Table of Contents

- COVER
- PREFACE
- ACKNOWLEDGMENTS
- PART ONE: Information Governance Concepts, Definitions, and Principles
  - CHAPTER 1: The Information Governance Imperative
    - Early Development of IG
    - Big Data Impact
    - Defining Information Governance
    - IG Is Not a Project, But an Ongoing Program
    - Why IG Is Good Business
    - Failures in Information Governance
    - Form IG Policies, Then Apply Technology for Enforcement
    - Notes
  - CHAPTER 2: Information Governance, IT Governance, Data Governance: What's the Difference?
    - Data Governance
    - Data Governance Strategy Tips
    - IT Governance
    - IT Governance Frameworks
    - Information Governance
    - Impact of a Successful IG Program
    - Summing Up the Differences
    - Notes
  - CHAPTER 3: Information Governance Principles
    - The Sedona Conference® Commentary on Information Governance
    - Smallwood IG Principles
    - Accountability Is Key
    - Generally Accepted Recordkeeping Principles®
    - Assessment and Improvement Roadmap
    - Information Security Principles
    - Privacy Principles
    - Who Should Determine IG Policies?
    - Notes
- PART TWO: Information Governance Risk Assessment and Strategic Planning

CHAPTER 4: Information Asset Risk Planning and Management
The Information Risk Planning Process
Create a Risk Profile
Information Risk Planning and Management Summary
Notes
CHAPTER 5: Strategic Planning and Best Practices for Information Governance
Crucial Executive Sponsor Role
Evolving Role of the Executive Sponsor
Building Your IG Team
Assigning IG Team Roles and Responsibilities
Align Your IG Plan with Organizational Strategic Plans
Survey and Evaluate External Factors
Formulating the IG Strategic Plan
Notes
CHAPTER 6: Information Governance Policy Development
The Sedona Conference IG Principles
A Brief Review of Generally Accepted Recordkeeping Principles®
IG Reference Model
Best Practices Considerations
Standards Considerations
Benefits and Risks of Standards
Key Standards Relevant to IG Efforts
Major National and Regional ERM Standards
Making Your Best Practices and Standards Selections to Inform Your IG Framework
Roles and Responsibilities
Program Communications and Training
Program Controls, Monitoring, Auditing, and Enforcement
Notes
PART THREE: Information Governance Key Impact Areas
CHAPTER 7: Information Governance for Business Units
Start with Business Objective Alignment
Which Business Units Are the Best Candidates to Pilot an IG Program?
What Is Infonomics?
How to Begin an IG Program

Business Considerations for an IG Program
Changing Information Environment
Calculating Information Costs
Big Data Opportunities and Challenges
Full Cost Accounting for Information
Calculating the Cost of Owning Unstructured Information
The Path to Information Value
Challenging the Culture
New Information Models
Future State: What Will the IG-Enabled Organization Look Like?
Moving Forward
Notes
CHAPTER 8: Information Governance and Legal Functions
Introduction to E-Discovery: The Revised 2006 and 2015 Federal Rules of Civil Procedure Changed Everything
Big Data Impact
More Details on the Revised FRCP Rules
Landmark E-Discovery Case: Zubulake v. UBS Warburg
E-Discovery Techniques
E-Discovery Reference Model
The Intersection of IG and E-Discovery
Building on Legal Hold Programs to Launch Defensible Disposition
Destructive Retention of E-Mail
Newer Technologies That Can Assist in E-Discovery
Defensible Disposal: The Only Real Way to Manage Terabytes and Petabytes
Notes
CHAPTER 9: Information Governance and Records and Information Management Functions
Records Management Business Rationale
Why Is Records Management So Challenging?
Benefits of Electronic Records Management
Additional Intangible Benefits
Inventorying E-Records
RM Intersection with Data Privacy Management
Generally Accepted Recordkeeping Principles®
E-Records Inventory Challenges

Records Inventory Purposes
Records Inventorying Steps
Appraising the Value of Records
Ensuring Adoption and Compliance of RM Policy
Sample Information Asset Survey Questions
General Principles of a Retention Scheduling
Developing a Records Retention Schedule
Why Are Retention Schedules Needed?
What Records Do You Have to Schedule? Inventory and Classification
Rationale for Records Groupings
Records Series Identification and Classification
Retention of E-Mail Records
How Long Should You Keep Old E-Mails?
Destructive Retention of E-Mail
Legal Requirements and Compliance Research
Event-Based Retention Scheduling for Disposition of E-Records
Prerequisites for Event-Based Disposition
Final Disposition and Closure Criteria
Retaining Transitory Records
Implementation of the Retention Schedule and Disposal of Records
Ongoing Maintenance of the Retention Schedule
Audit to Manage Compliance with the Retention Schedule
Notes
CHAPTER 10: Information Governance and Information Technology Functions
Data Governance
Steps to Governing Data Effectively
Data Governance Framework
Information Management
IT Governance
IG Best Practices for Database Security and Compliance
Tying It All Together
Notes
CHAPTER 11: Information Governance and Privacy and Security Functions
Information Privacy

Generally Accepted Privacy Principles
Fair Information Practices (FIPS)
OCED Privacy Principles
Madrid Resolution 2009
EU General Data Protection Regulation
GDPR: A Look at Its First Year
Privacy Programs
Privacy in the United States
Privacy Laws
Cybersecurity
Cyberattacks Proliferate
Insider Threat: Malicious or Not
Information Security Assessments and Awareness Training
Cybersecurity Considerations and Approaches
Defense in Depth
Controlling Access Using Identity Access Management
Enforcing IG: Protect Files with Rules and Permissions
Challenge of Securing Confidential E-Documents
Apply Better Technology for Better Enforcement in the Extended Enterprise
E-Mail Encryption
Secure Communications Using Record-Free E-Mail
Digital Signatures
Document Encryption
Data Loss Prevention (DLP) Technology
Missing Piece: Information Rights Management (IRM)
Embedded Protection
Hybrid Approach: Combining DLP and IRM Technologies
Securing Trade Secrets After Layoffs and Terminations
Persistently Protecting Blueprints and CAD Documents
Securing Internal Price Lists
Approaches for Securing Data Once It Leaves the Organization
Document Labeling
Document Analytics
Confidential Stream Messaging
Notes

PART FOUR: Information Governance for Delivery Platforms
CHAPTER 12: Information Governance for E-Mail and Instant Messaging
Employees Regularly Expose Organizations to E-Mail Risk
E-Mail Polices Should Be Realistic and Technology Agnostic
E-Record Retention: Fundamentally a Legal Issue
Preserve E-Mail Integrity and Admissibility with Automatic Archiving
Instant Messaging
Best Practices for Business IM Use
Technology to Monitor IM
Tips for Safer IM
Team and Channel Messaging Solutions Emerge
Notes
CHAPTER 13: Information Governance for Social Media
Types of Social Media in Web 2.0
Additional Social Media Categories
Social Media in the Enterprise
Key Ways Social Media Is Different from E-Mail and Instant Messaging
Biggest Risks of Social Media
Legal Risks of Social Media Posts
Tools to Archive Social Media
IG Considerations for Social Media
Key Social Media Policy Guidelines
Records Management and Litigation Considerations for Social Media
Emerging Best Practices for Managing Social Media Records
Notes
CHAPTER 14: Information Governance for Mobile Devices
Current Trends in Mobile Computing
Security Risks of Mobile Computing
Securing Mobile Data
Mobile Device Management (MDM)
IG for Mobile Computing
Building Security into Mobile Applications
Best Practices to Secure Mobile Applications
Developing Mobile Device Policies
Notes

CHAPTER 15: Information Governance for Cloud Computing
Defining Cloud Computing
Key Characteristics of Cloud Computing
What Cloud Computing Really Means
Cloud Deployment Models
Benefits of the Cloud
Security Threats with Cloud Computing
Managing Documents and Records in the Cloud
IG Guidelines for Cloud Computing Solutions
IG for SharePoint and Office365
Notes
CHAPTER 16: Leveraging and Governing Emerging Technologies
Data Analytics
Descriptive Analytics
Diagnostic Analytics
Predictive Analytics
Prescriptive Analytics
Which Type of Analytics Is Best?
Artificial Intelligence
The Role of Artificial Intelligence in IG
Blockchain: A New Approach with Clear Advantages
Breaking Down the Definition of Blockchain
The Internet of Things: IG Challenges
IoT as a System of Contracts
IoT Basic Risks and IG Issues
IoT E-Discovery Issues
Why IoT Trustworthiness Is a Journey and Not a Project
Governing the IoT Data
IoT Trustworthiness
Information Governance Versus IoT Trustworthiness
IoT Trustworthiness Journey
Conclusion
Notes
PART FIVE: Long-Term Program Issues
CHAPTER 17: Long-Term Digital Preservation
Defining Long-Term Digital Preservation

Key Factors in Long-Term Digital Preservation
Threats to Preserving Records
Digital Preservation Standards
PREMIS Preservation Metadata Standard
Recommended Open Standard Technology–Neutral Formats
Digital Preservation Requirements
Long-Term Digital Preservation Capability Maturity Model®
Scope of the Capability Maturity Model
Digital Preservation Capability Performance Metrics
Digital Preservation Strategies and Techniques
Evolving Marketplace
Looking Forward
Conclusion
Notes
CHAPTER 18: Maintaining an Information Governance Program and Culture of Compliance
Monitoring and Accountability
Change Management—Required
Continuous Process Improvement
Why Continuous Improvement Is Needed
Notes
APPENDIX A: Information Organization and Classification: Taxonomies and Metadata
Importance of Navigation and Classification
When Is a New Taxonomy Needed?
Taxonomies Improve Search Results
Metadata and Taxonomy
Metadata Governance, Standards, and Strategies
Types of Metadata
Core Metadata Issues
International Metadata Standards and Guidance
Records Grouping Rationale
Business Classification Scheme, File Plans, and Taxonomy
Classification and Taxonomy
Prebuilt Versus Custom Taxonomies
Thesaurus Use in Taxonomies
Taxonomy Types

Business Process Analysis
Taxonomy Testing: A Necessary Step
Taxonomy Maintenance
Social Tagging and Folksonomies
Endnotes
APPENDIX B: Laws and Major Regulations Related to Records Management
United States
Gramm-Leach-Bliley Act
Healthcare Insurance Portability and Accountability Act of 1996 (HIPAA)
PATRIOT Act (Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001)
Sarbanes-Oxley Act (SOX)
SEC Rule 17A-4
CFR Title 47, Part 42—Telecommunications
CFR Title 21, Part 11—Pharmaceuticals
US Federal Authority on Archives and Records: National Archives and Records Administration (NARA)
US Code of Federal Regulations
Canada*
United Kingdom
Australia
Identifying Records Management Requirements in Other Legislation
Notes
APPENDIX C: Laws and Major Regulations Related to Privacy
United States
European Union General Data Protection Regulation (GDPR)
Major Privacy Laws Worldwide, by Country
Notes
GLOSSARY
Notes
ABOUT THE AUTHOR
ABOUT THE MAJOR CONTRIBUTORS
INDEX
END USER LICENSE AGREEMENT

List of Tables

Chapter 3

Table 3.1 Generally Accepted Recordkeeping Principles Levels

Table 3.2 Improvement Areas for Generally Accepted Recordkeeping Principles

Table 3.3 Assessment Report and Roadmap

Chapter 4

Table 4.1 Risk Assessment

Chapter 7

Table 7.1 Key Steps in the IG Process

Chapter 8

Table 8.1 IG Impact on E-Discovery

Chapter 13

Table 13.1 Social Media Categories and Examples of Tools Commonly Used by Marke...

Table 13.2 Social Media Tools Used by NARA

Table 13.3 Additional Social Media Tools by Application Type

Table 13.4 Social Media Archiving and Management Software

Chapter 17

Table 17.1 Recommended Open Standard Technology-Neutral Formats

Table 17.2 Digital Preservation Performance Metrics

List of Illustrations

Chapter 6

Figure 6.1 Information Governance Reference Model

Chapter 7

Figure 7.1 Key Factors Driving Cost

Chapter 8

Figure 8.1 Electronic Discovery Reference Model

Chapter 9

Figure 9.4 Sample Records Retention Schedule

Figure 9.5 Excerpt from Canadian Records Retention Database

Chapter 10

Figure 10.1 DGI Data Governance Framework™

Figure 10.2 Key Steps from Data Modeling to Integration
Figure 10.3 Categories of Data
Chapter 17
Figure 17.1 Open Archival Information System Reference Model
Figure 17.2 PREMIS Data Model
Figure 17.3 Five Levels of Digital Preservation Capabilities
Figure 17.4 Digital Preservation Capability Maturity Model
Appendix A
Figure A.1 Metadata Link to Taxonomy Example
Figure A.2 Application of Metadata to Taxonomy Structure
Figure A.3 Mapping the Records Retention Schedule to the Taxonomy
Figure A.4 Library of Congress Subject Headings
Figure A.5 Yellow Pages Example
Figure A.6 Community Government Business-Unit Taxonomy
Figure A.7 State Government Regulatory Agency Functional Taxonomy
Figure A.8 Metadata Cross-Referencing within a Taxonomy
Figure A.9 Basic Accounting Business-Unit Taxonomy
Figure A.10 Business Process Example—Travel Expense Process

Founded in 1807, John Wiley & Sons is the oldest independent publishing company in the United States. With offices in North America, Europe, Asia, and Australia, Wiley is globally committed to developing and marketing print and electronic products and services for our customers' professional and personal knowledge and understanding.

The Wiley CIO series provides information, tools, and insights to IT executives and managers. The products in this series cover a wide range of topics that supply strategic and implementation guidance on the latest technology trends, leadership, and emerging best practices.

Titles in the Wiley CIO series include:

The Agile Architecture Revolution: How Cloud Computing, REST-Based SOA, and Mobile Computing Are Changing Enterprise IT by Jason Bloomberg

Architecting the Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS, and IaaS) by Michael Kavis

Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses by Michael Minelli, Michele Chambers, and Ambiga Dhiraj

The Chief Information Officer's Body of Knowledge: People, Process, and Technology by Dean Lane

Cloud Computing and Electronic Discovery by James P. Martin and Harry Cendrowski

Confessions of a Successful CIO: How the Best CIOs Tackle Their Toughest Business Challenges by Dan Roberts and Brian Watson

CIO Best Practices: Enabling Strategic Value with Information Technology (Second Edition) by Joe Stenzel, Randy Betancourt, Gary Cokins, Alyssa Farrell, Bill Flemming, Michael H. Hugos, Jonathan Hujsak, and Karl Schubert

The CIO Playbook: Strategies and Best Practices for IT Leaders to Deliver Value by Nicholas R. Colisto

Decoding the IT Value Problem: An Executive Guide for Achieving Optimal ROI on Critical IT Investments by Gregory J. Fell

Enterprise Performance Management Done Right: An Operating System for Your Organization by Ron Dimon

Information Governance: Concepts, Strategies and Best Practices by Robert F. Smallwood

IT Leadership Manual: Roadmap to Becoming a Trusted Business Partner by Alan R. Guibord

Leading the Epic Revolution: How CIOs Drive Innovation and Create Value

Across the Enterprise by Hunter Muller

Managing Electronic Records: Methods, Best Practices, and Technologies by Robert F. Smallwood

On Top of the Cloud: How CIOs Leverage New Technologies to Drive Change and Build Value Across the Enterprise by Hunter Muller

Straight to the Top: CIO Leadership in a Mobile, Social, and Cloud-based World (Second Edition) by Gregory S. Smith

Strategic IT: Best Practices for Managers and Executives by Arthur M. Langer and Lyle Yorks

Trust and Partnership: Strategic IT Management for Turbulent Times by Robert Benson, Piet Ribbers, and Ronald Billstein

Transforming IT Culture: How to Use Social Intelligence, Human Factors, and Collaboration to Create an IT Department That Outperforms by Frank Wander

Unleashing the Power of IT: Bringing People, Business, and Technology Together, Second Edition by Dan Roberts

The U.S. Technology Skills Gap: What Every Technology Executive Must Know to Save America's Future by Gary J. Beach

INFORMATION GOVERNANCE

CONCEPTS, STRATEGIES, AND BEST PRACTICES

SECOND EDITION

Robert F. Smallwood

WILEY

Copyright © 2020 by Robert F. Smallwood. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the Web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at www.wiley.com/go/permissions.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993, or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

***Library of Congress Cataloging-in-Publication Data*:**

Names: Smallwood, Robert F., 1959- author.

Title: Information governance: concepts, strategies, and best practices / Robert F. Smallwood.

Description: Second edition. | Hoboken, New Jersey: John Wiley & Sons, Inc., [2020] | Series: The Wiley CIO series | Includes index. |

Identifiers: LCCN 2019015574 (print) | LCCN 2019017654 (ebook) | ISBN 9781119491415 (Adobe PDF) | ISBN 9781119491408 (ePub) | ISBN 9781119491446 (hardback)

Subjects: LCSH: Information technology—Management. | Management information systems. | Electronic records—Management.

Classification: LCC HD30.2 (ebook) | LCC HD30.2 .S617 2020 (print) | DDC 658.4/038—dc23

LC record available at https://lccn.loc.gov/2019015574

Cover Design: Wiley

Cover Image: ©style_TTT/Shutterstock

For my sons
and the next generation of tech-savvy managers

PREFACE

In the five plus years since the first edition of this book was published, information governance (IG) has matured as a discipline, and business executives and managers at leading enterprises now see IG programs as increasingly valuable. A combination of factors have created an imperative for IG programs: new, tightened regulations; the continuing deluge of Big Data; and the realization that new value can be gained from information stores using analytics have all combined to raise the profile of IG programs across the globe.

In particular, new privacy legislation, including the EU General Data Protection Regulation and the California Consumer Privacy Act, helped foster a newfound awareness of data protection issues, and organizations worldwide scrambled to inventory and gain insight into their information stores. This is often a first step in IG programs, and so the realization of IG as a needed and valued undertaking set in. Enterprises began to see IG not only as a cost center and risk reduction activity, but also as one that can add value to the enterprise, in some cases even monetizing information.

This book clarifies and codifies what IG is—and what it is not—and how to launch, control, and manage IG programs. Based on exhaustive research, and with the contributions of a number of industry pioneers and experts, this book lays out IG as a complete discipline, fully updated, including an expanded section on information privacy and new material on managing emerging technologies.

IG is a "super-discipline" of sorts in that it includes components of privacy, cybersecurity, infonomics, law and e-discovery, records management, compliance, risk management, information technology (IT), business operations, and more. This unique blend calls for a new breed of information professional who is competent across these complex disciplines. Training and education are key to IG program success, and this book provides the fundamentals as well as advanced concepts to enable organizations to train a new generation of IG professionals. The book is being used to guide IG programs at major corporations, as well as to educate graduate students in information science, computer science, law, and business.

Practitioners in the component areas of IG will find the book useful in expanding their knowledge and helping them understand the linkages between the various facets of IG. And how breaking down existing siloed approaches and leveraging information as an asset across the enterprise is critical to gaining the full benefits of IG programs.

The book strives to offer clear and concise IG concepts, actionable strategies, and proven best practices in an understandable and digestible way; a concerted effort was made to simplify language and offer examples. There are summaries of key points throughout the book and at the end of each chapter to help the reader retain key points. The text is organized into five parts: (1) IG Concepts, Definitions, and

Principles; (2) IG Risk Assessment and Strategic Planning; (3) IG Key Impact Areas; (4) IG for Information Delivery Platforms, including a new section on emerging technologies; and (5) Long-Term Program Issues.

No other book offers comprehensive coverage of the complex and challenging field of IG with such clarity. Use the insights and advice contained in these pages and your IG program will have lower risks and costs, and produce better and more measurable results.

Robert Smallwood

ACKNOWLEDGMENTS

I would like to gratefully thank my colleagues for the support and generous contributions of their expertise and time, which made this updated and comprehensive text possible.

Many thanks to Lori Ashley, Jason R. Baron, Barb Blackburn, Barclay Blair, Robert Bogue, Charmaine Brooks, Baird Brueseke, Ken Chasse, Monica Crocker, Charles Dollar, Mark Driskill, Seth Early, Sam Fossett, Dr. Patricia Franks, Randy Kahn, Dennis Kessler, Darra Hoffman, Doug Laney, Paula Lederman, Reynold Leming, Barry Murphy, Robert Seiner, Teresa Schoch, Andrew Ysasi, and Bassam Zarkout.

I am truly honored to include their insightful work and owe them a great debt of gratitude.

PART ONE

Information Governance Concepts, Definitions, and Principles

CHAPTER 1
The Information Governance Imperative

Effective **information governance** (IG) programs improve operational efficiency and compliance capabilities while leveraging information as an asset to maximize their value. Active IG programs are the hallmark of well-managed organizations, and increasingly IG has become an imperative, especially for global enterprises.

A "perfect storm" of compliance pressures, cybersecurity concerns, Big Data volumes, and the increasing recognition that information itself has value have contributed to a substantial increase in the number of organizations implementing IG programs.

Most significantly, the European Union (EU) General Data Protection Regulation (GDPR), which went into effect May 25, 2018, left companies across the globe scrambling to gain control over the consumer data they had housed. The GDPR legislation applies to all citizens in the EU and the European Economic Area (EEA), regardless of where they reside, and also visitors and temporary residents of the EU. So any global enterprise doing business with EU/EEA citizens—or even visitors—must comply with the legislation or face a major fine. The primary goal of GDPR is to give citizens control over their personal data.

Brought about in part because of GDPR compliance concerns, membership in the International Association of Privacy Professionals (IAPP) grew from around 25,000 members in 2017 to over 40,000 members in 2018, and it continues to grow.

A first step in the GDPR compliance process is to conduct an inventory of an enterprise's information assets to create a data map showing where all incidences of data are housed. This is commonly the first major implementation step in IG programs, so the IG discipline and support for IG programs made substantial strides in 2018 with the lead-up to GDPR going into effect. Then California passed its California Consumer Privacy Act (CCPA), which borrowed many concepts from GDPR and required that any company (of a certain size) handling the personally identifiable information (PII) of California residents (in specified volumes) comply by January 1, 2020. Suddenly US-based companies could no longer ignore privacy regulations, and the momentum for IG programs that could manage privacy compliance requirements accelerated.

During this same time frame, data breaches and ransomware attacks became more prevalent and damaging, and organizations adopted IG programs to reduce the likelihood of cyberattacks, and their impact. IG programs implement effective risk reduction countermeasures.

A first step in the GDPR compliance process is to conduct an inventory of an enterprise's information assets to create a data map.

Added to that has been the continued massive increase on overall data volumes that organizations must manage, which results in managing a lot of unknown "dark data," which lacks metadata and has not been classified. Organizations also retain large volumes of redundant, outdated, and trivial (ROT) information that needs to be identified and disposed of. Cleaning up the ROT that organizations manage reduces their overall storage footprint and costs, and makes information easier to fine, leading to improved productivity for knowledge workers.

IG programs are also about optimizing and finding new value in information. The concept of managing and monetizing information is core to the emerging field of **infonomics**, which is the discipline that assigns "economic significance" to information and provides a framework to manage, measure, and monetize information.[1] Gartner's former analyst Doug Laney published a groundbreaking book in 2018, *Infonomics,* which delineates infonomics principles in great detail, providing many examples of ways organizations have harvested new value by finding ways to monetize information or leverage its value.

> Infonomics is the discipline that assigns "economic significance" to information and provides a framework to manage, measure, and monetize information.

Early Development of IG

IG has its roots in the United Kingdom's healthcare system. Across the pond, the government-funded National Health Service (NHS) has focused on IG to ensure data quality and protect patient data since 2002. Although IG was mentioned in journals and scholarly articles decades ago, the UK is arguably the home of healthcare IG, and perhaps the IG discipline.[2] Could this be the reason the UK leads the world in healthcare quality? Certainly, it must be a major contributing factor.

The United States has the most expensive healthcare in the world, the most sophisticated equipment, the most advanced medicines, the best-trained doctors—yet in a recent study of healthcare quality, the United States came in dead last out of 11 civilized nations.[3] The UK, Switzerland, and Sweden topped the list.

The U.S. healthcare problem is not due to poor training, inferior equipment, inferior medicines, or lack of financial resources. No, the problem is likely primarily *a failure to get the right information to the right people at the right time*; that is, caregivers must have accurate, current clinical information to do their jobs properly. These are IG issues.

Since 2002 each UK healthcare organization has been tasked with completing the IG Toolkit, managed by NHS Digital for the UK Department of Health. Although the IG Toolkit has evolved over the years, its core has remained constant. However, in April 2018 it was replaced with a new tool, the Data Security and Protection Toolkit, based around 10 National Data Security Standards that have been formulated by the

UK's National Data Guardian.[4]

Big Data Impact

According to the research group Gartner, Inc., Big Data is defined as "... high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."[5] A practical definition should also include the idea that the amount of information—both **structured data** (in databases) and **unstructured information** (e.g. e-mail, scanned documents, PDFs, MS Office documents) is so massive that it cannot be processed using traditional database tools (e.g. DB2, SQL) and analytic software techniques.[6]

In today's information overload era of Big Data—characterized by massive growth in business data volumes and velocity—the ability to distill key insights from enormous amounts of data is a major business differentiator and source of sustainable competitive advantage. In fact, a report by the World Economic Forum stated that data is a new asset class and personal data is "the new oil."[7] And we are generating more than we can manage effectively with current methods and tools.

The Big Data numbers are overwhelming: Estimates and projections vary, but it has been stated that 90% of the data existing worldwide today was created in the past two years,[8] and that every two days more information is generated than was from the dawn of civilization until 2003.[9] This trend will continue.

Certainly, there are new and emerging opportunities arising from the accumulation and analysis of all that data we are busy generating and collecting. New enterprises are springing up to capitalize on data mining and business analytics opportunities. Back in 2012, the US federal government joined in, announcing $200 million in Big Data research programs.[10]

> The onslaught of Big Data necessitates that IG be implemented to discard unneeded data in a legally defensible way.

However, established organizations, especially larger ones, are being crushed by this onslaught of Big Data: it is just too expensive to keep all the information that is being generated, and unneeded and ROT information becomes a sort of irrelevant sludge of data debris for decision makers to wade through. They have difficulty knowing which information is accurate and meaningful "signal," and which is simply irrelevant "noise." This means they do not have the precise information on which they can base good business decisions.

And it has real costs: the burden of massive stores of information has increased storage costs dramatically, caused overloaded systems to fail, and increased legal discovery costs.[11] Furthermore, the longer that data is kept the more likely that it will need to be migrated to newer computing platforms, driving up conversion costs;

and legally, there is the risk that somewhere in that mountain of data an organization keeps is a piece of information that represents a significant legal liability.[12]

This is where the worlds of Big Data and business collide. For Big Data proponents, more data is always better, and there is no perceived downside to the accumulation of massive amounts of data. In the business world, though, the realities of legal **e-discovery** mean the opposite is true.[13] To reduce risk, liability, and costs, it is critical for unneeded or useless information to be disposed of in a systematic, methodical, and "legally defensible" (justifiable in legal proceedings) way, when it no longer has legal, regulatory, or business value.

> Big Data values massive accumulation of data whereas in business, e-discovery realities and potential legal liabilities dictate that data be culled down to only that which has clear business value.

Organizations are struggling to reduce and right-size their information footprint by discarding superfluous and redundant data, **e-documents**, and information. *But the critical issue is devising policies, methods, and processes, and then deploying information technology (IT) to sort through the information and determine what is valuable and what no longer has value and can be discarded.*

IT, compliance, and legal representatives in organizations have a clear sense that most of the information stored is unneeded, raises costs, and poses risks. According to a survey by the Compliance, Governance and Oversight Council (CGOC), respondents estimated that approximately one-quarter of information stored in organizations has real business value, while 5% must be kept as business records, and about 1% is retained due to a litigation hold.[14] This means that [about] 69% of information in most companies has no business, legal or regulatory value. "Companies that are able to dispose of this debris return more profit to shareholders, can use more of their IT budgets for strategic investments, and can avoid excess expense in legal and regulatory response" [italics added].

> Only about one-quarter of the information that organizations are managing has real business value.

*With a smaller **information footprint**, organizations can more easily find what they need and derive business value from it.*[15] They must eliminate the data debris regularly and consistently, and to do this, processes and systems must be in place to cull out valuable information and discard the data debris. An IG program sets the framework to accomplish this.

The business environment has also underscored the need for IG. According to Ted Friedman at Gartner, "The recent global financial crisis has put information

governance in the spotlight.... [it] is a priority of IT and business leaders as a result of various pressures, including regulatory compliance mandates and the urgent need for improved decision-making."[16]

And IG mastery is critical for executives: many CIOs in regulated industries have been fired from their jobs for failed IG initiatives.[17]

> With a smaller information footprint, it is easier for organizations to find the information they need and derive business value from it.

Defining Information Governance

Information governance is a sort of "super discipline" that has emerged as a result of new and tightened legislation governing businesses, privacy concerns, legal demands, external pressures such as hacking and data breaches, and the recognition that multiple overlapping disciplines were needed to address today's information management challenges in an increasingly regulated and litigated business environment.[18]

IG is a subset of corporate governance, and includes key concepts from information security, data privacy and protection, records and information management (RIM), content management, IT and data governance, risk management, litigation readiness, regulatory compliance, **long-term digital preservation** (LTDP), and even analytics and information economics, (infonomics). This also means that it includes related technology and discipline subcategories such as **document management**, enterprise search, knowledge management, and **disaster recovery (DR)/business continuity (BC)**.

> Information governance is a subset of corporate governance.

Practicing good IG is the essential foundation for building legally defensible disposition practices to discard unneeded information, and to secure confidential, sensitive, and secret information, which may include trade secrets, strategic plans, price lists, blueprints, or personal information subject to privacy laws. Good IG provides the basis for consistent, reliable methods for managing, securing, controlling, and optimizing information.

Having trusted and reliable records, reports, data, and databases allows managers to make key decisions with confidence.[19] And accessing that information and data analytics insights in a timely fashion can yield a long-term sustainable competitive advantage, creating more agile enterprises.

IG is a sort of "super discipline" that encompasses a variety of key concepts from a variety of related disciplines.

To do this, organizations must standardize and systematize their handling of information, and audit their processes to ensure so. They must analyze and optimize how information is accessed, controlled, managed, shared, stored, preserved, and audited. They must have complete, current, and relevant policies, processes, and technologies to manage and control information, including *who* is able to access what information, and *when,* to meet external legal and regulatory demands and internal governance policy requirements. The idea is to provide the right information to the right people at the right time—securely. *Security, control, and optimization of information;* this, in short, is IG.

Practicing good IG is the essential foundation for building legally defensible disposition practices to discard unneeded information.

Information governance is a subset of corporate governance, which has been around as long as corporations have existed. IG is a rather new multidisciplinary field that is still being defined, but has gained significant traction in the past several years. The focus on IG comes not only from privacy, cybersecurity, compliance, legal, and records management functionaries, but also from executives who understand they are accountable for the governance of information, and that theft or erosion of information assets has real costs and consequences. It can cause corporate brand equity to collapse, and stock price to tumble.

IG is an all-encompassing term for *how an organization manages the totality of its information*.

Information governance programs are about minimizing information risks and costs, while maximizing its value. In short, IG is the security, control, and optimization of information.

Information governance programs are about minimizing information risks and costs, while maximizing its value. IG is control of information to meet business, legal, regulatory, and risk demands.

Stated differently, information governance is "a quality-control discipline for managing, using, improving, and protecting information."[20]

Unpacking the definition further: "Information governance is policy-based management of information designed to lower costs, reduce risk, and ensure compliance with legal, regulatory standards, and/or corporate governance."[21] IG necessarily incorporates not just policies and processes, but information

technologies to audit and enforce them. The IG team must be cognizant of information lifecycle issues, and be able to apply the proper retention and disposition policies, including digital preservation, where e-records of documents need to be maintained for long periods.

> Information governance is how an organization maintains security, complies with regulations, and meets ethical standards when managing information.

IG Is Not a Project, But an Ongoing Program

IG is an ongoing program, not a one-off project. IG provides a policy umbrella to manage and control information. Since technologies change so quickly, it is necessary to have overarching policies that can manage the various information technology (IT) platforms that an organization may use.

Compare it to a workplace safety program; every time a new location, team member, piece of equipment, or toxic substance is acquired by the organization, the workplace safety program should dictate how that is handled, and, if it doesn't, the workplace safety policies/procedures/training that are part of the workplace safety program need to be updated. Regular reviews are conducted to ensure the program is being followed and make adjustments based on your findings. *The effort never ends.*[22] The same is true for IG programs. They should continually be evaluated against established metrics, and should continue to expand and extend deeper into the enterprise.

IG is not only a tactical program to meet regulatory, compliance, and litigation demands. It can be strategic, in that it is the necessary underpinning for a management strategy that maximizes knowledge worker productivity, while minimizing risk and costs. Further, it treats information as an asset and seeks to maximize its value—perhaps even finding and harvesting newfound value.

> IG is a multidisciplinary program that requires an ongoing effort.

Why IG Is Good Business

IG is a tough sell. It can be difficult to make the **business case** for it, unless there has been some major compliance sanction, fine, legal loss, or colossal data breach. Doug Laney calls this "blunderfunding" in that organizations wait until a major blunder before they fund a program. In fact, *the largest impediment to IG adoption is simply identifying its benefits and costs,* according to The Economist Intelligence Unit. Sure, the enterprise needs better control over its information, but how much better? At what cost? What is the payback period and the **return on investment** (ROI)?[23]

It is challenging to make the business case for IG, yet making that case is fundamental to getting IG efforts off the ground.

Here are 10 reasons why IG makes good business sense:

1. **We can't keep everything forever.** IG makes sense because it enables organizations to get rid of unnecessary information in a defensible manner. Organizations need a sensible way to dispose of information in order to reduce the cost and complexity of the IT environment. Having unnecessary information around only makes it more difficult and expensive to harness information that has value.
2. **We can't throw everything away.** IG makes sense because organizations can't keep everything forever, nor can they throw everything away. We need information—the right information, in the right place, at the right time. Only IG provides the framework to make good decisions about what information to keep.
3. **E-discovery.** IG makes sense because it reduces the cost and pain of discovery. Proactively managing information reduces the volume of information exposed to e-discovery and simplifies the task of finding and producing responsive information.
4. **Your employees are screaming for it—just listen**. IG makes sense because it helps knowledge workers separate "signal" from "noise" in their information flows. By helping organizations focus on the most valuable information, IG improves information delivery and improves productivity.
5. **It ain't gonna get any easier.** IG makes sense because it is a proven way for organizations to respond to new laws and technologies that create new requirements and challenges. The problem of IG will not get easier over time, so organizations should get started now.
6. **The courts will come looking for IG**. IG makes sense because courts and regulators will closely examine your IG program. Falling short can lead to fines, sanctions, loss of cases, and other outcomes that have negative business and financial consequences.
7. **Manage risk: IG is a big one**. Organizations need to do a better job of identifying and managing risk. The risk of information management failures is a critical risk that IG helps to mitigate.
8. **E-mail: reason enough**. IG makes sense because it helps organizations take control of e-mail. Solving e-mail should be a top priority for every organization.[24]
9. **Privacy compliance.** With the advent of the EU GDPR legislation, and increasing privacy concerns globally, forward-thinking enterprises are implementing IG programs.
10. **Infonomics**. Enterprises are looking for innovative ways to leverage information as an asset and to find new value.

Failures in Information Governance

Associates in Psychiatry and Psychology

This first example is a result of weaknesses in cybersecurity and IG procedures, but some good steps were taken in advance to reduce the impact of any cyberattack. And the response is perhaps a model for how organizations *should* handle a ransomware attack.

Associates in Psychiatry and Psychology (APP) in Rochester, Minnesota, revealed that a ransomware attack occurred in March 2018. The ransomware attack affected patient information for over 6,500 individuals, although, in the preliminary investigation, it appeared that the information was not in a "human-readable" format and that the protected health information wasn't accessed or copied by the attackers.

APP had a prompt response to the attack, taking their systems offline. Doing so in a timely manner likely stopped the spread of the attack and limited possible encryption of personal data and data theft, completing the "ransom" aspect of the ransomware attack.

APP, in a Q&A regarding the incident, reported that it was a "Triple-M" ransomware attack. This variation uses the RSA-2048 encryption protocol, which utilizes long keys in order to encrypt the data. A ransom was paid, as the **backups** with the restore files couldn't be accessed based on the attack. The initial ransom demand of 4 Bitcoin ($30,000) was not paid and instead negotiated down to .5 BTC ($3,800). With the systems and data now restored, APP installed additional layers of security as well as new remote-access policies.

Ransomware attacks are not unique, especially within the healthcare sector. What was fascinating about this attack is the amount of information shared with affected patients and the openness with which APP talked about the breach. Most breaches go unnoticed in the public eye because very little information is shared with the general public, even those directly affected, especially if the data wasn't accessed or copied. APP's transparency provides affected parties the ability to understand how the breach affects them and what they can do to protect themselves.

Other organizations should stand up and take notice: APP's response should become the standard.

> Associates in Psychiatry and Psychology suffered a ransomware attack but steps taken in advance and in response to the attack should serve as a model for other organizations.

Chipotle Mexican Grill

Chipotle Mexican Grill strives to serve "food with integrity" that is fresh, not genetically modified, and never frozen. It is their corporate mantra. This is why the

multiple reports of foodborne illnesses at some Chipotle stores from 2016–2018 were so damning. And the stock market value reflected this, dropping over 40% in just three months during 2016. That's something like $5 billion in value that vanished due to the reputational damage wrought on the Chipotle brand. The stock had hit a high of about $750 in 2015, but as of August 2018 it had yet to break $500, after hitting a low of $255 earlier in the year.

Billions in value lost. That is *huge. And it was the result of poor IG practices at Chipotle.* But Chipotle did finally recover as of late 2019 when the stock was back up over $800.

There were reports that the food poisonings may have been the result of industrial espionage, and they may well be true. It is not so far-fetched—just such a scenario was played out in the Showtime series *Billions* where a new soft drink was tainted. Multiple parties benefited from Chipotle's losses at the time, so who had the most to gain? A hedge fund manager shorting the stock may have raked in billions. Competitors were able to improve their market standing during Chipotle's losses. And even some alternative (traditional) suppliers gained ground. Knowing that this risk was out there should have forced Chipotle executives to focus on taking measures to reduce that risk, but they did not do so quickly or completely enough until 2019.

And what was the crux of the matter? Information risk coming home to roost from poor IG. Chipotle managers did not have the proper level of detailed information to track exactly *where* in their supply chain their ingredients have been contaminated, which is a result of weaknesses in their IG and recordkeeping practices. In their 2016 investigation, the FBI pointed out that Chipotle's recordkeeping system was actually hindering the health authorities' investigation in locating the sources of the various infections.

All of this could have been prevented not only with improved food quality testing, but the detailed tracking of ingredient lot numbers and video surveillance. The fact that this information was not available to Chipotle managers was the result of a failure to construct proper records and information management (RIM) systems.

It was a failure of IG-but Chipotle eventually got back on track.

> Chipotle Mexican Grill suffered major declines in market value due to poor recordkeeping and IG failures when food poisonings occurred.

Anthem, Inc.

In 2016, a year after the largest healthcare data breach to date, where almost 80 million confidential records of members and employees at Anthem, Inc. were hacked, little had been learned about the nature, motivations, implications, and real costs of the breach.[25] According to Anthem the data breach affected several of its brands, including Anthem Blue Cross, Anthem Blue Cross and Blue Shield, Blue

Cross and Blue Shield of Georgia, Empire Blue Cross and Blue Shield, Amerigroup, Caremore, and UniCare.

Anthem, the nation's second largest health insurer, had insurance themselves—cyber-insurance. Perhaps that was why executives felt confident prior to the attack. Most of the initial costs were likely absorbed by a $100 million AIG cyber-insurance policy. But many class action lawsuits were filed, and "unresolved legal issues likely have stifled further disclosure of what is known."[26]

By law, Anthem was not required to encrypt the **personally identifiable information** (PII), although this is a standard industry best practice. Certainly, victims sued Anthem just on the basis that Anthem did not take proper care of their PII while in their custody.

The PII compromised included names, addresses, birthdates, social security numbers, medical IDs, e-mail addresses, and salary and employment information.[27] Anthem provided two years of credit monitoring for those who were affected. This was a mild measure, as hackers usually wait years to sell compromised data.

Certainly, Anthem's reputation was damaged, and the massive breach led to acquisition target Cigna questioning Anthem's IG posture, data privacy and security measures, and the resultant legal impact. In a letter, Cigna's CEO and former Board Chairman wrote, "Trust with customers and providers is critical in our industry, and Anthem has yet to demonstrate a path towards restoring this trust. We need to understand the litigation and potential liabilities, operational impact and long-term damage to Anthem's franchise as a result of this unprecedented data breach, as well as the governance and controls that resulted in this system failure."[28]

Anthem took steps to shore up its information security practices (blunderfunding), hiring cybersecurity firm Mandiant just after the attack. In addition, the National Association of Insurance Commissioners (NAIC) commissioned a "market conduct and financial exam" of the breach, but the report is classified.

Ford Motor Company

The failure to implement and enforce IG can lead to vulnerabilities that can have dire consequences. Ford Motor Company is reported to have suffered a loss estimated at $50–$100 million as a result of the theft of confidential documents by one of its own employees. A former product engineer who had access to thousands of trade secret documents and designs sold them to a competing Chinese car manufacturer. A strong IG program would have controlled and tracked access and prevented the theft while protecting valuable intellectual property.[29]

FBI

Law enforcement agencies have also suffered from poor IG. In a rather frivolous case in 2013 that highlighted the lack of policy enforcement for the mobile environment, it was reported that US FBI agents used government-issued mobile phones to send explicit text messages and nude photographs to coworkers. The incidents did not have a serious impact, but did compromise the agency and its

integrity, and "adversely affected the daily activities of several squads."[30] Proper mobile communications policies were obviously not developed and enforced.

Accenture

IG is also about information security and privacy, and serious thought must be given when creating policies that allow access to highly confidential information, as some schemes to compromise or steal information can be quite deceptive and devious, masked by standard operating procedures—if proper IG controls and monitoring are not in place. To wit: granting remote access to confidential information assets for key personnel is common. Granting medical leave is also common. But a deceptive and dishonest employee could feign a medical leave while downloading volumes of confidential information assets for a competitor—and that is exactly what happened at Accenture, a global consulting firm. During a fraudulent medical leave, an employee was allowed access to Accenture's Knowledge Exchange (KX), a detailed knowledge base containing previous proposals, expert reports, cost-estimating guidelines, and case studies. This activity could have been prevented by monitoring and analytics which would have shown an inordinate number of downloads—especially for an "ailing" employee. The employee then went to work for a direct competitor and continued to download the confidential information from Accenture, estimated to be as many as 1,000 critical documents. While the online access to KX was secure, the use of the electronic documents could have been restricted even *after* the documents were downloaded, if IG measures were in place and newer technologies (such as information rights management software or IRM) were deployed to secure them directly. With IRM, software security protections can be employed to seal the documents and control their use—even after they leave the organization. More detail on IRM technology and its capabilities is presented further on in this book.

Ford's loss from stolen documents in a single case of IP theft was estimated at $50–$100 million.

The list of breaches and IG failures could go on and on, more than filling the pages of this book. It is clear that it is occurring and that it will continue. *IG controls to safeguard confidential information assets and protect privacy cannot rely solely on the trustworthiness of employees and basic security measures*. It takes up-to-date IG policies and enforcement efforts and newer technology sets. It takes active, consistent monitoring and program adjustments to continue to improve.

Executives and senior managers can no longer avoid the issue, as it is abundantly clear that the threat is real and the costs of taking such avoidable risks can be high. A single security breach is an information governance failure and can cost the entire business. When organizations suffer high-profile data breaches, particularly when they involve consumer privacy, they suffer serious reputational damage, losses in market value, and are faced with potential fines or other sanctions.[31]

IG controls to safeguard confidential information assets and protect privacy cannot rely solely on the trustworthiness of employees and basic security measures.

Form IG Policies, Then Apply Technology for Enforcement

Typically, some policies governing the use and control of information and records may have been established for financial and compliance reports, and perhaps e-mail. But they are often incomplete and out-of-date, and have not been adjusted for changes in the business environment, such as new technology platforms (e.g., Web 2.0, social media), changing laws (e.g., California Consumer Privacy Act, U.S. FRCP 2006, 2015 changes), and additional regulations.

Further adding to the challenge is the rapid proliferation of mobile devices like tablets and smartphones used in business—information can be more easily lost or stolen—so IG efforts must be made to preserve and protect the enterprise's information assets.

Lasting and durable IG requires that policies are flexible enough not to hinder the proper flow of information in the heat of the business battle, yet strict enough to control and audit for misuse, policy violations, or security breaches. This is a continuous iterative policy-making process, which must be monitored and fine-tuned. Even with the absolute best efforts, some policies will miss the mark and need to be reviewed and adjusted.

Getting started with IG awareness is the crucial first step. It may have popped up on an executive's radar at one point or another and an effort might have been made, but many organizations leave these policies on the shelf and do not revise them on a regular basis.

IG is the necessary underpinning for a legally defensible disposition program that discards data debris and helps narrow the search for meaningful information on which to base business decisions. IG is also necessary to protect and preserve critical information assets, before their value can be exploited. An IG strategy should aim to minimize exposure to risk, at a reasonable cost level, while maximizing productivity and improving the quality of information delivered to knowledge users.

But a reactive, tactical *project* approach is not the way to go about it—haphazardly swatting at technological, legal, and regulatory flies. A proactive, strategic *program,* driven from the top-down with a clear, accountable executive sponsor and IG lead and an ongoing plan, auditing, and regular review process, is the only way to continuously adjust IG policies to keep them current so that they best serve the organization's needs.

Getting started with IG awareness is the crucial first step.

Some organizations have created formal governance bodies to establish strategies, policies, and procedures surrounding the distribution of information inside and outside the enterprise. These IG governance bodies, steering committees, or teams may include members from many different functional areas, since *proper IG necessitates cross-functional input from a variety of stakeholders*. Representatives from privacy, security, legal, IT, records management, risk management, compliance, operations, legal, finance, and perhaps analytics/data science, knowledge management, and human resources (for training and communications) are typically a part of IG teams. Often these efforts are jump-started and organized by an executive sponsor who utilizes third-party consulting resources that specialize in IG efforts, especially considering the newness of IG and its emerging **best practices**.

In this era of ever-growing privacy and security concerns, increased regulation, Big Data volumes, and infonomics opportunities, IG programs are playing an increasing role. Leveraging IG policies to focus on retaining the information that has real business value, while discarding the majority of information that has no value and carries associated increased costs and risks, is critical to success for modern enterprises. This must be accomplished in a systematic, consistent, and legally defensible manner for IG programs to succeeed.

CHAPTER SUMMARY: KEY POINTS

- A “perfect storm” of compliance pressures, cybersecurity concerns, Big Data volumes, and the increasing recognition that information itself has value have contributed to a significant increase in the number of organizations implementing IG programs.
- Infonomics, which is the discipline that asserts “economic significance” to information and provides a framework to manage, measure, and monetize information.
- IG has its roots in the United Kingdom's healthcare system.
- A first step in the GDPR compliance process is to conduct an inventory of an enterprise's information assets to create a data map.
- The onslaught of Big Data necessitates that IG be implemented to discard unneeded data in a legally defensible way.
- Big Data values massive accumulation of data whereas in business, e-discovery realities and potential legal liabilities dictate that data be culled down to only that which has clear business value.
- Only about one-quarter of information that organizations are managing has real business value.

- With a smaller information footprint, it is easier for organizations to find the information they need and derive business value from it.
- IG is a subset of corporate governance, and *encompasses the policies and leveraged technologies meant to manage what corporate information is retained, where, and for how long, and also how it is retained*.
- IG is a sort of "super discipline" that encompasses a variety of key concepts from a variety of related and overlapping disciplines.
- Practicing good IG is the essential foundation for building legally defensible disposition practices to discard unneeded information.
- Information Governance programs are about minimizing information risks and costs, while maximizing its value. In short, IG is the security, control, and optimization of information.
- Information governance is how an organization maintains security, complies with regulations and laws, and meets ethical standards when managing information.
- Associates in Psychiatry and Psychology suffered a ransomware attack but steps taken in advance and in response to the attack should serve as a model for other organizations.
- Chipotle Mexican Grill suffered major declines in market value due to poor recordkeeping and IG failures when food poisonings occurred.
- IG is a multidisciplinary program that requires an ongoing effort that requires active participation of a broad cross-section of functional groups and stakeholders.
- IG controls to safeguard confidential information assets and protect privacy cannot rely solely on the trustworthiness of employees and basic security measures.
- Getting started with IG awareness is the crucial first step.

Notes

1. Doug Laney, *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage* (New York: Bibliomotion/Taylor & Francis, 2018), 9.
2. Andrew Harvey and Barry Moult, e-mail to author, February 25, 2018.
3. Jenn Christensen and Elizabeth Cohen, CNN Health, May 4, 2016, http://edition.cnn.com/2016/05/03/health/medical-error-a-leading-cause-of-death/.
4. Ibid.

5. Gartner, Inc., "IT Glossary," www.gartner.com/it-glossary/big-data/ (accessed April 15, 2013).

6. Webopedia, "Big Data," www.webopedia.com/TERM/B/big_data.html (accessed April 15, 2013).

7. Personal Data: The Emergence of a New Asset Class, An Initiative of the World Economic Forum, January 2011, http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.

8. Deidra Pacnad, "Defensible Disposal: You Can't Keep All Your Data Forever," July 17, 2012, www.forbes.com/sites/ciocentral/2012/07/17/defensible-disposal-you-cant-keep-all-your-data-forever/.

9. Susan Karlin, "Earth's Nervous System: Looking at Humanity through Big Data," www.fastcocreate.com/1681986/earth-s-nervous-system-looking-at-humanity-through-big-data#1 (accessed March 5, 2013).

10. Steve Lohr, "How Big Data Became So Big," *New York Times,* August 11, 2012, www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?_r=2&smid=tw-share&.

11. Kahn Consulting, "Information Governance brief" sponsored by IBM, www.delve.us/downloads/Brief-Defensible-Disposal.pdf (accessed March 4, 2013).

12. Barclay T. Blair, "Girding for Battle," *Law Technology News,* October 1, 2012, www.law.com/jsp/lawtechnologynews/PubArticleLTN.jsp?id=1202572459732&thepage=1.

13. Ibid.

14. Deidra Pacnad, "Defensible Disposal: You Can't Keep All Your Data Forever," July 17, 2012, www.forbes.com/sites/ciocentral/2012/07/17/defensible-disposal-you-cant-keep-all-your-data-forever/.

15. Randolph A. Kahn, November 28, 2012, https://twitter.com/InfoParkingLot/status/273791612172259329.

16. "Gartner Says Master Data Management Is Critical to Achieving Effective Information Governance," January 19, 2012, www.gartner.com/newsroom/id/1898914.

17. Ibid.

18. Monica Crocker, e-mail to author, June 21, 2012.

19. The Economist Intelligence Unit, "The Future of Information Governance," www.emc.com/leadership/business-view/future-information-governance.htm (accessed March 10, 2012).

20. Arvind Krishna, "Three Steps to Trusting Your Data in 2011," CTO Edge, posted

March 9, 2011, www.ctoedge.com/content/three-steps-trusting-your-data-2011.

21. Laura DuBois and Vivian Tero, IDC White Paper, sponsored by EMC Corp., "Practical Information Governance: Balancing Cost, Risk, and Productivity," August 2010, https://docplayer.net/6122601-Emc-perspective-emc-sourceone-email-management.html.

22. Monica Crocker, e-mail to author, June 21, 2012.

23. Barclay T. Blair, "Making the Case for Information Governance: Ten Reasons IG Makes Sense," ViaLumina Ltd, 2010. Online at http://barclaytblair.com/making-the-case-for-ig-ebook/.

24. Barclay T. Blair, "8 Reasons Why Information Governance (IG) Makes Sense," posted June 29, 2009, http://aiim.typepad.com/aiim_blog/2009/06/8-reasons-why-information-governance-ig-makes-sense.html.

25. Bob Herman, "Details of Anthem's Massive Cyberattack Remain in the Dark a Year Later," *Modern Healthcare,* March 30, 2016, www.modernhealthcare.com/article/20160330/NEWS/160339997.

26. Ibid.

27. https://en.wikipedia.org/wiki/Anthem_medical_data_breach.

28. Bob Herman, "Details of Anthem's Massive Cyberattack Remain in the Dark a Year Later," *Modern Healthcare,* March 30, 2016, www.modernhealthcare.com/article/20160330/NEWS/160339997.

29. Peter Abatan, "Corporate and Industrial Espionage to Rise in 2011," Enterprise Digital Rights Management, www.enterprisedrm.info/post/2742811887/corporate-espionage-to-rise-in-2011 (accessed March 9, 2012).

30. BBC News, "FBI Staff Disciplined for Sex Texts and Nude Pictures," February 22, 2013, www.bbc.co.uk/news/world-us-canada-21546135.

31. "Gartner Says Master Data Management Is Critical to Achieving Effective Information Governance," January 19, 2012, www.gartner.com/newsroom/id/1898914.

CHAPTER 2
Information Governance, IT Governance, Data Governance: What's the Difference?

There has been a great deal of confusion around the term *information governance* (IG), and how it is distinct from other similar industry terms such as *information technology (IT) governance* and *data governance*. Some books, articles, and blogs have compounded the confusion by offering a limited definition of IG, or sometimes offering a definition of IG that is just plain incorrect, often confusing it with data governance. Even so-called "experts" confuse the terms!

So in this chapter we will spell out the differences and include examples in hopes of clarifying what the meaning of each is, and how they are related.

All three terms are a subset of corporate governance, and in the above sequence, become increasingly broad in their approach. Data governance can be seen as part of IT governance, which is also a part of a broader program of information governance.

We will now delve into more detailed definitions and a comparison of the three.

Data Governance

Data governance expert Robert Seiner, author of the book *Non-Invasive Data Governance,* and also the editor of *The Data Administration Newsletter* for over 20 years, pioneered the concept of "non-invasive data governance." In his approach, Seiner focuses on what can get done toward improving data governance without major disruptions to the business or redesigning business processes. Bob offers his definition of data governance: "Data governance is the execution and enforcement of authority over the definition, production and usage of data."[1] He goes on to say, "My definition intentionally has some grit and some teeth —I fully stand behind having strong definition especially if it catches people's attention and opens the door for greater discussion. At the end of the day, true governance over data or information requires executed and enforced authority."

But his clients sometimes like to tone it down, softening the definition. Seiner notes, "Some of my clients ponder that the definition is too aggressive. These clients do not like the words 'execution and enforcement' so they tame it down to something less aggressive like 'formalized behavior for the management of data.' That is my definition of **data stewardship**."

> "Data governance is the execution and enforcement of authority over the definition, production and usage of data." —Robert Seiner

Data governance involves processes and controls to ensure that data at the most basic level—raw data that the organization is gathering and inputting—is true and accurate, and unique (not redundant). It involves **data cleansing (**or **data scrubbing**) to strip out corrupted, inaccurate, or extraneous data and **de-duplication,** to eliminate redundant occurrences of data. It also usually involves implementing Master Data Management (MDM, which is discussed in more detail in Chapter 10 on IG for IT).

Data governance focuses on data quality "from the ground up" at the lowest or root level, so that subsequent reports, analyses, and conclusions are based on clean, reliable, trusted data (or records) in database tables. Data governance is the most fundamental level at which to implement information governance. Data governance efforts seek to assure that formal management controls—systems, processes, and policies—are implemented to govern critical data assets to improve data quality and to avoid negative downstream effects of poor data. DG efforts also hold data stewards accountable for information quality and accuracy.

> Data governance uses techniques like data cleansing and de-duplication to improve data and reduce redundancies.

Data governance is a newer, hybrid quality control discipline that includes elements of data quality, data management, IG policy development, business process improvement (BPI), and compliance and risk management.

Data Governance Strategy Tips

Everyone in an organization wants good quality data to work with. But it isn't so easy to implement a data governance program. First of all, data is at such a low level that executives and board members are typically unaware of the details of the "smoky back room" of data collection, cleansing, normalization, and input. So it is difficult to gain an executive sponsor and funding to initiate the effort.[2] And if a data governance program does move forward, there are challenges in getting business users to adhere to new policies. This is a crucial point, since much of the data is being generated by business units. But there are some general guidelines that can help improve a data governance program's chances for success:

- *Identify a measurable impact:* You must be able to demonstrate the business value of a data governance program, or it will not get the executive sponsorship and funding it needs to move forward. A readiness assessment should capture the current state of data quality and whether or not an enterprise or business unit level effort is warranted. Other key issues include: Can the organization save hard costs by implementing data governance? Can it reach more customers, or increase revenue generated from existing customers?[3]
- *Assign accountability for data quality to business units, not IT:* Typically, IT has had responsibility for data quality, yet it is mostly not under their control, since most of the data is being generated out in the business units. So a pointed effort must be made to push responsibility and ownership for data to the business units that create and use the data.
- *Recognize the uniqueness of data as an asset:* Unlike other assets like people, factories, equipment, and even cash, data is largely unseen, out of sight, and intangible. It changes daily. It spreads throughout business units. It is copied and deleted. Data growth can spiral out of control, obscuring the data that has true business value. So data has to be treated differently and its unique qualities must be considered.
- *Forget the past, implement a "going forward" strategy:* It is a significantly greater task to try to improve data governance across the enterprise for existing data. Remember, you may be trying to fix decades of bad behavior, mismanagement, and lack of governance. Taking an "incremental approach with an eye to the future" provides for a clean starting point and can substantially reduce the pain required to implement. So a "from this point on" strategy where new data governance policies for handling data are implemented beginning on a certain date is a proven best practice.
- *Manage the change:* Educate, educate, educate. People must be trained to understand why the data governance program is being implemented and how it will benefit the business. The new policies represent a cultural change and supportive program messages and training are required to make the shift.[4]

Good data governance ensures that downstream negative effects of poor data are avoided and that subsequent reports, analyses and conclusions are based on reliable, trusted data.

IT Governance

IT governance is the primary way that stakeholders can ensure that investments in IT create business value and contribute toward meeting business objectives.[5] This strategic alignment of IT with the business is challenging, yet essential. IT governance programs go further and aim to "improve IT performance, deliver optimum business value and ensure regulatory compliance."[6]

Although the CIO typically has line responsibility for implementing IT governance, the CEO and board of directors must receive reports and updates to discharge their responsibilities for IT governance and to see that the program is functioning well and providing business benefits.

Typically, in past decades, board members did not get involved in overseeing IT governance. But today it is a

critical and unavoidable responsibility. According to the IT Governance Institute's Board Briefing on IT Governance, "IT governance is the responsibility of the board of directors and executive management. It is an integral part of enterprise governance and consists of the leadership and organizational structures and processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives."[7]

> IT governance seeks to align business objectives with IT strategy to deliver business value.

The focus is on the actual software development and maintenance activities of the IT department or function, and IT governance efforts focus on making IT efficient and effective. That means minimizing costs by following proven software development methodologies and best practices, principles of data governance and information quality, and project management best practices, while aligning IT efforts with the business objectives of the organization.

IT Governance Frameworks

There are several IT governance frameworks that can be used as a guide to implementing an IT governance program. (They are introduced below in a cursory, way, as a detailed discussion of them is best suited for other books focused solely on IT governance.)

Although frameworks and guidance like COBIT® and ITIL have been widely adopted, there is no absolute standard IT governance framework; the combination that works best for your organization depends on business factors, corporate culture, IT maturity, and staffing capability. The level of implementation of these frameworks will also vary by organization.

COBIT®

COBIT (Control Objectives for Information and [related] Technology) is a process-based IT governance framework that represents a consensus of experts worldwide. It was codeveloped by the IT Governance Institute and ISACA and first released in 1996, as a set of control objectives to assist auditors. COBIT 5 was released in 2012, and the current version is COBIT 2019.

COBIT is a high-level framework and de facto standard to guide software development efforts. It holds IT departments accountable for contributing to business objectives. COBIT has been harmonized with other standards and best practices contained in ITIL (IT Infrastructure Library), COSO (Committee of Sponsoring Organizations of the Treadway Commission), ISO 27001/2, PMBOK (Project Management Book of Knowledge), and other "accepted practices" in IT development and operations.[8] COBIT addresses business risks, control requirements, compliance, and technical issues.[9]

COBIT offers IT controls that:

- Cut IT risks while gaining business value from IT under an umbrella of a globally accepted framework.
- Assist in meeting regulatory compliance requirements.
- Utilize a structured approach for improved reporting and management decisionmaking
- Provide solutions to control assessments and project implementations to improve IT and information asset control.[10]

COBIT consists of detailed descriptions of processes required in IT and also tools to measure progress toward maturity of the IT governance program. It is industry agnostic and can be applied across all vertical industry sectors, and it continues to be revised and refined.[11]

COBIT is broken out into three basic organizational levels and their responsibilities: (1) Board of directors and executive management; (2) IT and business management; and, (3) line level governance, security, and control knowledge workers.[12]

The COBIT model draws upon the traditional "plan, build, run, monitor" paradigm of traditional IT management, only with variations in semantics. There are four IT domains in the COBIT framework, which contain 34 IT processes and 210 control objectives that map to the four IT processes: (1) plan and organize, (2) acquire and implement, (3) deliver and support, and, (4) monitor and evaluate. Specific goals and

metrics are assigned, and responsibilities and accountabilities are delineated.

Val IT®

Val IT is a newer value-oriented framework that is compatible with and complementary to COBIT. Its principles and best practices focus is on leveraging IT investments to gain maximum value. 40 key Val IT essential management practices (analogous to COBIT's control objectives) support three main processes: Value Governance, Portfolio Management, and Investment Management. Val IT and COBIT "provide a full framework and supporting tool set" to help managers develop policies to manage business risks and deliver business value while addressing technical issues and meeting control objectives in a structured, methodic way.[13]

> COBIT is process-oriented and has been widely adopted as an IT governance framework. Val IT is value-oriented and compatible with and complementary to COBIT, yet focuses on value delivery.

ITIL

ITIL (Information Technology Infrastructure Library) is a set of process-oriented best practices and guidance originally developed in the UK to standardize delivery of IT service management. ITIL is applicable to both the private and public sectors and is the "most widely accepted approach to IT service management in the world." [14] Again, as with other IT governance frameworks, ITIL provides essential guidance for delivering business value through IT, and it "provides guidance to organizations on how to use IT as a tool to facilitate business change, transformation, and growth."[15]

ITIL best practices form the foundation for ISO/IEC 20000 (previously BS15000), the International Service Management Standard for organizational certification and compliance.[16] ITIL 2011 is the latest revision (as of this printing), and it consists of five core published volumes that map the IT service cycle in a systematic way:

1. ITIL Service Strategy
2. ITIL Service Design
3. ITIL Service Transition
4. ITIL Service Operation
5. ITIL Continual Service Improvement[17]

> ITIL is the "most widely accepted approach to IT service management in the world." [18]

ISO 38500

ISO/IEC 38500:2015 is an international standard that provides high-level principles and guidance for senior executives and directors, and those advising them, for the effective and efficient use of IT.[19] Based primarily on AS 8015, the Australian IT governance standard, it "applies to the governance of management processes" that are performed at the IT service level, but the guidance assists executives in monitoring IT and ethically discharging their duties with respect to legal and regulatory compliance of IT activities.

The ISO 38500 standard comprises three main sections:

1. Scope, Application and Objectives
2. Framework for Good Corporate Governance of IT
3. Guidance for Corporate Governance of IT

It is largely derived from AS 8015, the guiding principles of which were to:

- Establish responsibilities
- Plan to best support the organization
- Acquire validly
- Ensure performance when required
- Ensure conformance with rules
- Ensure respect for human factors

The standard also has relationships with other major ISO standards, and embraces the same methods and approaches. It is certain to have a major impact upon the IT governance landscape.[20]

ISO 38500 is an international standard that provides high level principles and guidance for senior executives and directors responsible for IT governance.

Information Governance

Corporate governance is the highest level of governance in an organization and a key aspect of it is information governance (IG). According to the Sedona Conference, IG programs are about *minimizing information risks and costs and maximizing information value*.[21] This is a compact way to convey the key aims of IG programs, and it is what should be emphasized when the merits of an IG program are discussed. The definition of IG can be distilled further to a more succinct "elevator pitch" definition of IG, which is "security, control, and optimization" of information. (See Chapter 1 for more detailed definitions.)

IG processes are higher level than the details of IT governance, and much higher level than data governance, but both of the aforementioned can be (and should be) a part of an overall IG program. In fact, often IG programs are launched from successful (and funded) data governance programs.

IG programs are driven from the top down but implemented from the bottom up.

The IG approach to governance focuses not on detailed IT or data capture and quality processes, but rather on controlling the information that is generated by IT, office systems, and external systems, that is, the *output of IT*. IG efforts seek to manage and control information assets to lower risk, ensure compliance with regulations, and to improve information quality and accessibility while implementing information security measures to protect and preserve information that has business value.[22]

IG programs focus on breaking down traditional functional group "siloed" approaches to maximize the value of information. Mature IG programs employ the principles of infonomics to measure and monetize information. But these programs rely on robust, effective data governance programs to provide good, clean data so that calculations and analytics that are applied yield true and accurate results.

Information governance is how an organization maintains security, complies with regulations and laws, and meets ethical standards when managing information.

Impact of a Successful IG Program

When making the business case for IG, and articulating its benefits, it is useful to focus on its central impact. If there is a business case to apply infonomics and gain new value from information, the benefits may be quite clear in terms of monetizing information, or leveraging it in a barter transaction. However, *putting*

cost-benefit numbers to IG programs often is difficult, unless you also consider the worst-case scenario of loss or misuse of corporate or agency records. What is losing the next big lawsuit worth? How much are confidential merger and acquisition (M&A) documents worth? How much are customer records worth? How much could a GDPR or HIPAA fine be, and what is the risk?

Frequently, executives and managers do not understand the value of IG until it is a crisis, an expensive legal battle is lost, heavy fines are imposed for noncompliance, or executives go to jail.

There are some key outputs from implementing an IG program. A successful IG program should enable organizations to:

- *Improve collaboration between business units*. IG programs require cross-functional collaboration, so that IG team leaders can foster an environment of information sharing and leverage across the entire enterprise.
- *Use common terms across the enterprise*. This means that departments must agree on how they are going to classify document types, which relies on a cross-functional effort. With common enterprise terms, searches for information are more productive and complete. This begins with developing a standardized corporate taxonomy, which defines the terms (and substitute terms in a custom corporate thesaurus), document types, and their relationships in a hierarchy.
- *Map information creation and usage*. This effort can be buttressed with the use of technology tools such as data mapping and **data loss prevention (DLP)**, which can also be used to discover the flow of information within and outside of the enterprise. You must first determine *who* is accessing *which* information *when,* and *where* it is going. Then these information flows can be monitored and analyzed. The goal is to stop the erosion or misuse of information assets, and to stem data breaches with monitoring and security technology.
- *Comply with data protection regulations*. Once a data map is created, organizations are better able to govern data and to comply with requirements like **digital subject access requests (dSAR)** under GDPR.
- *Obtain "information confidence."* That is, the assurance that information has integrity, validity, accuracy, and quality; this means being able to *prove* that the information is reliable, and its access, use, and storage meets compliance and legal demands.
- *Harvest and leverage information*. Using techniques and tools like business intelligence and advanced analytics (descriptive, diagnostic, predictive, prescriptive), new insights may be gained that provide an enterprise with a sustainable competitive advantage over the long term, since managers will have more and better information as a basis for business decisions.[23]
- *Monetize information*. Applying infonomics principles and specific formulas can allow an organization to find real, tangible value in their information, which had not been capitalized upon before.

Summing Up the Differences

IG consists of the overarching polices and processes to optimize and leverage information, while controlling its access, keeping it secure, and meeting legal and privacy obligations, in alignment with stated organizational business objectives.

IT Governance consists of following established frameworks and best practices to gain the most leverage and benefit out of IT investments and support accomplishment of business objectives.

Data governance is the execution and enforcement of authority over the definition, production, and usage of data,[24] and consists of the processes, methods, and techniques to ensure that data at the root level is of high quality, reliable, and unique (not duplicated), so that downstream uses in reports and databases are more trusted and accurate.

CHAPTER SUMMARY: KEY POINTS

- IG, IT governance, and data governance are all a subset of corporate governance.
- Data governance is the "execution and enforcement of authority over the definition, production and usage of data," according to expert Bob Seiner.
- Data governance uses techniques like data cleansing and de-duplication to improve data and reduce redundancies.
- Good data governance ensures that downstream negative effects of poor data are avoided and that subsequent reports, analyses, and conclusions are based on reliable, trusted data.
- IT governance seeks to align business objectives with IT strategy to deliver business value.
- COBIT2019 is process-oriented and has been widely adopted as an IT governance framework. Val IT is value-oriented and compatible and complementary with COBIT yet focuses on value delivery.
- ITIL is the "most widely accepted approach to IT service management in the world."[25]
- ISO 38500 is an international standard that provides high-level principles and guidance for senior executives and directors responsible for IT governance.
- Information governance is how an organization maintains security, complies with regulations and laws, and meets ethical standards when managing information.
- According to the Sedona Conference, IG is about minimizing information risks and costs and maximizing information value.
- IG, in short, is "security, control, and optimization of information."
- IG programs allow organizations to improve collaboration between business units; use common terms across the enterprise; map information creation and usage; comply with data protection regulations; obtain information confidence; harvest and leverage information; and monetize information, using Infonomics principles and techniques.

Notes

1. "Talking Data Governance with Thought Leader Bob Seiner," *Information Governance World*, Fall issue, 2018, 58.
2. "New Trends and Best Practices for Data Governance Success," SearchDataManagement.com eBook, http://viewer.media.bitpipe.com/1216309501_94/1288990195_946/Talend_sDM_SO_32247_EBook_1104.pdf (accessed March 11, 2013).
3. Ibid.
4. Ibid.
5. M. N. Kooper, R. Maes, and E. E. O. Roos Lindgreen, "On the Governance of Information: Introducing a New Concept of Governance to Support the Management of Information," *International Journal of Information Management* 31 (2011), 195–20, www.scribd.com/doc/63866742/On-the-Governance-of-Information.
6. Nick Robinson, "The Many Faces of IT Governance: Crafting an IT Governance Architecture," *ISACA Journal* 1 (2007), www.isaca.org/Journal/Past-Issues/2007/Volume-1/Pages/The-Many-Faces-of-IT-Governance-Crafting-an-IT-Governance-Architecture.aspx.
7. Bryn Phillips, IT Governance for CEOs and Members of the Board (CreateSpace Independent Publishing Platform, 2012), 18.
8. Ibid.
9. Ibid., 26.

10. "Control Objectives for Information and Related Technology (COBIT®) Internationally Accepted Gold Standard for IT Controls & Governance," IBM Global Business Services—Public Sector, www-304.ibm.com/industries/publicsector/fileserve?contentid=187551 (accessed March 11, 2013).

11. Bryn Phillips, 26.

12. "Control Objectives for Information and Related Technology (COBIT®) Internationally Accepted Gold Standard for IT Controls & Governance," IBM Global Business Services—Public Sector, http://www-304.ibm.com/industries/publicsector/fileserve?contentid=187551 (accessed March 11, 2013).

13. Ibid.

14. www.itil-officialsite.com/ (accessed March 12, 2013).

15. www.itil-officialsite.com/AboutITIL/WhatisITIL.aspx (accessed March 12, 2013).

16. Ibid.

17. Ibid.

18. www.itil-officialsite.com/ (accessed March 12, 2013).

19. www.iso.org/iso/catalogue_detail?csnumber=51639 (accessed March 12, 2013).

20. http://www.38500.org/ (accessed March 12, 2013).

21. https://thesedonaconference.org/publications (accessed October 11, 2017).

22. www.naa.gov.au/records-management/agency/digital/digital-continuity/plan/information-governance.aspx (accessed March 12, 2013).

23. Arvind Krishna, "Three Steps to Trusting Your Data in 2011," CTO Edge, March 9, 2011, www.ctoedge.com/content/three-steps-trusting-your-data-2011.

24. Bob Seiner, e-mail to author, July 24, 2018.

25. www.itil-officialsite.com/ (accessed March 12, 2013).

CHAPTER 3
Information Governance Principles

Using **guiding principles** to drive your information governance (IG) program can help educate stakeholders, focus efforts, and maintain consistency.

The Sedona Conference® Commentary on Information Governance

The Sedona Conference is a group of mostly legal and technology professionals that meets periodically and develops commentary and guidance on e-discovery, electronic records, privacy, risk, IG, and related issues. They have developed 11 general principles of IG,[1] which provide guidance on the expectations and aims of IG programs. These principles can further an IG team's understanding of IG and can be used in an introductory "IG Awareness Training" session in the early stages of your program launch. A good exercise is to have team members rewrite these principles in their own words, and then hold discussions about how each of these principles would apply to their departmental IG efforts, and the overall IG program. The Sedona Conference Commentary on IG, formed as principles, are:

1. Organizations should consider implementing an IG program to make coordinated decisions about information for the benefit of the overall organization that address information-related requirements and manage risks while optimizing value.
2. An IG program should maintain sufficient independence from any particular department or division to ensure that decisions are made for the benefit of the overall organization.
3. All information stakeholders should participate in the IG program.
4. The strategic objectives of the IG program should be based upon a comprehensive assessment of information-related practices, requirements, risks, and opportunities.
5. An IG program should be established with the structure, direction, resources, and accountability to provide reasonable assurance that the program's objectives will be achieved.
6. The effective, timely, and consistent disposal of physical and electronic information that no longer needs to be retained should be a core component of any IG program.
7. When IG decisions require an organization to reconcile conflicting laws or obligations, the organization should act in good faith and give due respect to considerations such as privacy, data protection, security, records and information management, risk management, and sound business practices.
8. If an organization has acted in good faith in its attempt to reconcile conflicting laws and obligations, a court or other authority reviewing the organization's actions should do so under a standard of reasonableness according to the circumstances at the time such actions were taken.
9. An organization should consider reasonable measures to maintain the integrity and availability of long-term information assets throughout their intended useful life.
10. An organization should consider leveraging the power of new technologies in its IG program.
11. An organization should periodically review and update its IG program to ensure that it continues to meet the organization's needs as they evolve.

Smallwood IG Principles

The following 11 IG principles are the result of the author's research and consulting efforts over the past decade or so, where a great deal of practical information on IG program successes, failures, and best practices was synthesized, analyzed, and distilled.

These 11 IG principles must be adhered to as general guidelines for IG programs to succeed:

1. *Value information as an asset.* Just as any organization has physical assets like buildings, furniture, fixtures, equipment, computers, software, and vehicles, and so on, that have value, information collected and analyzed also has value. The formal management of information assets with the goal of monetizing and leveraging that information is clearly outlined in the book *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage* (Bibliomotion/Taylor & Francis, 2018), written by Doug Laney. It is necessary to identify and map out information assets so that confidential information including personally identifiable information (PII), credit card information (PCI), and protected health information (PHI), may be secured directly, so that if hackers are able to get inside the organization's firewall, this information is encrypted and unreadable. The IG steering committee must also explore which analytic tools could help to maximize information value, which may come in the form of reducing uninformed or poor decisions, improving customer satisfaction, improving operational efficiency, reducing legal costs, improving compliance capabilities, and other related benefits. In addition, clear policies must be established for the secure access and use of information, and those policies must be communicated regularly and crisply to employees, with constant reinforcement. This includes conveying the value and risk of information and the consequences of violating IG policies.
2. *Stakeholder consultation.* IG programs are, by nature, cross-functional efforts. Those who work most closely to information are the ones who best know why it is needed and how to manage it, so business units must be consulted in IG policy development. Records and information management (RIM) professionals know the details and nuances of managing records. They should be deeply involved in electronic records management (ERM) and document management governance efforts, which should be at

the core of IG. Effective ERM governance also leads to reduced costs and improved operational efficiency. Therefore, RIM professionals must be an active part of IG programs and their input into the policymaking process is critical. RIM professionals must also work hand-in-hand with privacy and legal professionals to ensure customer (and employee) privacy is protected vigilantly. Privacy has become even more important globally with the implementation of Europe's General Data Protection Regulation (GDPR), which applies to any organization conducting business with European citizens, regardless of location. In-house legal council should be a key player in IG programs and the legal team must be consulted on a variety of legal, regulatory, privacy, and litigation issues as IG program efforts involve all these areas. Further, IG programs can cut electronic discovery collection and review costs and make the legal hold notification (LHN) process more streamlined and effective. The IT department must play a major role as technology is leveraged in IG program efforts. The increased level and sophistication of cybersecurity attacks underscores the need for a robust cybersecurity program, including security awareness training (SAT) to offset information risks. It is clear that cross-functional stakeholder consultation is a necessary component of IG programs.

3. *Information integrity.* The business-to-customer relationship is based on trust, and that trust includes ensuring that accurate customer information is created and also kept secure. IG programs focus heavily on information quality, from the ground up, beginning with data governance. Data governance techniques and tools focus on creating clean, accurate, nonduplicate data in database tables so that downstream reports and analyses are more trusted and accurate. Information integrity considers the consistency of methods used to create, retain, preserve, distribute, and track information. Information integrity means there is the assurance that information is accurate, correct, and authentic. From a legal standpoint, enabling information technologies and data stewardship polices must support the effort to meet legal standards of admissibility and preserve the integrity of information to guard against claims that it has been altered, tampered with, or deleted (called "spoliation"). Audit trails must be kept and monitored to ensure compliance with IG policies to assure information integrity.
4. *Information organization and classification.* This means that not only must customer and business operations records be organized in a standardized taxonomy with a specified metadata approach, but that all information must be organized in a standardized way, categorizing all information and semantically linking it to related information. It also means creating a **records retention schedule** (RRS) that spells out how long the PII/PCI/PHI as well as business information (e.g., e-mail, e-documents, spreadsheets, reports) should be retained and how it is to be disposed of or archived (disposition). Further, it means developing departmental file plans that are logical and help end users to conduct more complete and accurate searches for information.
5. *Information security and privacy.* This again focuses on the trust proposition between customer and the business or governmental agency. Information security must be in place before information privacy can be assured. This means that every attempt must be made to secure PII, PCI, and PHI in all three states: at rest, in motion, and in use. It means that the organization should conduct regular security awareness training, which can include staged phishing and spear phishing attacks to see if employees handle them properly, and to coach them on mistakes they may make during the test. Ransomware is also a problem. When rogue players launch ransomware attacks, they typically encrypt the storage drives of the organization and demand a modest payment by Bitcoin. To offset this risk a complete backup of your entire information system must be made daily and kept separate from your network, offline. Additional cybersecurity hygiene measures are needed to protect information from damage, theft, or alteration by malicious outsiders and insiders as well as nonmalicious (accidental) actions that may compromise information. For instance, an employee may lose a laptop with confidential information, but if proper IG policies are enforced using security-related information technologies, the information can be kept secure. This can be done by **access control** methods, data or document encryption, deploying **information rights management** (IRM) software, using remote digital shredding capabilities, and implementing enhanced auditing procedures. Information **privacy awareness training** (PAT) should also be conducted, including updates on federal, provincial, state, and even possibly municipal legal requirements. Information privacy is closely related to information security and is critical when dealing with confidential information and other sensitive information such as race or religion.
6. *Information accessibility.* Information accessibility must be balanced with information security concerns. Information accessibility includes making the information as simple as possible to locate and access, which involves not only an intuitive user interface but also utilizing enterprise search principles, technologies, and tools. It further includes basic access controls, such as password management, identity and access management (IAM), and delivering information to a variety of hardware devices. Accessibility to information is essential not only in the short term but also over time. Maintaining records for perhaps decades requires consideration of **long-term digital preservation** (LTDP) planning, tools, and methods in accordance with international, technology-neutral standards. Today, LTDP capabilities can be provided through cloud services providers that keep a number of copies of the information (typically five to six) on Amazon or Microsoft cloud servers, spread around the world, to reduce the risk of loss. There are privacy implications to this global approach, especially with GDPR legislation, and they must be researched.
7. *Information control.* An enterprise RRS is a key foundational element of IG programs. Non-record information must also be categorized and scheduled. Then a standardized, automated LHN process must be put in place to assign data stewards and lock down information that may be requested in legal proceedings. In addition, key information control technologies must be deployed to control the access, creation, updating, and printing of data, documents, and reports. These technologies include several types of software: enterprise content management (ECM) and enterprise file synch and share (EFSS), document management, document analytics, report management, and workflow. Additional security software including encryption should be deployed to protect confidential or sensitive information.
8. *Information governance monitoring and auditing.* Early on in the development of an IG program a concerted effort must be made to develop metrics to objectively measure program progress and employee conformance with IG polices. To ensure that guidelines and policies are being followed, especially regarding customer privacy and cybersecurity hygiene, information access and use must be monitored. To guard against claims of legal spoliation, the use of e-mail, social media, cloud computing, and report generation should be logged (recorded or archived) in real time and maintained as an audit record. Technology tools such as document analytics can track how many documents users access and print and how long they spend doing so.

9. *Executive sponsorship*. Once again, due to the cross-functional, collaborative nature of IG programs, this is the most crucial factor in IG program success. This is especially true in the healthcare arena, where various clinical specialties that may have their own proprietary information systems are represented. No IG effort will survive and be successful if it does not have an accountable, responsible executive sponsor. The sponsor must develop the business case for IG early on, establish a budget, then assemble the steering committee and drive the effort. The executive sponsor must pay periodic attention to the IG program, monitoring progress based on metrics and milestones. The IG program lead, or perhaps even a chief IG officer, manages the IG program on a day-to-day basis, bringing in the executive sponsor only when support is needed for a particular issue. The executive sponsor must clear obstacles for the IG program lead and IG steering committee while actively communicating the goals and business objectives that the IG program addresses, and at the same time keeping upper management informed on progress, particularly when accomplishing milestones.
10. *Change management*. IG programs require leveraging information as an asset and breaking down functional siloed approaches to managing information. IG programs also often involve changing the way users interact with systems to streamline search and retrieval and improve productivity. These changes must be "sold" to the stakeholders, with an emphasis on how the IG program efforts will help the organization achieve its business objectives. These changes must be clearly communicated and reinforced to employees and change must occur at the core of the organization or target department. This all requires that a purposeful change management initiative accompany the IG program implementation.
11. *Continuous improvement*. IG programs are not one-time projects but rather ongoing programs, akin to a workplace safety program. (In fact, the information security aspects of an IG program could actually be termed "information safety.") The IG program is a major change management effort, which requires a major training and communications effort. Progress in the IG program must be reviewed periodically and adjusted to account for gaps or shortcomings as well as changes in the business environment, technology usage, or business strategy.

Using these 11 principles as guidelines will help to communicate with stakeholders and IG steering committee what IG is, why it is needed, what it involves, and how to fashion an IG program that is successful. It is essential to continually reinforce the importance of these principles during the course of an IG program, and measure how well the organization is doing in these 11 critical areas.

There are also other sets of principles that apply to IG efforts and can help provide a more complete understanding of IG programs, especially early in the IG program development process. These IG principles reflect, reinforce, and expand on the previous set.

> Principles of successful IG programs are emerging. They include executive sponsorship, information classification, integrity, security and privacy, accessibility, control, monitoring, and auditing; also policy development and continuous improvement.

Accountability Is Key

According to Debra Logan at Gartner Group, *none of the proffered definitions of IG include "any notion of coercion, but rather ties governance to accountability* [italics added] that is designed to encourage the right behavior.... The word that matters most is *accountability*" [italics in the original]. The root of many problems with managing information is the "fact that there is no accountability for information as such."[2]

Establishing policies, procedures, processes, and controls to ensure the quality, integrity, accuracy, and security of business records are the fundamental steps needed to reduce the organization's risk and cost structure for managing these records. Then, it is essential that IG efforts are supported by information technologies (IT). The auditing, testing, maintenance, and improvement of IG is enhanced by using electronic records management (ERM) software, along with other complementary technology sets such as workflow and **business process management system (BPMS)** software and digital signatures.

> Accountability is a key aspect of IG.

Generally Accepted Recordkeeping Principles®

Contributed by Charmaine Brooks

A major part of an IG program is managing formal business records. Although they account for only about 7–9% of the total information that an organization holds, they are the most critically important subset to manage, as there are serious compliance and legal ramifications.

Records and recordkeeping are inextricably linked with any organized business activity. Through the information that an organization uses and records, creates, or receives in the normal course of business, it knows what has been done and by whom. This allows the organization to effectively demonstrate compliance with applicable standards, laws, and regulations, as well as plan what it will do in the future to meet its mission and strategic objectives.

Standards and principles of recordkeeping have been developed by **records and information management** (RIM) practitioners to establish benchmarks for how organizations of all types and sizes can build and sustain compliant, defensible **records management** (RM) programs.

The Principles

In 2009 ARMA International published a set of eight **Generally Accepted Recordkeeping Principles®**, known as "The

Principles"[3] (or sometimes "GAR Principles") to foster awareness of good recordkeeping practices. These principles and associated metrics provide an **information governance** (IG) framework that can support continuous improvement.

The eight Generally Accepted Recordkeeping Principles are:

1. *Accountability*. A senior executive (or person of comparable authority) oversees the recordkeeping program and delegates program responsibility to appropriate individuals. The organization adopts policies and procedures to guide personnel, and ensure the program can be audited.
2. *Transparency*. The processes and activities of an organization's recordkeeping program are documented in a manner that is open and verifiable and is available to all personnel and appropriate interested parties.
3. *Integrity*. A recordkeeping program shall be constructed so the records and information generated or managed by or for the organization have a reasonable and suitable guarantee of authenticity and reliability.
4. *Protection*. A recordkeeping program shall be constructed to ensure a reasonable level of protection to records and information that are private, confidential, privileged, secret, or essential to business continuity.
5. *Compliance*. The recordkeeping program shall be constructed to comply with applicable laws and other binding authorities, as well as the organization's policies.
6. *Availability*. An organization shall maintain records in a manner that ensures timely, efficient, and accurate retrieval of needed information.
7. *Retention*. An organization shall maintain its records and information for an appropriate time, taking into account legal, regulatory, fiscal, operational, and historical requirements.
8. *Disposition*. An organization shall provide secure and appropriate disposition for records that are no longer required to be maintained by applicable laws and the organization's policies."[4]

The Generally Accepted Recordkeeping Principles consists of eight principles that provide an IG framework that can support continuous improvement.

The Generally Accepted Recordkeeping Principles apply to all sizes of organizations, in all types of industries, and in both the private and public sectors, and can be used to establish consistent practices across business units. The Principles are an IG maturity model and this is used as a preliminary evaluation of recordkeeping programs and practices.

Interest and the application of The Principles for assessing an organization's recordkeeping practices have steadily increased since its establishment. It is an accountability framework that includes the processes, roles, standards, and metrics that ensure the effective and efficient use of records and information in support of an organization's goals and business objectives.

As shown in Table 3.1, the Generally Accepted Recordkeeping Principles maturity model associates characteristics that are typical in five levels of recordkeeping capabilities that range from 1 (substandard) to 5 (transformational). The levels are both descriptive and color coded for ease of understanding. The eight principles and levels (metrics) are applied to the current state of an organization's recordkeeping capabilities and can be cross-referenced to the policies and procedures. *While it is not unusual for an organization to be at differing levels of maturity in the eight principles, the question "How good is good enough?" must be raised and answered*; a rating of less than "transformational" may be acceptable, depending on the organization's tolerance for risk and an analysis of the costs and benefits of moving up each level.

Table 3.1 Generally Accepted Recordkeeping Principles Levels

Source: Based on data from ARMA.

Level 1 Substandard	**Characterized by an environment where recordkeeping concerns are either not addressed at all or are addressed in an ad hoc manner.**
Level 2 In Development	Characterized by an environment where there is a developing recognition that recordkeeping has an impact on the organization, and the organization may benefit from a more defined information governance program.
	Characterized by an environment where defined policies and procedures exist that address the minimum or essential legal and regulatory requirements, but more specific actions need to be taken to improve recordkeeping.
	Characterized by an environment where information governance issues and considerations are integrated into business decisions on a routine basis, and the organization consistently meets its legal and regulatory obligations.
	Characterized by an environment that has integrated information governance into its corporate infrastructure and business processes to such an extent that compliance with program requirements is routine.

The Generally Accepted Recordkeeping Principles maturity model measures recordkeeping maturity in five levels.

The maturity levels define the characteristics of evolving and maturing records management programs. The assessment should reflect the current RM environment and practices. The principles and maturity level definitions, along with improvement recommendations (roadmap), outline the tasks required to proactively approach addressing systematic records management

practices and reach the next level of maturity for each principle. While the Generally Accepted Recordkeeping Principles are broad in focus, they illustrate the requirements of good records management practices. The Principles Assessment can also be a powerful communication tool to promote cross-functional dialogue and collaboration among business units and staff.

Accountability

The principle of **accountability** covers the assigned responsibility for RM at a senior level to ensure effective governance with the appropriate level of authority. A senior-level executive must be high enough in the organizational structure to have sufficient authority to operate the records management program effectively. The primary role of the senior executive is to develop and implement records management policies, procedures and guidance, and to provide advice on all record-keeping issues. The direct responsibility for managing or operating facilities or services may be delegated.

The senior executive must possess an understanding of the business and legislative environment within which the organization operates; business functions and activities; and the required relationships with key external stakeholders. This person must also understand how records management contributes to achieving the corporate mission, aims, and objectives.

It is important for top-level executives to take ownership of the records management issues of the organization identifying corrective actions required for mitigation or ensuring resolution of problems and recordkeeping challenges. An executive sponsor should identify opportunities to raise awareness of the relevance and importance of RM and effectively communicate the benefits of good records management to staff and management.

The regulatory and legal framework for records management must be clearly identified and understood. The senior executive must have a sound knowledge of the organization's information and technological architecture and actively participate in strategic decisions for information technology systems acquisition and implementation.

The senior executive is responsible for ensuring the processes, procedures, governance structures, and related documentation are developed. The policies should identify the roles and responsibilities at all levels of the organization.

An audit process must be developed to cover all aspects of RM within the organization, including substantiating that sufficient levels of accountability have been assigned and accountability deficiencies are identified and remedied. Audit processes should include compliance with the organization policies and procedures for all records, regardless of format or media. Accountability audit requirements for electronic records include employing appropriate technology to audit the information architecture and systems. Accountability structures must be updated and maintained as changes occur in the technology infrastructure.

> An audit process must be developed to cover all aspects of RM in the organization.

The audit process must reinforce compliance and hold individuals accountable. The results should be constructive, encourage continuous improvement, but not be used as a means of punishment. *The audit should contribute to records program improvements in risk mitigation, control, and governance issues, and have the capacity to support sustainability*.

Transparency

Policies are broad guidelines for the operation of the organization and provide a basic guide to action that prescribes the boundaries within which business activities are to take place. They state the course of action to be followed by the organization, business unit, department, and employees.

Transparency of recordkeeping practices includes documenting processes and promoting an understanding of the roles and responsibilities of all stakeholders. *To be effective, policies must be formalized and integrated into business processes.* Business rules and recordkeeping requirements need to be communicated and socialized at all levels of the organization.

Senior management must recognize that transparency is fundamental to IG and compliance. Documentation must be consistent, current, and complete. A review and approval process must be established to ensure the introduction of new programs or changes can be implemented and integrated into business processes.

> To be effective, policies must be formalized and integrated into business processes.

Employees must have ready access to RM policies and procedures. They must receive guidance and training to ensure they understand their roles and requirements for records management. Recordkeeping systems and business processes must be designed and developed to clearly define the records lifecycle.

In addition to policies and procedures, the development of guidelines and operational instructions, diagrams and flowcharts, system documentation, and user manuals must include clear guidance on how records are to be created, retained, stored, and dispositioned. The documentation must be readily available and incorporated in communications and training provided to staff.

Integrity

Record-generating systems and repositories must be assessed to determine recordkeeping capabilities. *A formalized process must be in place for acquiring or developing new systems, including requirements for capturing the metadata required for lifecycle management of records in the systems*. In addition, the record must contain all the necessary elements of an official record, including structure, content, and context. **Records integrity**, reliability, and trustworthiness are confirmed by ensuring that a record was created by a competent authority according to established processes.

Maintaining the integrity of records means that they are complete and protected from being altered. The authenticity of a record

is ascertained from internal and external evidence, including the characteristics, structure, content, and context of the record to verify they are genuine and not corrupted or altered. In order to trust that a record is authentic, organizations must ensure that recordkeeping systems that create, **capture**, and manage electronic records are capable of protecting records from accidental or unauthorized alteration or deletion while the record has value.

Protection

*Organizations must ensure the **protection** of records and ensure they are unaltered through loss, tampering, or corruption.* This includes technological change or the failure of digital storage media and protecting records against damage or deterioration.

This principle applies equally to physical and electronic records, each having unique requirements and challenges.

Access and security controls need to be established, implemented, monitored, and reviewed to ensure business continuity and minimize business risk. Restrictions on access and disclosure include the methods for protecting personal privacy and proprietary information. Access and security requirements must be integrated into the business systems and processes for the creation, use, and storage of records.

Long-term digital preservation (LTDP) is a series of managed activities required to ensure continued access to digital materials for as long as necessary. Electronic records requiring long-term retention may require conversion to a medium and format suitable to ensure long-term access and readability. Cloud-based services for file conversion and long-term storage have emerged that have simplified the LTDP process and made it more affordable for organizations.

Compliance

Records management programs include the development and training of the fundamental components, including **compliance monitoring** to ensure sustainability of the program.

Monitoring for compliance involves reviewing and inspecting the various facets of records management, including ensuring records are being properly created and captured, implementation of user permissions and security procedures, work flow processes through sampling to ensure adherence to policies and procedures, ensuring records are being retained following disposal authorities, and documentation of records destroyed or transferred to determine whether **destruction**/transfer was authorized in accordance with disposal instructions.

Compliance monitoring can be carried out by an internal audit, external organization, or records management and must be done on a regular basis.

Availability

Organizations should evaluate how effectively and efficiently records and information are stored and retrieved using present equipment, networks, and software. The evaluation should identify current and future requirements and recommend new systems as appropriate. Certain factors should be considered before upgrading or implementing new systems. These factors are practicality, cost, and effectiveness of new configurations.

A major challenge for organizations is ensuring that timely and reliable access to and use of information and records are accessible and usable for the entire length of the retention period. Rapid changes and enhancements to both hardware and software compound this challenge.

Retention

***Retention** is the function of preserving and maintaining records for continuing use.* The **records retention schedule** identifies the actions needed to fulfill the requirements for the retention and disposal of records and provides the authority for employees and systems to retain, destroy, or transfer records. The records retention schedule documents the recordkeeping requirements and procedures, identifying how records are to be organized and maintained, what needs to happen to records and when, who is responsible for doing what, and who to contact with questions or guidance.

Organizations must identify the scope of their recordkeeping requirements for documenting business activities based on regulated activities and jurisdictions that impose control over records. This includes business activities regulated by the government for every location or jurisdiction in which you do business. Other considerations for determining retention requirements include operational, legal, fiscal, and historical.

Records appraisal is the process of assessing the value and risk of records to determine their retention and disposition requirements. Legal research is outlined in appraisal reports. This may be accomplished as a part of the process of developing the records retention schedules, as well as conducting a regular review to ensure that citations and requirements are current.

*The **records retention period** is the length of time that records should be retained and the actions taken for them to be destroyed or preserved.* The retention periods for different records should be based on legislative or regulatory requirements as well as on administrative and operational requirements. It is important to document the legal research conducted and used to determine whether the law or regulation has been reasonably applied to the recordkeeping practices and provide evidence to regulatory officials or courts that due diligence has been conducted in good faith to comply with all applicable requirements.

Disposition

***Disposition** is the last stage in the **information life cycle**.* When the retention requirements have been met and the records no longer serve a useful business purpose, they may be destroyed. Records requiring long-term or permanent retention should be transferred to an **archive** for **preservation**. The timing of the **transfer** of physical or electronic records should be determined through the records retention schedule process. Additional methods are often required to preserve electronic records, which may include migration or conversion.

Disposition is the last stage in the life cycle of records. Disposition is not synonymous with destruction, though destruction may be one disposal option.

Records must be destroyed in a controlled and secure manner and in accordance with authorized disposal instructions. The **destruction** of records must be clearly documented to provide evidence of destruction according to an agreed-on program.

Destruction of records must be undertaken by methods appropriate to the confidentiality of the records and in accordance with disposal instructions in the records retention schedule. An audit trail documenting the destruction of records should be maintained and **certificates** of destruction obtained for destruction undertaken by third parties. In the event disposal schedules are not in place, the written authorization should be obtained prior to destruction. Procedures should specify who must supervise the destruction of records. Approved methods of destruction must be specified for each media type to ensure that information cannot be reconstructed.

Disposition is not synonymous with destruction, though destruction may be one disposal option. Destruction of records must be carried out under controlled, confidential conditions by shredding or permanent disposition. This includes the destruction of confidential microfilm, microfiche, computer cassettes, and computer tapes, as well as paper.

Methods of Disposition

- *Discard*. The standard destruction method for nonconfidential records. If possible, all records should be shredded prior to recycling. Note that transitory records can also be shredded.
- *Shred*. Confidential and sensitive records should be processed under strict security. This may be accomplished internally or by secure on-site shredding by a third-party vendor who provides certificates of secure destruction. The shredded material is then recycled.
- *Archive*. This designation is for records requiring long-term or permanent preservation. Records of enduring legal, fiscal, administrative, or historical value are retained.
- *Imaging*. Physical records converted to digital images, after which the original paper documents are destroyed.
- *Purge*. This special designation is for data, documents, or records sets that need to be purged by removing material based on specified criteria. This often applies to structure records in databases and applications.

Assessment and Improvement Roadmap

The Generally Accepted Recordkeeping Principles maturity model can be leveraged to develop a current state assessment of an organization's recordkeeping practices and resources, identify gaps and assess risks, and develop priorities for desired improvements.

The Principles were developed by ARMA International to identify characteristics of an effective recordkeeping program. Each of the eight principles identifies issues and practices that, when evaluated against the unique needs and circumstances of an organization, can be applied to improvements for a recordkeeping program that meets recordkeeping requirements. The principles identify requirements and can be used to guide the incremental improvement in the management and governance of the creation, organization, security, maintenance, and other activities over a one- to five-year period. Fundamentally, records management and information governance are business disciplines that must be tightly integrated with operational policies, procedures, and infrastructure.

The Principles can be mapped to the four improvement areas in Table 3.2.

As an accepted industry guidance maturity model, The Principles provide a convenient and complete framework for assessing the current state of an organization's recordkeeping and developing a roadmap to identify improvements that will bring the organization into compliance. An assessment/analysis of the current record management practices, procedures, and capabilities together with current and future state practices provides two ways of looking at the future requirements of a complete RM (see Table 3.3).

Table 3.2 Improvement Areas for Generally Accepted Recordkeeping Principles

Improvement Area	Accountability	Transparency	Integrity	Protection	Compliance	Availability	Retention	Disposition
Roles and responsibilities	◇				◇		◇	
Policies and procedures	◇	◇	◇	◇	◇	◇	◇	◇
Communication and training	◇	◇		◇	◇		◇	
Systems and automation	◇			◇	◇	◇	◇	◇

Table 3.3 Assessment Report and Roadmap

Principle	Level	Findings	Requirements to Move to the Next Step

Accountability	Level 1 Substandard	■ No senior executive (or person of comparable authority) is responsible for the records management program. ■ The records manager role is largely nonexistent or is an administrative and/or clerical role distributed among general staff.	1. Assign records management responsibilities to senior executive. 2. Hire or promote records manager.
Transparency	Level 1 Substandard	■ It is difficult to obtain information about the organization or its records in a timely fashion. No clear documentation is readily available. ■ There is no emphasis on transparency. ■ Public requests for information, discovery for litigation, regulatory responses, or other requests (e.g. from potential business partners, investors, or buyers) cannot be readily accommodated. ■ The organization has not established controls to ensure the consistency of information disclosure. ■ Business processes are not well defined.	1. Develop policies and procedures. 2. Develop training for all levels of staff. 3. Identify requirements for records findability and accessibility. 4. Define business processes.
Integrity	Level 1 Substandard	■ There are no systematic audits or defined processes for showing the origin and authenticity of a record. ■ Various organizational functions use ad hoc methods to demonstrate authenticity and chain of custody, as appropriate, but their trustworthiness cannot easily be guaranteed.	1. Develop audit process. 2. Identify business activities for creation and storage of records.
Protection	Level 1 Substandard	■ No consideration is given to record privacy. ■ Records are stored haphazardly, with protection taken by various groups and departments with no centralized access controls. ■ Access controls, if any, are assigned by the author.	1. Assess security and access controls. 2. Develop access and security control scheme.
Compliance	[redacted]	■ The organization has identified all relevant compliance laws and regulations. ■ Record creation and capture are systematically carried out in accordance with records management principles. ■ The organization has a strong code of business conduct which is integrated into its overall information governance structure and recordkeeping policies. ■ Compliance and the records that demonstrate it are highly valued and measurable. ■ The hold process is integrated into the organization's information management and discovery processes for the “most critical” systems. ■ The organization has defined specific goals related to compliance.	1. Implement systems to capture and protect records. 2. Develop metadata scheme. 3. Develop remediation plan and implement corrective actions.
Availability	**Level 2 In Development**	■ Record retrieval mechanisms have been implemented in certain areas of the organization. ■ In those areas with retrieval mechanisms, it is possible to distinguish between official records, duplicates, and non-record materials. ■ There are some policies on where and how to store official records, but a standard is not imposed across the organization. ■ Legal discovery is complicated and costly due to the inconsistent treatment of information.	1. Develop enterprise classification scheme. 2. Identify user search and retrieval requirements. 3. Develop standards for managing the lifecycle of records.
Retention	**Level 2 In Development**	■ A retention schedule is available, but does not encompass all records, did not go through official review, and is not well known around the organization. ■ The retention schedule is not regularly updated or maintained. ■ Education and training about the retention policies are not available.	1. Develop enterprise wide functional retention schedule. 2. Map retention schedule to classification scheme. 3. Implement an annual review process for

			record series and legal research. 4. Develop training for classification scheme and retention schedule.
Disposition	**Level 2 In Development**	■ Preliminary guidelines for disposition are established. ■ There is a realization of the importance of suspending disposition in a consistent manner, repeatable by certain legal groupings. ■ There may or may not be enforcement and auditing of disposition.	1. Develop procedures for disposition of records. 2. Implement disposition processes. 3. Develop audit trails for records transfers and destruction.
Overall	Level 1 Substandard		

Information Security Principles

The information security aspects of your IG program should be guided by established principles.

Principle of Least Privilege

The Principle of Least Privilege (POLP) is an important cybersecurity maxim that means users should only be given access to the bare minimum permissions and information needed to do their job.[5] Under POLP, users are only given access to the files needed to perform their job function. POLP should be used to control who has access to which information, on which devices, and when.

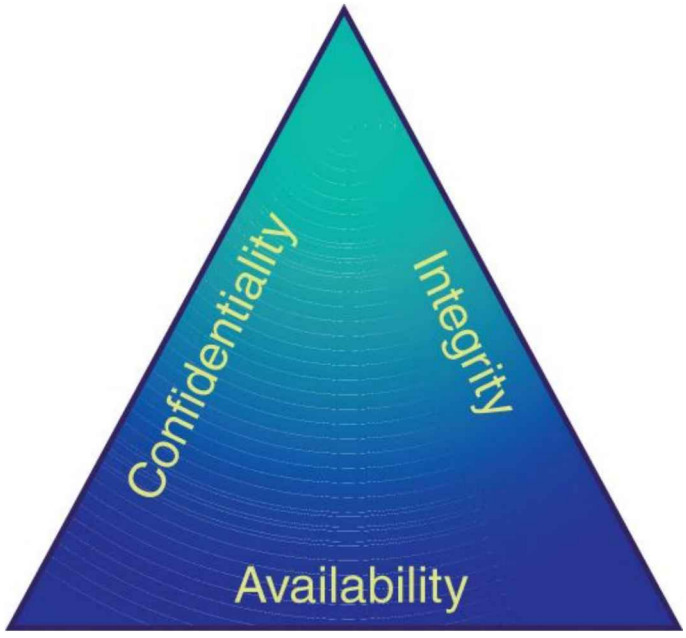
The CIA Triad

The **CIA triad** (sometimes referred to as the AIC triad to avoid confusion with the US government spy agency) depicts the three "most crucial components" of information security.[6]

Confidentiality (roughly equivalent to Generally Accepted Recordkeeping Principle® #4, Protection) means that access to private and sensitive is tightly controlled so that only authorized personnel have access to it. **Integrity** (the same as GAR Principle® #3) means that information has a reasonable assurance of being accurate, reliable, and trusted, throughout its lifecycle. **Availability (**the same as GAR Principle® #6) is the concept that information can be reliably and consistently accessed and retrieved by authorized employees, which requires software patches and updates are implemented in a timely way, and that hardware is maintained regularly.

Privacy Principles

The **Generally Accepted Privacy Principles** (GAPP) were developed jointly by the Canadian Institute of Chartered Accountants (CICA) and the American Institute of Certified Public Accountants (AICPA) through the AICPA/CICA Privacy Task Force. These principles can be used to guide the privacy aspects of an IG program. The field of information privacy is rapidly changing, and the International Association of Privacy Professionals (IAPP) is quite active globally with conferences, workshops, and training. IAPP's membership exploded in 2017–2018, when GDPR came into effect. Nevertheless, the 10 Generally Accepted Privacy Principles have been accepted by the privacy profession. The 10 Generally Accepted Privacy Principles and their criteria are:[7]



1. *Management*
   - The organization defines, documents, communicates, and assigns accountability for its privacy policies and procedures.
   - *Criteria*:
     - Privacy policies define and document all 10 GAPP.
     - Review and approval of changes to privacy policies are conducted by management.
     - Risk assessment process is in place to establish a risk baseline and regularly identify new or changing risks to personal data.
     - Infrastructure and systems management take into consideration impacts on personal privacy.
     - Privacy awareness training
2. *Notice*
   - The organization provides notice of its privacy policies and procedures. The organization identifies the purposes for which personal information is collected, used, and retained.
   - *Criteria*:
     - Communication to individuals
     - Provision of notice
     - Use of clear and conspicuous language
3. *Choice and consent*
   - The organization describes the choices available to the individual. The organization secures implicit or explicit consent regarding the collection, use, and disclosure of the personal data.
   - *Criteria*:
     - Communicating the consequences of denying/withdrawing consent
     - Consent for new purposes/uses of the personal data
     - Explicit consent for sensitive data
     - Consent for online data transfer
4. *Collection*
   - Personal information is only collected for the purposes identified in the notice (see #2).
   - *Criteria*:
     - Document and describe types of information collected and methods of collection
     - Collection of information by fair and lawful means, including collection from third parties
     - Inform individuals if information is developed or additional information is acquired
5. *Use, retention, and disposal*

- The personal information is limited to the purposes identified in the notice the individual consented to. The organization retains the personal information only for as long as needed to fulfill the purposes, or as required by law. After this period, the information is disposed of appropriately.
- *Criteria*:
  - Systems and procedures in place to ensure personal information is used, retained, and disposed appropriately

6. *Access*
   - The organization provides individuals with access to their personal information for review or update.
   - *Criteria*:
     - Confirmation of individual's identity before access is given to personal information
     - Personal information presented in understandable format
     - Access provided in reasonable time frame and at a reasonable cost
     - Statement of disagreement; the reason for denial should be explained to individuals in writing
7. *Disclosure to third parties*
   - Personal information is disclosed to third parties only for the identified purposes and with implicit or explicit consent of the individual.
   - *Criteria*:
     - Communication with third parties should be made known to the individual.
     - Information should only be disclosed to third parties that have equivalent agreements to protect personal information.
     - Individuals should be aware of any new uses/purposes for the information
     - The organization should take remedial action in response to misuse of personal information by a third party.
8. *Security for privacy*
   - Personal information is protected against both physical and logical unauthorized access.
   - *Criteria*:
     - Privacy policies must address the security of personal information.
     - Information security programs must include administrative, technical, and physical safeguards.
     - Logical access controls in place
     - Restrictions on physical access
     - Environmental safeguards
     - Personal information is protected when being transmitted (e.g. mail, Internet, public, or other nonsecure networks).
     - Security safeguards should be tested for effectiveness at least once annually.
9. *Quality*
   - The organization maintains accurate, complete, and relevant personal information that is necessary for the purposes identified.
   - *Criteria:*
     - Personal information should be relevant for the purposes it is being used.
10. *Monitoring and enforcement*
    - The organization monitors compliance with its privacy policies and procedures. It also has procedures in place to address privacy-related complaints and disputes.
    - *Criteria:*
      - Individuals should be informed on how to contact the organization with inquiries, complaints, and disputes.
      - Formal process is in place for inquires, complaints, or disputes.
      - Each complaint is addressed and the resolution is documented for the individual.
      - Compliance with privacy policies, procedures, commitments, and legislation is reviewed, documented, and reported to management.

Source: American Institute of Certified Public Accountants, https://www.aicpa.org/InterestAreas/InformationTechnology/Resources/PRIVACY/DownloadableDocuments/10252-346_Records%20Management-PRO.pdf.

These 10 privacy principles can be applied by organizations to establish and maintain the privacy aspects of their IG programs.

Utilizing the various sets of complementary IG principles to help educate stakeholders and guide the IG program will help to keep the scope of the program focused by providing some guidelines to keep it on track that help assure the success of the program.

Who Should Determine IG Policies?

When forming an IG steering committee or board, it is essential to include representatives from cross-functional groups, and at differing levels of the organization. It must be driven by an executive sponsor (see later chapter on securing and managing executive sponsorship), and include active members from key business units, as well as other departments or functions including privacy, cybersecurity, IT, legal, risk management, compliance, records management, and possibly finance. Then, corporate training/education and communications must be involved to keep employees trained and current on IG policies. This function may be performed by an outside consulting firm if there is no corporate education staff.

When forming an information governance steering committee or board, it is essential to include representatives from cross-functional group.

Knowledge workers, those who work with records and sensitive information in any capacity, best understand the nature and value of the records they work with as they perform their day-to-day functions. IG policies must be developed and also communicated clearly and consistently. *Policies are worthless if people do not know or understand them, or how to comply*. And training is a crucial element that will be examined in any compliance hearing or litigation that may arise. "Did senior management not only create the policies, but provide adequate training on them, on a consistent basis?" This will be a key question raised. For these reasons, a training plan is a necessary piece of IG, and education should be heavily emphasized.[8]

Knowledge workers, those who work with records in any capacity, best understand the nature and value of the records they work with.

The need for IG is increasing due to increased and tightened regulations, increased litigation, increased data volumes, and the increased incidence of theft and misuse of internal documents and records. *Organizations that do not have active IG programs should reevaluate IG policies and their internal processes following any major loss of records, the inability to produce accurate records in a timely manner, or any document security breach or theft*. If IG teams include a broad cross-section of critical players on the IG committee, and strong executive sponsorship, they will be better preparing the organization for legal and regulatory rigors, as well as unlocking new value in their information.

CHAPTER SUMMARY: KEY POINTS

- The Sedona Conference Commentary on Information Governance provides 11 principles to consider when implementing IG programs.
- Organizations without active IG programs should reevaluate IG policies and their internal processes following any major loss of records, the inability to produce accurate records in a timely manner, or any document security breach or theft.
- Principles of successful IG programs are emerging. They include executive sponsorship, information classification, integrity, security, accessibility, control, monitoring, and auditing. In addition, policy development and continuous improvement are included.
- Cross-functional collaboration is needed for IG policies to hit the mark and be effective.
- Lines of authority, accountability, and responsibility must be clear for the IG program to succeed.
- Accountability is a key aspect of IG.
- The Generally Accepted Recordkeeping Principles consists of eight principles that provide an IG framework that can support continuous improvement.
- An audit process must be developed to cover all aspects of IG in the organization.
- To be effective, policies must be formalized and integrated into business processes.
- Disposition is the last stage in the life cycle of records. Disposition is not synonymous with destruction, though destruction may be one disposal option.
- Knowledge workers, those who work with records in any capacity, best understand the nature and value of the records they work with.
- When forming an information governance steering committee or board, it is essential to include representatives from cross-functional groups.
- No IG effort will survive and be successful if it does not have an accountable, responsible executive sponsor.
- IG programs are not one-time projects but rather ongoing programs.
- The Principle of Least Privilege (POLP) is an important cybersecurity maxim that means users should only be given access to the bare minimum permissions and information needed to do their job.
- The CIA information security triad includes confidentiality, integrity, and availability, three principles that can be mapped back to three of the Generally Accepted Recordkeeping Principles.
- Privacy is a major issue today and a key aspect of IG programs. Privacy considerations should be injected into daily business processes.
- The 10 Generally Accepted Privacy Principles provide guidance for privacy programs.

Notes

1. https://thesedonaconference.org/publication/The%20Sedona%20Conference%C2%AE%20Commentary%20on%20Information%20Go (accessed October 14, 2017).
2. Debra Logan, "What Is Information Governance? And Why Is It So Hard?" posted January 11, 2010, http://blogs.gartner.com/debra_logan/2010/01/11/what-is-information-governance-and-why-is-it-so-hard/.
3. ARMA International, Generally Accepted Recordkeeping Principles, www.arma.org/garp/copyright.cfm (accessed May 8, 2012).
4. ARMA International, Information Governance Maturity Model, www.arma.org/garp/Garp%20maturity%20Model%20Grid%20(11x23).pdf (accessed June 12, 2012).
5. http://searchsecurity.techtarget.com/definition/principle-of-least-privilege-POLP (accessed March 12, 2018).
6. http://whatis.techtarget.com/definition/Confidentiality-integrity-and-availability-CIA (accessed March 12, 2018).
7. https://www.cippguide.org/2010/07/01/generally-accepted-privacy-principles-gapp/ (accessed October 31, 2017).
8. "Governance Overview (SharePoint Server 2010)," http://technet.microsoft.com/en-us/library/cc263356.aspx (accessed April 19, 2011).

PART TWO

Information Governance Risk Assessment and Strategic Planning

CHAPTER 4
Information Asset Risk Planning and Management

Information asset risk planning is a key information governance (IG) program activity. In fact, much of IG is about managing information risk, and often, information risk analysis is a regulatory obligation.[1] Many times organizations have identified risks to information, but have not taken the appropriate risk assessment and mitigation steps to counter those risks.

There are various types of risks to information assets, including the risk of noncompliance with legal regulations; technology risks centered around cybersecurity and system maintenance; external and internal data breaches; management risks related to managing change, system planning, and providing proper training; and even natural disasters or rare disasters caused by humans, such as the 9/11 attacks in New York City.

Information asset risk planning requires that the organization take a number of specific steps in identifying, analyzing, and countering information risks:

1. *Identify risks*. Conduct a formal process of identifying potential vulnerabilities and threats (both external and internal) to information assets.
2. *Assess impact*. Determine the potential financial and operational impact of the identified adverse events.
3. *Determine probability*. Weigh the likelihood that the identified risk events materialize.
4. *Countermeasures*. Create high-level strategic plans to mitigate the greatest risks.
5. *Create policy*. Develop strategic plans into specific policies.
6. *Establish metrics*. Determine metrics to measure risk reductions from mitigation efforts.
7. *Assign responsibilities*. Identify those who are accountable for executing the new risk mitigating processes and maintaining the processes in place.
8. *Execute plan*. Execute the information risk mitigation plan.
9. *Audit, review, adjust*. Audit the information risk mitigation plan and make adjustments.

Critically, these risk mitigation efforts must be tested and audited periodically not only to ensure conformance to the policies, but also to provide a feedback loop to revise and fine-tune policies and optimize **business processes**.

Information asset risk planning requires that the organization take a number of specific steps in identifying, analyzing, and countering information risks.

Some key benefits that flow from this information risk planning process include:

- Protection and preservation of information assets
- Protection of the organization's reputation, brand, and equity value
- Organizational “defense in depth” for privacy and cybersecurity
- A direct connection to enterprise information security practices which help to assure consumer or customer privacy
- Privacy controls that are clearly defined which reduce risks and support compliance efforts
- Privacy requirements that are measurable and enforceable
- Accountability in cybersecurity and privacy processes[2]

Depending on the jurisdiction, information is required by specific laws and regulations to be retained for specified periods (a compliance risk), and to be produced in specified situations. To determine which laws and regulations apply to your organization's information, research into the legal and regulatory requirements for information in the jurisdictions in which you operate must be conducted.

The Information Risk Planning Process

The risk planning steps, delineated in more detail, are as follows.

Step 1: Conduct a Formal Process of Identifying Potential Vulnerabilities and Threats

Some primary threats to information assets include:

Breaches. A key threat to all organizations is major data breaches. Breaches can not only compromise confidential information, but they also can represent a breach of consumer trust, which damages the organization's reputation and equity value. It also negatively affects employee morale and the ability to retain employees. *It is costly* to prepare for and prevent breaches, so threats must be prioritized based on the specific organization's risk profile. According to a recent study by the Ponemon Institute, the cost of a data breach averages almost $4 million, and the average cost for each lost or stolen record containing sensitive and confidential information was $148.[3]

Ransomware. Ransomware is a newer type of risk that organizations face. Ransomware attacks typically occur when hackers intrude computer systems and lock down crucial files with encryption, and then demand a fairly modest (although

this has been increasing) ransom payment to unlock the files. Perhaps the most widespread ransomware attack to date occurred in May 2017 with the Wannacry attacks that infected over 200,000 Windows systems including computers at 48 hospital trusts in the United Kingdom, crippling operations. The attack spread to European countries and to the United States, and even included attacks that compromised medical devices.[4] Hackers know that daily hospital operations depend on IT systems and that often management will decide to pay rather than disrupt operations. Rogue hackers are getting more sophisticated and savvy. They recently introduced **Ransomware-as-a-service** kits that they sell to other rogue operators, which can be customized to hit a particular target. Often the developer of the kit will take a percentage of the proceeds of successful ransomware attacks.[5]

Noncompliance fines and sanctions. Another major risk is compliance violations, and the potential for large fines. A major violation of the EU GDPR privacy regulation can result in a fine of as much as 4% of total annual revenues. In the United States, the new (2020) California Consumer Privacy Act also will require strict privacy requirements. Further, Sarbanes-Oxley violations for public companies can run into the millions of dollars, as can HIPAA violations for healthcare institutions. These actions have not only immediate financial impact but also can erode the organization's reputation in the marketplace, which would impact future revenues and even shareholder equity value.

Other compliance and legal risks. There are additional compliance and legal risks to identify and research. Federal, provincial, state, and even municipal laws and regulations may apply to the retention period for business or consumer information. Organizations operating in multiple jurisdictions must maintain compliance with laws and regulations that may cross national, state, or provincial boundaries. Legally required privacy actions and retention periods must be researched for each jurisdiction (country, province, state, and even city) in which the business operates, so that it complies with all applicable laws. Legal counsel and records managers (or the IG lead) must conduct their own legislative research to apprise themselves of mandatory information retention requirements, as well as privacy considerations and requirements, especially in regard to PII and PHI. This regulatory information must be analyzed and structured and then presented to legal staff for discussion. Then further legal and regulatory research must be conducted, and firm legal opinions must be rendered by the organization's legal counsel regarding information retention and privacy and security requirements in accordance with laws and regulations. This is an absolute requirement. The legal staff or outside legal counsel should provide input as to the **legal hold notification** (LHN) process, provide opinions and interpretations of law that applies to a particular organization, provide input on the value of formal records to arrive at a consensus on records that have legal value to the organization, and construct an appropriate retention schedule.

Legal requirements take priority over all others. The retention period for confidential data or a particular type of record series must meet minimum retention, privacy, and security requirements as mandated by law. Business needs

and other considerations are secondary. So, legal research is required before determining and implementing retention periods, privacy policies, and security measures. In identifying information requirements and risks, legal requirements trump all others.

In order to locate the regulations and citations relating to retention of records, there are two basic approaches. The first approach is to use a records retention citation service, which publishes in electronic form all of the retention-related citations. These services are usually purchased on a subscription basis, as the citations are updated on an annual or more frequent basis as legislation and regulations change.

Another approach is to search the laws and regulations directly using online or print resources. Records retention requirements for corporations operating in the United States may be found in the **Code of Federal Regulations** (CFR). "The Code of Federal Regulations (CFR) annual edition is the codification of the general and permanent rules published in the *Federal Register* by the departments and agencies of the federal government. It is divided into 50 titles that represent broad areas subject to federal regulation."[6]

For governmental agencies, a key consideration is complying with requests for information as a result of the US Freedom of Information Act (FOIA), Freedom of Information Act 2000 (in the UK), and similar legislation in other countries. So the process of governing information is critical to meeting these requests by the public for governmental records.

Major risks that organizations face include data breaches, ransomware, compliance violations, and legal risks.

Step 2: Determine the Potential Financial and Operational Impact of the Identified Adverse Events

Benchmarking data from peer organizations provides reasonable projections of potential financial and operational impact. For instance, when banks of similar size and business model are fined for noncompliance, others in that market can expect the regulators to come knocking. That is, if Banks A and B have been fined over $1 billion for noncompliance, then Bank C can reasonably presume that they are facing the same size financial risk. Also, a list of major breaches and ransomware attacks at peer organizations and their estimated costs should be considered in the calculations of potential financial impact. These estimates should then be normalized and brought into line with the size of an organization, with considerations given to the competitive, regulatory, and economic environment within which it operates.

Step 3: Weigh the Likelihood That the Identified Risk Events Materialize

In this step, percentages are assigned to the potential adverse events that have been identified. Whereas a major breach event could cost the organization, say, $5 million dollars, its likelihood may be low, in the 3–5% range. Risk management professionals use certain tried and true methodologies to assess the likelihood that an event may occur, which may be leveraged. Or possibly, senior management may have internal models developed to assess risk likelihood that are specific to the organization. Absent standard methodologies, the IG Steering Committee should utilize their experience and information from external input to assess the likelihood that an adverse event may occur.

Once percentages have been assigned, an **expected value** (EV) calculation can be made. For instance, if a major breach would cost an estimated $5 million, and its likelihood is 5 percent, then the expected value of the financial impact of that event is:

$$EV = \$5,000,000 \times 5\% = \$250,000$$

If the exposure from a compliance violation has led to fines at peer organizations in the $2 million range, and your organization holds a fairly weak compliance posture, perhaps the likelihood is 10 percent. The EV calculation would then be:

$$EV = \$2,000,000 \times 10\% = \$200,000$$

And in like manner the potential financial impact of other identified risk events may be calculated, so they can then be ranked and prioritized. This gives executive management the information they need to make budget decisions. Clearly, the risks that are most likely to have a greater financial impact are those that must be mitigated as a priority.

Many organizations create a formalized **risk profile** to more accurately assess risks the organization faces.

> Expected value calculations can help IG program managers rank and prioritize risks.

Create a Risk Profile

According to ISO, risk is defined as "the effect of uncertainty on objectives" and a *risk profile* is "a description of a set of risks."[7] Creating a risk profile is a basic building block in **enterprise risk management** (yet *another* ERM acronym), which assists executives in understanding the risks associated with stated business objectives, and allocating resources, within a structured evaluation approach or framework.

There are multiple ways to create a risk profile, and how often it is done, the external sources consulted and stakeholders who have input will vary from

organization to organization.[8] A key tenet to bear in mind is that simpler is better, and that sophisticated tools and techniques should not make the process overly complex. Creating a risk profile involves identifying, documenting, assessing, and prioritizing risks that an organization may face in pursuing its business objectives. Those associated risks can be evaluated and delineated within an IG framework.

> The risk profile is a high-level, executive decision input tool.

The corporate risk profile should be an informative tool for executive management, the CEO, and the board of directors, so it should reflect that tone. In other words, it should be clear, succinct, and simplified. A risk profile may also serve to inform the head of a division or subsidiary, in which case it may contain more detail. The process is applicable and can also be applied to public and nonprofit entities.

The time horizon for a risk profile varies, but looking out three to five years is a good rule of thumb.[9] The risk profile typically will be created annually, although semiannually would serve the organization better and account for changes in today's dynamic business and regulatory environment. But if an organization is competing in a market sector with rapid business cycles or volatility, the risk profile should be generated more frequently, perhaps quarterly.

There are different types of risk profile methodologies, with a **"Top 10"** list, **risk map**, and **heat map** being commonly used. The first is a simple identification and ranking of the 10 greatest risks in relation to business objectives. The risk map is a visual tool that is easy to grasp, with a grid depicting a likelihood axis and an impact axis, usually rated on a scale of 1–5. In a risk assessment meeting, stakeholders can weigh in on risks, using voting technology to generate a consensus. A heat map is a color-coded matrix generated by stakeholders voting on risk level by color (e.g. red being highest).

> A common risk profile method is to create a prioritized or ranked "Top 10" list of greatest risks to information. Other methods include risk maps and heat maps.

Information gathering is a fundamental activity in building the risk profile. Surveys are good for gathering basic information, but for more detail, a good method to employ is direct, person-to-person interviews, beginning with executives and risk professionals.[10] Select a representative cross-section of functional groups to gain a broad view. Depending on the size of the organization, you may need to conduct 20–40 interviews, with one person asking the questions and probing, while another team member takes notes and asks occasionally for clarification or elaboration. Conduct the interviews in a compressed timeframe—knock them out within one to three weeks and do not drag the process out, as

business conditions and personnel can change over the course of months.

There are a number of helpful considerations to conducting successful interviews. First, prepare some questions for the interviewee in advance, so they may prepare and do some of their own research. Second, schedule the interview close to their office, and at their convenience. Third, keep the time as short as possible, but long enough to get the answers you will need: approximately 20–45 minutes. Be sure to leave some open time between interviews to collect your thoughts and prepare for the next one. And follow up with interviewees after analyzing and distilling the notes to confirm you have gained the correct insights.

The information you will be harvesting will vary depending on the interviewee's level and function. You will need to look for any hard data or reports that show performance and trends related to information asset risk. There may be benchmarking data available as well. Delve into information access and security policies, policy development, policy adherence, and the like. Ask questions about retention of e-mail and legal hold processes. Ask about records retention and disposition policies. Ask about privacy policies. Ask about their data deletion policies. Ask about long-term preservation of digital records. Ask for documentation regarding IG-related training and communications. Dig into policies for access to confidential data and vital records. Try to get a real sense of the way things are run, what is standard operating procedure, and also how workers might get around overly restrictive policies or operate without clear policies. Learn enough so that you can firmly grasp the management style, corporate culture, and appetite for risk, and then distill that information into your findings.

> Once a list of risks is developed, grouping them into basic categories helps stakeholders to more easily grasp them and consider their likelihood and impact.

Key events and developments must also be included in the risk profile. For instance, a major data breach, the loss or potential loss of a major lawsuit, pending regulatory changes that could impact your IG policies, or a change in business ownership or structure must all be accounted for and factored into the information asset risk profile. Even changes in governmental leadership should be considered, if they might impact regulations impacting the organization. These types of developments should be tracked on a regular basis, and should continue to feed into the risk equation.[11] You must observe and incorporate an analysis of key events in developing and updating the risk profile.

When you get to this point, it should be possible to generate a list of specific potential risks. *It may be useful to group or categorize the potential risks into clusters such as technology, natural disaster, regulatory, safety, competitive, management policy, and so forth.* Armed with this list of risks, you should solicit input from stakeholders as to likelihood and timing of the threats or risks. As the

organization matures in its risk identification and handling capabilities, a good practice is to look at the risks and their ratings from the previous years to attempt to gain insights into change and trends—both external and internal—that affected the risks.

Step 4: Create High-Level Strategic Plans to Mitigate the Greatest Risks

After identifying the major risk events the organization faces, and calculating the potential financial impact, the IG Steering Committee must develop possible countermeasures to reduce the risks, and their impact if they do occur. This means creating an **information asset risk mitigation plan**. Various risk mitigation options should be explored for each major risk, and then the required tasks to reduce the specified risks and improve the odds of achieving business objectives should be delineated.[12] Considering all the documentation that has been collected and analyzed in creating the risk profile and risk assessment, the specific tasks and accountabilities should be laid out and documented. *The information asset risk mitigation plan must include key milestones and metrics and a timetable for implementation of the recommended risk mitigation measures*. Some of the major tasks will include developing a robust and consistent security awareness training program, implementing a privacy awareness training program, acquiring new IT tools, developing risk countermeasure implementation plans, assigning roles and responsibilities to carry them out, and developing an audit and verification process to ensure mitigation actions are being taken and are effective.

> The information asset risk mitigation plan develops risk reduction options and tasks to reduce specified risks and improve the odds for achieving business objectives.

A helpful exercise and visual tool is to draw up a table of top risks, their potential impact, actions that have been taken to mitigate the risk, and suggested new risk countermeasures, as in Table 4.1.

Table 4.1 Risk Assessment

What Are the Risks?	How Might They Impact Business Objectives?	Actions and Processes Currently in Place?	Additional Resources Needed to Manage This Risk?	Action by Whom?	Action by When?	Done
Breach of Confidential Documents	Compromise confidential information Compromise competitive position Compromise business negotiations	Utilizing ITIL and CobiT IT Frameworks Annual CIS Top 20 security assessment Published security policies Published privacy policies Annual security Audits	Hold security awareness training Implement newer technologies including Information Rights Management (IRM) Privacy Awareness training Implement Quarterly Audits	IT Staff, Chief Info Security Officer Chief Privacy Officer IT Audit	01/10/2021	06/1/2021

Step 5: Develop Strategic Plans into Specific Policies

The strategic plans will be high level and must be forged into everyday policies to embed IG considerations into daily operations. Creating or updating policies in multiple areas will be required, along with metrics to measure how well the information asset risk mitigation is being implemented. Some policy areas that may need to be reviewed include: privacy notice and privacy policy; e-mail policies, specifically when handling PII, confidential, or sensitive information; text messaging and instant messaging (IM) policies when handling confidential information; social media use policies; mobile device policies, especially when handling PII, confidential, or sensitive information; policies for the use of cloud computing platforms; if the organization uses SharePoint there must be updated governance policies on the appropriate use of the portal; personnel policies, especially when handling PII, confidential or sensitive information, and other areas as needed by a specific organization.

Step 6: Determine Metrics to Measure Risk Reductions from Mitigation Efforts

The IG program must be measured and controlled. Objective ways to measure

conformance and performance of the program must be developed. This requires quantitative measures that are meaningful and measure progress. Stakeholder consultation is required in order that meaningful metrics are created.

Determining relevant ways of measuring progress will allow executives to see progress, as, realistically, *reducing risk is not something anyone can see or feel*—it is only in the failure to do so, when the risk comes home to roost, when the painful realizations are made. Also, tracking valid metrics help to justify investment in the IG program.

Assigning some quantitative measures that are meaningful and do, in fact, measure progress may take some serious effort and consultation with stakeholders. Determining relevant ways of measuring progress will allow executives to see progress, as, realistically, reducing risk is not something anyone can see or feel—it is only in the failure to do so, when the risk comes home to roost, when the painful realizations are made. Also, valid metrics help to justify investment in the IG program.

The proper metrics will vary from organization to organization, but some examples of specific metrics may be:

- Reduce the number of stolen or misplaced laptops by 50% over the previous fiscal year
- Reduce the number of hacker intrusion events by 75% over the previous fiscal year
- Reduce e-discovery collection and review costs per GB by 25% over the previous fiscal year
- Reduce the number of adverse findings in the risk and compliance audit by 50% over the previous fiscal year
- Provide security awareness training (SAT) to 100% of the headquarters workforce this fiscal year, and maintain a continual education program
- Roll out the implementation of IRM software to protect confidential e-documents to 50 users this fiscal year
- Provide confidential “vanishing” messaging services for the organization's 20 top executives this fiscal year
- Reduce the number of medical errors due to poor or untimely information by 10 percent over the previous fiscal year.

Your organization's metrics should be tailored to address the primary goals of your IG program, and should tie directly to stated business objectives.

Metrics are required to measure progress in the information risk mitigation plan.

Step 7: Identify Those Who Are Accountable for Executing the New Risk Mitigating Processes and Maintaining the Processes in Place

From the IG Steering Committee, specific individuals should be assigned to be held accountable for specific tasks that are set out in the information asset risk mitigation plan. Sometimes this may mean a small team; a subgroup of the larger IG Steering Committee is assigned accountability, since IG crosses functional boundaries. Generally, those individuals who have expertise in a certain area are best to assign accountabilities in their area. For instance, if the organization is planning on encrypting all PII when transmitted, as well as when at rest (stored), the RIM manager and Chief Information Security Officer (CISO) or IT representative will need to work together to roll out the new capability and to develop a training and communications plan. A good tool to use is a **RACI** (responsible, accountable, consulted, informed) **matrix** that pinpoints key accountabilities.

Step 8: Execute the Risk Mitigation Plan

Executing the risk mitigation plan requires that regular project team meetings are set up, and key reports on the information asset risk mitigation metrics are tracked to manage the process. Standard project/program management tools and techniques should be utilized, as they are proven. Some additional tools may be needed, such as leveraging data mapping software or perhaps developing an **information asset register**, advanced analytics, collaboration software, artificial intelligence (AI) software, knowledge management software, or even social media within the organization.

The most important part of managing an IG program is clear and regular communications, to keep the team updated and motivated, and to smoke out any rising problems or challenges. Everyone on the IG team must be kept up to date on the progress being made in the information risk reduction effort. Teams go through various stages of conflict and compromise, so do not expect that things will always go smoothly. Encouraging healthy conflict can yield positive results. But the executive sponsor and IG lead must have high levels of emotional intelligence to manage conflict in a positive way so that team members do not feel slighted when their ideas are not adopted as they proposed. The sum of the team's efforts should be emphasized.

Step 9: Audit the Information Asset Risk Mitigation Plan

Using the metrics you have developed, there must be a process in place to separately and independently audit compliance to information risk mitigation measures, to see that they are being implemented. The result of the audit should be a useful input in improving and fine-tuning the program. It should not be viewed as an opportunity to cite shortfalls and implement punitive actions. It should primarily be a periodic and regular feedback loop into the IG program.

It may be wise to use an internal auditor, or even an external auditor or consultant

to measure the progress of the information asset risk mitigation plan, based on the metrics established by the IG team. The output of the audit process will provide useful input for fine-tuning and improving the program.

> The information risk mitigation plan must be audited to provide feedback to fine-tune the program.

Information Risk Planning and Management Summary

To summarize, an information risk assessment can be compressed into five basic steps:[13]

1. *Identify the risks:* This should be an output of creating a risk profile, but if conducting a risk assessment, first identify the major information-related risks.
2. *Determine potential impact:* If a calculation of a range of economic impact is possible (e.g. lose $10M in legal damages), then include it. If not, be as specific as possible as to how a negative event related to an identified risk can impact business objectives.
3. *Evaluate risk levels and probabilities and recommend action:* Based on a prioritized list of information risks, assign probabilities, determine the potential impact, and develop countermeasures. This may be in the form of recommending new procedures or processes, new investments in information technology, or other actions to mitigate identified risks.
4. *Create a report with recommendations and implement:* This may include a risk assessment table as well as written recommendations, then implement.
5. *Review periodically:* Audit annually or semiannually, as appropriate for your organization.

CHAPTER SUMMARY: KEY POINTS

- Information asset risk planning requires that the organization take a number of specific steps in identifying, analyzing, and countering information risks.
- Major risks that organizations face include data breaches, ransomware, compliance violations, and legal risks. Also, internal bad actors.
- Expected value calculations can help IG program managers rank risks.
- In identifying information requirements and risks, legal requirements take priority over all others.
- In the United States, the Code of Federal Regulations lists information retention requirements for businesses, divided into 50 subject matter areas.
- The risk profile is a high-level, executive decision input tool.
- A common risk profile method is to create a prioritized or ranked "Top 10" list of greatest risks to information. Other tools include risk maps and heat maps.
- Once a list of risks is developed, grouping them into basic categories helps stakeholders to more easily grasp them and consider their likelihood and impact.
- The risk mitigation plan develops risk reduction options and tasks to reduce specified risks and improve the odds for achieving business objectives.
- Metrics are required to measure progress in the risk mitigation plan.
- The information asset risk mitigation plan must be audited to provide feedback to fine-tune the program.

Notes

1. Elizabeth Snell, "The Role of Risk Assessments in Healthcare," http://healthitsecurity.com/features/the-role-of-risk-assessments-in-healthcare (accessed February 1, 2018).
2. Eric Basu, "Implementing a Risk Management Framework for Health Information Technology Systems—NIST RMF," *Forbes*, August 3, 2013, www.forbes.com/sites/ericbasu/2013/08/03/implementing-a-risk-management-framework-for-health-information-technology-systems-nist-rmf/#23e63d46523a.
3. "2018 Cost of a Data Breach Study by Ponemon,"

https://www.ibm.com/security/data-breach (accessed August 22, 2018).

4. Thomas Fox-Brewster, "Medical Devices Hit by Ransomware for the First Time in U.S. Hospitals," Forbes.com, May 17, 2017, https://www.forbes.com/sites/thomasbrewster/2017/05/17/wannacry-ransomware-hit-real-medical-devices/#2fc718b9425c.

5. Ibid.

6. U.S. Government Publishing Office (GPO), "Code of Federal Regulations," www.gpo.gov/help/index.html#about_code_of_federal_regulations.htm (accessed March 6, 2016).

7. "ISO 31000 2009 Plain English, Risk Management Dictionary," www.praxiom.com/iso-31000-terms.htm (accessed March 25, 2013).

8. John Fraser and Betty Simkins, eds., *Enterprise Risk Management: Today's Leading Research and Best Practices for Tomorrow's Executives* (Hoboken, NJ: John Wiley & Sons, 2010), 171.

9. Ibid., 172.

10. Ibid.

11. Ibid., 179.

12. Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK Guide),* 4th ed., ANSI/PMI 99-001-2008, pp. 273–312.

13. "Controlling the Risks in the Workplace," Health Safety Executive, www.hse.gov.uk/risk/controlling-risks.htm (accessed March 6, 2016).

CHAPTER 5
Strategic Planning and Best Practices for Information Governance

A **strategic plan** is the process of envisioning your organization's desired future state, developing business objectives that must be accomplished to progress toward it, and then determining the steps and milestones needed to achieve the desired future state. Your information governance (IG) strategic plan should support and be in alignment with the organization's overall strategic plan.

Securing a sponsor at the executive management level is always crucial to projects and programs, and this is especially true of any strategic planning effort. An executive must be on board and supporting the effort in order to garner the resources needed to develop and execute the strategic plan, and that executive must be held accountable for the development and execution of the plan. These axioms apply to the development of an IG strategic plan.

Also, resources are needed—time, human capital, and budget money. The first is a critical element: it is not possible to require managers to take time out of their other duties to participate in a project if there is no executive edict and consistent follow-up, support, and communication. Executive sponsorship is a best practice. And, of course, without an allocated budget, no program can proceed.

The higher your executive sponsor is in the organization, the better.[1] The implementation of an IG program may be driven at a high level by the general counsel (GC), chief risk officer, chief compliance officer, chief information officer (CIO), or, ideally, the chief executive officer (CEO). With CEO sponsorship come many of the key elements needed to complete a successful project, including allocated management time, budget money, and management focus.

It is important to bear in mind that this IG effort is truly a *change management* effort, in that it aims to change the structure, guidelines, and rules within which employees operate. *The change must occur at the very core of the organization's culture*. It must be embedded permanently, and for it to be, the message must be constantly and consistently reinforced. Achieving this kind of change requires commitment from the very highest levels of the organization.

> Executive sponsorship is critical to project success. There is no substitute. Without it, a project is at risk of failure.

If the CEO is not the sponsor, then another high-level executive must lead the effort and be accountable for meeting milestones as the program progresses. Programs with no executive sponsor or an unenthusiastic one can lose momentum

and focus, especially as competing projects and programs are evaluated and implemented. Program failure is a great risk without a strong executive sponsor. Such a program likely will fade or fizzle out or be relegated to the back burner. Without strong high-level leadership, when things go awry, finger pointing and political games may take over, impeding progress and cooperation.

The executive sponsor must be actively involved, tracking program objectives and milestones on a regular, scheduled basis and ensuring they are aligned with business objectives. He or she must be aware of any obstacles or disputes that arise, take an active role in resolving them, and push the program forward.

Crucial Executive Sponsor Role

The role of an executive sponsor is high level, requiring periodic and regular attention to the status of the program, particularly with budget issues, staff resources, and milestone progress. The role of a program or project manager (PM) is more detailed and day-to-day, tracking specific tasks that must be executed to make progress toward milestones. Both roles are essential. The savvy PM brings in the executive sponsor to push things along when more authority is needed but reserves such project capital for those issues that absolutely cannot be resolved without executive intervention. It is best for the PM to keep the executive sponsor fully informed but to ask for assistance only when absolutely needed.

At the same time, the PM must manage the relationship with the executive sponsor, perhaps with some gentle reminders, coaxing, or prodding, to ensure that the role and tasks of executive sponsorship are being fulfilled. “[T]he successful Project Manager knows that if those duties are not being fulfilled, it's time to call a timeout and have a serious conversation with the Executive Sponsor about the viability of the project.”[2]

The executive sponsor serves six key purposes on a project:

1. *Budget*. The executive sponsor ensures an adequate financial commitment is made to see the project through and lobbies for additional expenditures when change orders are made or cost overruns occur.
2. *Planning and control*. The executive sponsor sets direction and tracks accomplishment of specific, measureable business objectives.
3. *Decision making*. The executive sponsor makes or approves crucial decisions and resolves issues that are escalated for resolution.
4. *Expectation management*. The executive sponsor must manage expectation, since success is quite often a stakeholder perception.
5. *Anticipation*. Every project that is competing for resources can run into unforeseen blockages and objections. Executive sponsors run interference and provide political might for the PM to lead the project to completion, through a series of milestones.
6. *Approval*. The executive sponsor signs off when all milestones and objectives

have been met.

An eager and effective executive sponsor makes all the difference to an IG program—if the role is properly managed by the PM. It is a tricky relationship, since the PM is always below the executive sponsor in the organization's hierarchy, yet the PM must coax the superior into tackling certain high-level tasks. Sometimes a third-party consultant who is an expert in the specific project can instigate and support requests made of the sponsor and provide a solid business rationale.

> While the executive sponsor role is high level, the PM's role and tasks are more detailed and involve day-to-day management.

Evolving Role of the Executive Sponsor

The role of the executive sponsor necessarily evolves and changes over the life of the initial IG program launch, during the implementation phases, and on through the continued IG program.

To get the program off the ground, the executive sponsor must make the business case and get adequate budgetary funding. But an effort such as this takes more than money; it takes *time*—not just time to develop new policies and implement new technologies, but the time of the designated PM, program leaders, and needed program team members.

In order to get this time set aside, the IG program must be made a top priority of the organization. It must be recognized, formalized, and aligned with organizational business objectives. All this up-front work is the responsibility of the executive sponsor.

Once the IG program team is formed, team members must clearly understand why the new program is important and how it will help the organization meet its business objectives. This message must be regularly reinforced by the executive sponsor; he or she must not only paint the vision of the future state of the organization but articulate the steps in the path to get there.

When the formal program effort commences, the executive sponsor must remain visible and accessible. He or she cannot disappear into everyday duties and expect the program team to carry the effort through. The executive sponsor must be there to help the team confront and overcome business obstacles as they arise and must praise the successes along the way. This requires active involvement and a willingness to spend the time to keep the program on track and focused.

The executive sponsor must be the lighthouse that shows the way, even through cloudy skies and rough waters. This person is the captain who must steer the ship, even if the first mate (PM) is seasick and the deckhands (program team) are drenched and tired.

After the program is implemented, the executive sponsor is responsible for

maintaining its effectiveness and relevance. This is done through periodic compliance audits, testing and sampling, and scheduled meetings with the ongoing PM.

> The role of the executive sponsor changes during the inception, planning, and execution of the IG program.

Building Your IG Team

Who should make up the IG team? Although there are no set requirements or formulas, the complex nature of IG and the fact that it touches upon a number of specialized disciplines and functional areas dictates that a cross-functional approach be taken. Therefore you will need representatives from several departments. There are some absolutes: you must have an executive sponsor and an IG program manager, hopefully a chief IG officer. And based on the Information Governance Reference Model and empirical research, you'll need a representative from your legal staff or outside counsel, your information technology (IT) department, a **senior records officer (SRO)** or the equivalent, an information security professional, and hopefully a privacy professional, especially in this era of GDPR, California Consumer Privacy Act, and emerging privacy compliance legislation around the globe. In addition, there may be a need for input from your chief data officer (CDO), managers of compliance, risk management, human resources (for training and communications), and certain business units that could benefit most from IG. You also may want to recruit the CFO, based on the idea that preventing breaches and unauthorized access or misuse of information can damage the brand, and cause a loss in equity value, and also that the CFO can provide input into approaches to leveraging and monetizing information assets.

The most appropriate business units to participate are those with the most pressing IG issues. It could be the department with the most litigation, where litigation costs and risk could be substantially cut. Or the department where information is either inaccurate or not quickly found, which causes compliance violations, fines, or sanctions, or compromises in customer service. Or it could be the department with the greatest opportunities to monetize and leverage information as an asset.

Depending on the scope of the effort, other possible IG team members might include an analytics specialist; a change management specialist; an audit lead; the chief knowledge officer (CKO) for knowledge management (KM); the corporate or agency archivist, business analysts, litigation support head, business process specialist, project management professional, and other professionals in functions related to these areas.

The risk mitigation plan develops risk reduction options and tasks to reduce specified risks and improve the odds for achieving business objectives.

Assigning IG Team Roles and Responsibilities

The executive sponsor will need to designate an IG program manager (PM). Depending on the focus of the IG effort, that person could come from several areas, including legal, privacy, cybersecurity, compliance, risk management, records management, or IT.

In terms of breaking down the roles and responsibilities of the remainder of the IG team, the easy decision is to have IG team representatives take responsibility for the functional areas of their expertise. But there will be overlap, and it is best to have some pairs or small work groups teamed up to gain the broadest amount of input and optimum results. This will also facilitate cross training. For instance, inside legal counsel may be responsible for rendering the final legal opinions, but because they are not expert in records, document management, or risk management, they could benefit from input of others in specialized functional areas, which will inform them and help narrow and focus their legal research. Basic research into which regulations and laws apply to the organization regarding security, retention, and preservation of e-mail, e-records, and PII or PHI could be conducted by the SRO or records management head, in consultation with the corporate archivist and CIO, with the results of their findings and recommendations drafted and sent to the legal counsel. The draft report may offer up several alternative approaches that need legal input and decisions. Then the legal team lead can conduct its own focused research and make final recommendations regarding the organization's legal strategy, business objectives, financial position, and applicable laws and regulations.

The IG team must include a cross-functional group of stakeholders from various departments, including legal, records management, IT, and risk management.

The result of the research, consultation, and collaboration of the IG team should result in a final draft of the IG strategic plan. It will still need more input and development to align the plan with business objectives, an analysis of internal and external drivers, applicable best practices, competitive analysis, applicable IT trends, an analysis and inclusion of the organization's culture, and other factors.

Align Your IG Plan with Organizational Strategic Plans

The IG plan must support the achievement of the organization's business

objectives and therefore must be melded into the organization's overall strategic plan. Integration with the strategic plan means that the business objectives in the IG plan are consistent with, and in support of, the enterprise strategic plan.

So, for example, if the corporate strategy includes plans for acquiring smaller competitors and folding them into the organization's structure as operating divisions, then the IG plan must assist and contribute to this effort. Plans for standardizing operating policies and procedures must include a consistent, systematized approach to the components of IG, including stakeholder consultation, user training and communications, and compliance audits. The IG plan should bring a standard approach across the spectrum of information use and management within the organization and it must be forged to accommodate the new technology acquisitions. This means that e-mail policies, e-discovery policies, mobile device policies, social media policies, cloud collaboration and storage use, and even nitty-gritty details like report formats, data structures, document taxonomies, and metadata must be consistent and aligned with the overall strategic plan. In other words, the goal is to get all employees on the same page and working to support the business objectives of the strategic plan in everyday small steps within the IG plan.

> The IG strategic plan must be aligned and synchronized with the organization's overall strategic plans, goals, and business objectives.

The organization will also have an IT plan that must be aligned with the strategic plan to support overall business objectives. The IT strategy may be moving to a cloud-based approach, which means that cloud-based solutions should be considered first, to align with the IT plan. Or, the IT strategy could be to convert new acquisitions to the internal financial and accounting systems of the organization and to train new employees to use the existing software applications under the umbrella of the IG plan. Again, the IG plan needs to be integrated with the IT strategy and must consider the organization's approach to IT.

The result of the process of aligning the IG effort with the IT strategy and the organization's overall strategic plan will mean, ideally, that employee efforts are more efficient and productive since they are *consistently moving toward the achievement of the organization's overall strategic goals*. The organization will be healthier and will have less dissent and confusion with clear IG policies that leverage the IT strategy and help employees pursue overall business objectives.

Further considerations must be folded into the IG plan. As every corporate culture is different and has a real impact on decision-making and operational approaches, corporate culture must be included in the plan. Corporate culture includes the organization's appetite for risk, its use of IT (e.g. forward-thinking first adopter versus laggard), its capital investment strategies, and other management actions, which may be characterized as conservative, progressive/aggressive, or somewhere in between.

So, if the organization is conservative and risk averse, it may want to hold off on implementing some emerging content analytics or e-discovery technologies that can cut costs but also induce greater risk. Or if it is an aggressive, progressive, risk-taking organization, it may opt to test and adopt newer e-discovery technologies under the IT strategy and umbrella of IG policies. An example may be the use of blockchain technology to develop new applications. Or implementing artificial intelligence (AI), such as **predictive coding** technology in **early case assessment** (ECA). Predictive coding uses text auto-classification technology and neural technology with the assistance of human input to "learn" which e-documents might be relevant in a particular legal matter and which may not be. Through a series of steps of testing and checking subsets of the documents, humans provide input to improve the document or e-mail sorting and selection process. The software uses **machine learning** (a form of artificial intelligence whereby the software can change and improve on a particular task, as its decision engine is shaped and "trained" by input**)** to improve its ability to cull through and sort documents.

Predictive coding can reduce e-discovery costs, yet there are risks that the approach can be challenged in court and could, in fact, affect the case adversely. Thus, a decision on a technology like predictive coding can involve and include elements of the IG plan, IT strategy, and overall organizational strategic plan.

And there are resource issues to consider: How much management time, or bandwidth, is available to pursue the IG plan development and execution? Is there a budget item to allow for software acquisitions and training and communications to support the execution of the IG plan? Obviously, without the allocated management time and budget money, the IG plan cannot be executed.

Survey and Evaluate External Factors

The IG plan is now harmonized and aligned with your organization's strategic plan and IT strategy, but you are not finished yet, because the plan cannot survive in a vacuum: organizations must analyze and consider the external business, legal, and technological environment and fold their analysis into their plans.

Analyze IT Trends

IG requires IT to support and monitor implementation of polices, so it *matters* what is developing and trending in the IT space. What new technologies are coming online? Are you tracking developments in AI, blockchain, and the Internet of Things (IoT)? Why are they being developed and becoming popular? How do these changes in the business environment that created opportunities for new technologies to be developed affect your organization and its ability execute its IG plan? How can new technologies assist? Which ones are immature and too risky? These are some of the questions that must be addressed in regard to the changing IT landscape.

The IG strategic plan must be informed with an assessment of relevant technology trends.

Some changes in **information and communications technology** (ICT) are rather obvious, such as the trends toward mobile computing, tablet and smartphone devices, cloud storage, and social media use. Each one of these major trends that may affect or assist in implementing IG needs to be considered within the framework of the organization's strategic plan and IT strategy. If the corporate culture is progressive and supportive of remote work and telecommuting, and if the organizational strategy aims to lower fixed costs by reducing the amount of office space for employees and moving to a more mobile workforce, then trends in collaborative software, and in tablet and smartphone computing that are relevant to your organization, must be analyzed and considered. Is the organization going to provide mobile devices or support a bring-your-own-device environment? Which equipment will you support? Will you support iOS, Android, or both? What is your policy going to be on phone jacking (changing communications carrier settings)? What is the IG policy regarding confidential documents on mobile devices? Will you use encryption? If so, which software? Is your enterprise moving to the cloud computing model? Utilizing social media? What about **Big Data**? Are you going to consider deploying auto-classification and predictive coding technologies? What are the trends that might affect your organization?

Many, many questions must be addressed, but the evaluation must be narrowed down to those technology trends that specifically might impact the execution of your IG plan and rollout of new technology.

On a more granular level, you must evaluate even supported file and document formats. It gets that detailed when you are crafting IG policy. For instance, PDF/A-1 is the standard format for archiving electronic documents. So your plans must include **long-term digital preservation** (LTDP) standards and best practices for those records that must be stored to document the heritage of the organization.

Survey Business Conditions and the Economic Environment

If the economy is on a down cycle, and particularly if your business sector has been negatively affected, resources may be scarcer than in better times. Hence, it may be more difficult to get budget approval for necessary program expenses, such as new technologies, staff, training materials, communications, and so forth. This means your IG plan may need to be scaled back or its scope reduced. Implementing the plan in a key division rather than attempting an enterprise rollout may be the best tactic in tough economic times. Also, there are a number of activities that can be executed at a relatively low cost to move the IG program along, such as policy development, taxonomy development, updating departmental file plans, and so forth.

But if things are booming and the business is growing fast, budget money for

investments in the IG program may be easier to secure, and the goals may be expanded.

> Trends and conditions in the internal and external business environment must be included in the IG strategic plan.

IG *must be* an ongoing program, but it takes time to implement, and it takes temporal, human, and financial resources to execute, audit, and continue to refine. So an executive looking for a quick and calculable payback on the investment may want to focus on narrower areas. For instance, the initial focus may be entirely on shared drive cleanup of redundant, obsolete, and trivial (ROT) information. Or providing security awareness training (SAT) to employees who handle information to lower risk. Or it could focus on the legal hold and e-discovery process, with business objectives that include reducing pretrial costs and attorney fees by a certain percentage, ratio, or amount. Concrete results can be seen when focusing on e-discovery, since legal costs are real, and always will be there. The business case may be more difficult to make if the IG effort is broader in focus. If the focus is on improving search capabilities, for faster and more accurate retrieval, the organization will benefit as a whole, but it will take time to see results. When the results are evident, management decision making, as well as compliance capabilities, will be improved. Improved management decision making will improve the organization's competitiveness in the long term, but it may be difficult to cite specific examples where costs were saved or revenues were increased as a result of the “better decisions” that should come about through better IG.

Analyze Relevant Legal, Regulatory, and Political Factors

In consultation with your legal team or lead, the laws and regulations that affect your industry should be identified. Narrowing the scope of your analysis, those that specifically could impact your governance of information should be considered and analyzed. What absolute requirements do they impose? Where there is room for interpretation, where, legally, does your organization want to position itself? How much legal risk is acceptable? For instance, practical organizations may focus on those regulations that regulators are focusing on for that particular cycle. These are the types of questions you will have to look to your legal and risk management professionals to make. Again, *legal requirements take priority over all others*.

> Laws and regulations relevant to your organization's management and distribution of information in all jurisdictions must be considered and included in the IG strategic plan. Legal requirements take priority over all others.

Your decision process must include considerations for the future and anticipated future changes. Changes in the legal and regulatory environment happen based on the political leaders who are in place and any pending legislation. So you must go further and analyze the current political environment and make some judgments based on the best information you can gather, the organization's culture and appetite for risk, management style, available resources, and other factors. Generally, a more conservative environment means less regulation, and this analysis must also be folded into your IG strategic plan.

Survey and Determine Industry Best Practices

IG is a developing hybrid discipline. In a sense, it's a superset of records and information management (RIM) and a subset of **governance, risk management, and compliance** (GRC), a discipline that emerged to help executives manage risk and compliance at a high level.

IG developed due to the explosion in the amount of e-mail, records, documents, and data that must be managed in today's increasingly high-volume and velocity business environment and highly regulated and litigious compliance environment. As such, best practices are still being formed and added to. This process of testing, proving, and sharing IG best practices will continue for the next decade as the practices are expanded, revised, and refined.

The most relevant study of IG best practices is one that is conducted for your organization and surveys your industry and what some of your more progressive competitors are doing in regard to IG. Often the best way to accomplish such a study is by engaging a third-party consultant, who can more easily contact, study, and interview your competitors in regard to their practices. Business peer groups and trade associations also can provide some consensus as to emerging best practices.

> Include a best practices review in your IG strategic plan. The most relevant best practices in IG are those in your industry proven by peers and competitors.

Twenty-one examples of IG best practices covering a number of areas in which IG has an impact or should be a major consideration are listed next.

1. *Executive sponsorship is crucial.* Securing a committed, engaged executive sponsor at the senior management level is key to successful IG programs. It is not possible to require managers to take time out of their other duties to participate in a project if there is no executive edict. The executive sponsor must *own* the business case for the IG program, and have a long-term vested interest in its success. It is advisable to also have a deputy executive sponsor to help support the program and assure the durability of IG program leadership.
2. *Establish a cross-functional IG council or steering committee*. There must be

a holistic view of information use in the organization, which seeks to leverage it as an asset, and to reduce its risks and costs. At a minimum, there must be representation from Legal, IT, Privacy, Information Security, RIM, and possibly Finance, and Human Resources, and, depending on the organization and its focus, perhaps other key groups such as Risk Management, Data Governance, Analytics, Knowledge Management, and more.

3. *Create a formal IG Program Charter for guidance*. It should include the overall mission and goals of the IG program, and should list IG committee members and their basic responsibilities, as well as the meeting schedule. It also should show the reporting structure of the IG committee members and delineate their basic program responsibilities. *It is advisable to form a small, top-tier "decision committee" to facilitate decisions and recommendations made to the executive sponsor, otherwise, decision making can become slowed and ineffective*. The IG Program Charter should be signed off on by the executive sponsor.
4. *Develop an overall organizational strategy for the IG program*. This will ensure there is agreement on the aims and foci of the program, and help the various functional groups involved to collaborate and cooperate to execute the IG program strategy. "An overarching strategy is needed—including ... organizational performance and risk mitigation—to establish organization's goals and priorities, and consistently drive these through information systems and business processes."[3]
5. *IG is not a project, but rather an ongoing program*. IG programs are "evergreen" and should eventually become embedded into routine operations. True, there must be discrete projects executed under the overall IG program, which provides an umbrella of guidelines and policies. Performance is then monitored and enforced with the support of metrics, information technologies, and audit tools.

   Compare the IG program to a workplace safety program which is continuously improved, reinforced, and expanded; every time a new location, team member, piece of equipment, or toxic substance is acquired by the organization, the workplace safety program dictates how that is handled and, if it doesn't, workplace safety policies/procedures/training need to be updated. The program must be monitored and audited to ensure the program is followed and to make adjustments. The effort never ends.[4]
6. *Using an IG framework or maturity model is helpful in assessing and guiding IG programs*. Various models are offered. The Information Governance Reference Model, which grew out of the Electronic Discovery Reference Model (both found at EDRM.net),[5] can be used early on in developing IG programs to communicate the need for cross-functional collaboration, and to develop the core team. The Information Governance Process Maturity Model (IGPMM), from the Compliance, Governance, and Oversight Council (CGOC), is a comprehensive assessment tool that measures

IG program maturity in 22 core IG processes. The IGPMM was released in 2012 and updated and expanded in 2017 to include privacy and data protection obligations, a new data security cost lever, cloud computing safeguards, a greater focus on data governance, and other considerations. For analyzing records management program functions, the Generally Accepted Recordkeeping Principles® from ARMA International are useful and widely used (hence "recordkeeping" in its title).

7. *Business processes must be redesigned when implementing new technologies to streamline operations and maximize impact*. Implementing new technologies without redesigning processes will not provide the maximum benefit and impact to the organization.
8. *Leverage analytics to improve decision making and possibly find new value*. The entire range of analytics, from descriptive to diagnostic to predictive to prescriptive analytics, must be deployed to fully exploit data value.[6] It is crucial to have a robust data governance program in place to assure data quality so the analytics are accurate. Beyond that, the organization should look for ways to monetize data, either directly or indirectly.
9. *Focus data governance efforts heavily on data quality*. Improved data quality and availability will help reduce errors, improve decision making, improve customer satisfaction, improve the professional environment, and improve financial performance.
10. *Creating standardized metadata terms should be part of an IG effort that enables faster, more complete, and more accurate searches and retrieval of records*. This is important not only in everyday operations, but also for conducting analysis of content for new insights. Good metadata management also assists in the maintenance of corporate memory and improving accountability in business operations.[7] Using a standardized format and controlled vocabulary provides a "precise and comprehensible description of content, location, and value."[8] Using a controlled vocabulary means the organization has standardized a set of terms used for metadata elements describing records. This ensures consistency and helps with optimizing search and retrieval functions, as well as meeting e-discovery requests, compliance demands, and other legal and regulatory requirements.
11. *Defensible deletion of data debris and information that no longer has value is critical in the era of Big Data*. You must have IG polices in place and be able to prove that you follow them consistently and systematically in order to justify, to the courts and regulators, deletion of information. With a smaller information footprint, organizations can more easily find what they need and derive business value from it.[9] Data debris must be eliminated regularly and consistently, and to do this, processes and systems must be in place to cull out valuable information and discard the data debris. An IG program sets the framework to accomplish this.

12. *IG policies must be developed before enabling technologies are deployed to assist in enforcement*. After the policy-making effort, seek out the proper technology tools to assist in monitoring, auditing, and enforcement.
13. *To provide comprehensive e-document security throughout a document's life cycle, documents must be secured upon creation using highly sophisticated technologies, such as information rights management (IRM) technology*. IRM acts as a sort of "security wrapper" that denies access without proper credentials. Document access and use by individuals having proper and current credentials is also tightly monitored. IRM software controls the access, copying, editing, forwarding, and printing of documents using a policy engine that manages the rights to view and work on an e-document. Access rights are set by levels or "roles" that employees are responsible for within an organization.
14. *A records retention schedule and legal hold notification (LHN) process are two foundational elements of a fundamental IG program*. These are the basics. Implementation will require records inventorying, taxonomy development, metadata normalization and standardization, and a survey of LHN best practices.
15. *An information risk mitigation plan is a critical part of the IG planning process*. An information risk mitigation plan helps in developing risk mitigation options and tasks to reduce the specified risks and improve the odds of achieving business objectives.[10]
16. *Proper metrics are required to measure the conformance and performance of your IG program*. You must have an objective way to measure how you are doing, which means numbers and metrics. Assigning some quantitative measures that are meaningful before rolling out the IG program is essential.
17. *IG programs must be audited for effectiveness*. Periodic audits will tell you how your organization is doing and where to fine-tune your efforts. To keep an IG program healthy, relevant, and effective, changes and fine-tuning will always be required.
18. *Business processes must be redesigned to improve and optimize the management and security of information and especially the most critical information,* ***electronic records****,* before *implementing enabling technologies*. For instance, using **electronic records management** (ERM) and workflow software fundamentally changes the way people work, and greater efficiencies can be gained with business process redesign (versus simply using ERM systems as electronic filing cabinets to speed up poor processes).
19. *Personal archiving of e-mail messages should be disallowed*. Although users will want to save certain e-mail messages for their own reasons, control and management of e-mail archiving must be at the organization level or as high a level as is practical, such as division or region.

20. *Destructive retention of e-mail helps to reduce storage costs and legal risk while improving "findability" of critical records*. It makes good business sense to have a policy to, say, destroy all e-mail messages after 90 or 120 days that are not flagged as potential records (which, e.g., help document a transaction or a situation that may come into dispute in the future) or those that have a legal hold.
21. *Some digital information assets must be preserved permanently as part of an organization's documentary heritage*.[11] It is critical to identify records that must be kept for the long term as early in the process as possible; ideally, these records should be identified prior to or upon creation. LTDP applies to content that is born digital as well as content that is converted to digital form. Digital preservation is defined as long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span that the information is required to be retained. Dedicated repositories for historical and cultural memory, such as libraries, archives, and museums, need to move forward to put in place trustworthy digital repositories that can match the security, environmental controls, and wealth of descriptive metadata that these institutions have created for analog assets (such as books and paper records). Digital challenges associated with records management affect all sectors of society—academic, government, private, and not-for-profit enterprises—and ultimately citizens of all developed nations.

Formulating the IG Strategic Plan

Now comes the time to make sense of all the information and input your IG team has gathered and hammer it into a workable IG strategic plan. Doing this will involve some give-and-take among IG team members, each having their own perspective and priorities. Everyone will be lobbying for the view of their functional groups. It is the job of the executive sponsor to set the tone and to emphasize organizational business objectives so that the effort does not drag out or turn into a competition but is a well-informed consensus development process that results in a clear, workable IG strategic plan.

Synthesize Gathered Information and Fuse It into IG Strategy

At this point your IG team will have gathered a great deal of information that needs to be analyzed and distilled into actionable strategies. This process will depend on the expertise and input of the specialized knowledge your team brings to the table within your organizational culture. Team members must be able to make decisions and establish priorities that reflect organizational business objectives and consider a number of influencing factors.

Do not prolong the strategy development process—the longer it lasts, the more key factors influencing it can change. You want to develop a strategic plan that is durable enough to withstand changes in technology, legislation, and other key influencing factors, but it should be relevant to that snapshot of information that

was collected early on. When all the parts and pieces start changing and require reconsideration, a dated IG plan does not serve the organization well.

Develop IG strategies for each of the critical areas, including the legal hold process, e-discovery action plans, e-mail policy, mobile computing policy, IT acquisition strategy, confidential document handling, vital records and disaster planning, social media policy, and other areas that are important to your organization. To maintain focus, do this first without regard to the prioritization of these areas.

> Fuse the findings of all your analyses of external and internal factors into your IG strategic plan. Develop strategies and then prioritize them.

Then you must go through the hard process of prioritizing your strategies and aligning them to your organizational goal and objectives. This may not be difficult in the beginning—for instance, your IG strategies for legal holds and e-discovery readiness are likely going to take higher priority than your social media policy, and protecting vital records is paramount to any organization. As the process progresses, it will become more challenging to make trade-offs and establish priorities. Then you must tie these strategies to overall organizational goals and business objectives.

A good technique to keep goals and objectives in mind may be to post them prominently in the meeting room where these strategy sessions take place. This will help to keep the IG team focused.

Develop Actionable Plans to Support Organizational Goals and Objectives

Plans and policies to support your IG efforts must be developed that identify specific tasks and steps and define roles and responsibilities for those who will be held accountable for their implementation. This is where the rubber meets the road. But you cannot simply create the plan and marching orders: You must build in periodic checks and audits to test that new IG policies are being followed and that they have hit their mark. Invariably, there will be adjustments made continually to craft the policies for maximum effectiveness and continued relevance in the face of changes in external factors, such as legislation and business competition, and internal changes in management style and structure.

Create New IG Driving Programs to Support Business Goals and Objectives

You have got to get things moving, get employees motivated, and launching new subprograms within the overall IG program is a good way to start. For instance, a new security awareness training (SAT) program for knowledge workers which is fun, engaging, and gamified can energize the IG program and immediately reduce information risk, while demonstrating that senior management is prudent and

proactive.

An “e-discovery readiness” initiative can show almost immediate results if implemented properly, with the support of key legal and records management team members, driven by the executive sponsor. You may want to revamp the legal hold process to make it more complete and verifiable, assigning specific employees specific tasks to be accountable for. Part of that effort may be evaluating and implementing new technology-assisted review (TAR) processes and predictive coding technology. So you will need to bring in the IG team members responsible for IT and perhaps business analysis. Working cooperatively on smaller parts of the overall IG program is a way to show real results within defined time frames. Piecing together a series of program components is the best way to get started, and it breaks the overall IG program down into digestible, doable chunks. A small win early on is crucial to maintain momentum and executive sponsorship. E-discovery has real costs, yet progress can be measured objectively in terms of reducing the cost of activities such as early case assessment (ECA). Benefits can be measured in terms of reduced attorney review hours, reduced costs, and reduced time to accomplish pretrial tasks.

> Create supporting subprograms to jump-start your IG program effort. Smaller programs should be able to measure real results based on metrics that are agreed on in advance.

To be clear, you will need to negotiate and agree on the success metrics by which the program will be measured in advance.

There are other examples of supporting IG subprograms, such as shared drive ROT cleanup and remediation; updating departmental file plans and the records retention schedule (RRS); or e-mail management and archiving, where storage costs, search times, and information breaches can be measured in objective terms. Or you may choose to roll out new policies for the use of mobile devices within your organization, where adherence to policy can be measured by scanning mobile devices and monitoring their use.

Draft the IG Strategic Plan and Gain Input from a Broader Group of Stakeholders

Once you have the pieces of the plan drafted and the IG team is in agreement that it has been harmonized and aligned with overall organizational goals and objectives, you must test the waters to see if you have hit the mark. It is a good practice to expose a broader group of stakeholders to the plan to gain their input. Perhaps your IG team has become myopic or has passed over some points that are important to the broader stakeholder audience. Solicit and discuss their input, and to the degree that there is a consensus, refine the IG strategic plan one last time before finalizing it. But remember, it is a living document, a work in progress, which will require revisiting and updating to ensure it is in step with changing

external and internal factors. Periodic auditing and review of the plan will reveal areas that need to be adjusted and revised to keep it relevant and effective.

Get Buy-in and Sign-off and Execute the Plan

Take the finalized plan to executive management, preferably including the CEO, and present the plan and its intended benefits to them. Field their questions and address any concerns to gain their buy-in and the appropriate signatures. You may have to make some minor adjustments if there are significant objections, but, if you have executed the stakeholder consultation process properly, you should be very close to the mark. Then begin the process of implementing your IG strategic plan, including regular status meetings and updates, steady communication with and reassurance of your executive sponsor, and planned audits of activities.

CHAPTER SUMMARY: KEY POINTS

- Engaged and vested executive sponsors are necessary for IG program success. It is not possible to require managers to take time out of their other duties to participate in a project if there is no executive edict or allocated budget.
- The executive sponsor must be: (1) directly tied to the success of the program, (2) fully engaged in and aware of the program, and (3) actively eliminating barriers and resolving issues.
- The role of the executive sponsor evolves over the life of the IG program and IG program effort. Initially, the focus is on garnering the necessary resources, but as the program commences, the emphasis is more on supporting the IG program team and clearing obstacles. Once the program is implemented, the responsibilities shift to maintaining the effectiveness of the program through testing and audits.
- While the executive sponsor role is high level, the project manager's role and tasks involve more detailed and day-to-day management.
- The risk mitigation plan develops risk reduction options and tasks to reduce specified risks and improve the odds for achieving business objectives.
- The IG team must include a cross-functional group of stakeholders from various departments, including legal, records management, IT, and risk management.
- The IG strategic plan must be aligned and synchronized with the organization's overall strategic plans, goals, and business objectives.
- The IG strategic plan must include an assessment of relevant technology trends.
- Trends and conditions in the internal and external business environment

must be included in the IG strategic plan.

- Laws and regulations relevant to your organization's management and distribution of information in all jurisdictions must be considered and included in the IG strategic plan. Legal requirements trump all others.
- Include a best practices review in your IG strategic plan. The most relevant best practices in IG are those in your industry proven by peers and competitors. (Twenty-one IG best practices are listed in this chapter for the first time in print.)
- Fuse the findings of all your analysis of external and internal factors into your IG strategic plan. Develop strategies and then prioritize them.
- Creating supporting subprograms to jump-start your IG program effort. Smaller programs should be able to measure real results based on metrics that are agreed on in advance.
- Make sure to get executive sign-off on your IG strategic plan before moving to execute it.

Notes

1. Roger Kastner, "Why Projects Succeed—Executive Sponsorship," February 15, 2011, http://blog.slalom.com/2011/02/15/why-projects-succeed-%E2%80%93-executive-sponsorship/.
2. Ibid.
3. https://www.infogovbasics.com/best-practices/by-industry/healthcare/ (accessed February 7, 2018).
4. Monica Crocker, e-mail to author, June 21, 2012.
5. EDRM, "Information Governance Reference Model (IGRM) Guide," www.edrm.net/resources/guides/igrm (accessed November 30, 2012).
6. AHIMA Staff, "Use Cases Demonstrate Information Governance Best Practices," *Journal of AHIMA*, September 30, 2014, https://journal.ahima.org/2014/09/30/use-cases-demonstrate-information-governance-best-practices/.
7. Kate Cumming, "Metadata Matters," in *Managing Electronic Records*, eds. Julie McLeod and Catherine Hare (London: Facet Publishing, 2005), 34.
8. Minnesota State Archives, Electronic Records Management Guidelines, www.mnhs.org/preserve/records/electronicrecords/ermetadata.html (accessed March 6, 2016).
9. Randolph A. Kahn,

https://twitter.com/InfoParkingLot/status/273791612172259329 (November 28, 2012).

10. Project Management Institute, *A Guide to the Project Management Body of Knowledge* (*PMBOK Guide*), 4th ed. (Newtown Square, PA: Project Management Institute, 2008), ANSI/PMI 99-001-2008, pp. 273–312.

11. Charles Dollar and Lori Ashley, e-mail to author, August 10, 2012.

CHAPTER 6
Information Governance Policy Development

To develop overarching **information governance** (IG) policies, you must inform and frame them with established principles, models, internal and external frameworks, best practices, and standards—those that apply to your organization and the scope of its planned IG program. Best practices within your industry segment are the most relevant, and there may be some that have been established within your organization that can be leveraged.

Your IG policy framework will actually be a collection of linked and consistent policies across multiple areas of the organization.

In this chapter, we first present and discuss key IG frameworks and models and then identify key standards for consideration.

The Sedona Conference IG Principles

The Sedona Conference® Commentary on Information Governance

In Chapter 3, we introduced the Sedona Conference IG principles. These can help steer your program and educate program stakeholders, especially in the early stages. A good exercise is to take the 11 principles and have a group rewrite them in their own words, referencing the organization's business objectives and scenario.

The Sedona IG principles state that IG programs should look at information as an organization-wide asset with associated risks that must be managed, while finding value; should maintain independence; should include all information stakeholders; must have an assessment to form strategic objectives; should have the resources and accountability to ensure a reasonable chance at success; must include defensible disposition; should act in good faith in reconciling varying laws and obligations; must preserve long-term digital assets; should leverage new technologies; and should be reviewed periodically to ensure needs and objectives are being met. These principles are quite useful in guiding policy development in IG programs. To review The Sedona Conference Commentary on IG in more detail, please refer to Chapter 3.

A Brief Review of Generally Accepted Recordkeeping Principles®

In Chapter 3, we also introduced and discussed ARMA International's eight Generally Accepted Recordkeeping Principles, known as the Principles[1] (or sometimes as GAR Principles). These Principles and associated metrics provide a framework, particularly applied to **records management** processes, that can support continuous improvement.

To review, the eight Principles are:

1. Accountability
2. Transparency
3. Integrity
4. Protection
5. Compliance
6. Availability
7. Retention
8. Disposition[2]

The Principles establish benchmarks for how organizations of all types and sizes can build and sustain compliant, legally defensible records management (RM) programs. Using the maturity model (also presented in Chapter 3), organizations can assess where they are in terms of records management maturity, identify gaps, and take steps to improve across the eight areas the Principles cover.

Although the author and advocate of The Principles, ARMA International, promotes them as IG principles, they were, in fact, designed as the Generally Accepted *Recordkeeping* Principles®, and are best applied as

such.

IG Reference Model

In late 2012, with the support and collaboration of ARMA International and the Compliance, Governance and Oversight Council (CGOC), the Electronic Discovery Reference Model (EDRM) Project released version 3.0 of its Information Governance Reference Model (IGRM), which added **information privacy and security** "as primary functions and stakeholders in the effective governance of information."[3] These areas have grown in importance since then. The model is depicted in Figure 6.1.

The IGRM is aimed at fostering IG adoption by facilitating communication and collaboration between disparate (but overlapping) IG stakeholder functions, including information technology (IT), legal, RIM, privacy and security, and business unit stakeholders.[4] It is a good tool to use in the early stages of introducing an IG program to stakeholders. The Model also aims to provide a common, practical framework for IG that will foster adoption of IG in the face of new Big Data challenges and increased legal and regulatory demands. It is a clear snapshot of where IG fundamentally "lives" and shows critical interrelationships and unified governance.[5] It can help organizations to forge policy in an orchestrated way and embed critical elements of IG policy across functional groups. Ultimately, implementation of IG helps organizations leverage information value, reduce risk, and address legal demands.

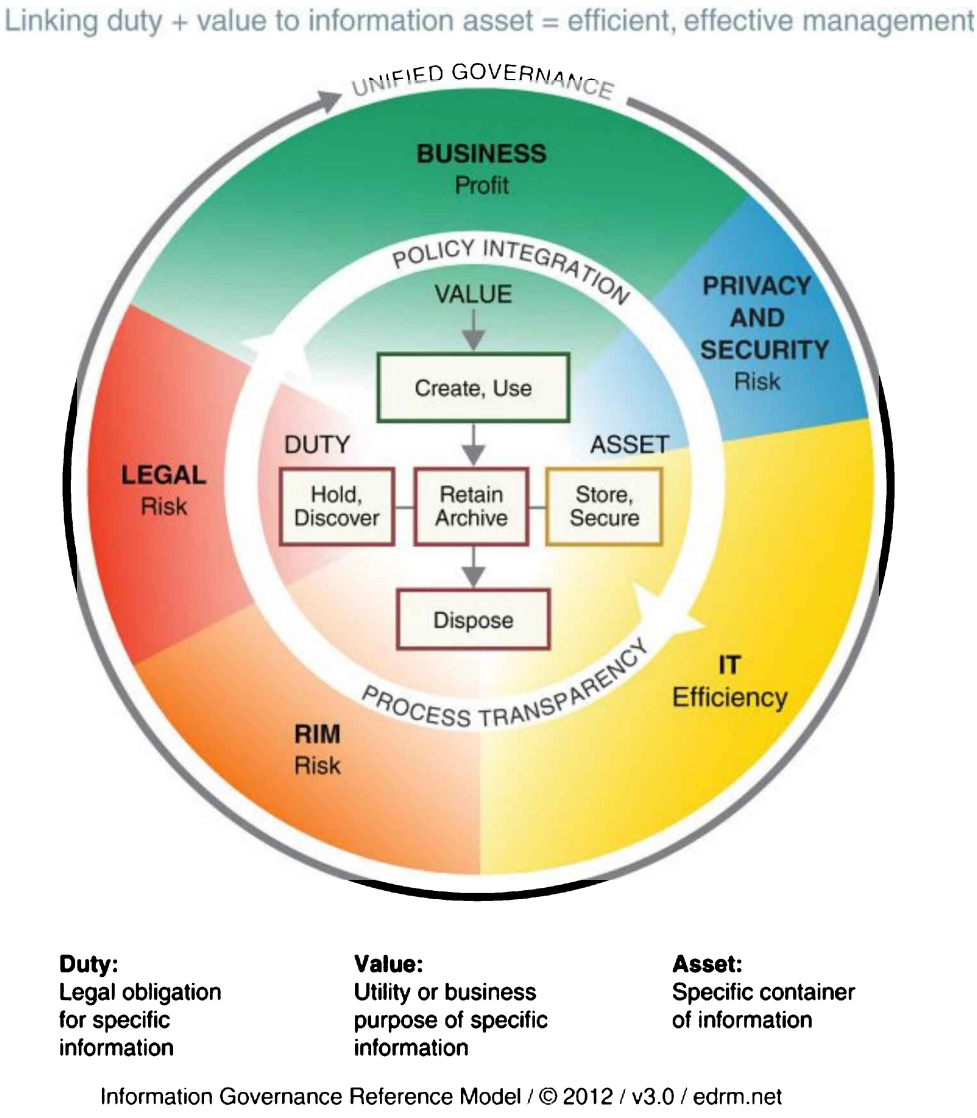


Figure 6.1 Information Governance Reference Model

Source: EDRM.net.

The growing CGOC community (3,800-plus members and rising) has widely adopted the IGRM and developed the IG Process Maturity Model (IGPMM) that leverages IGRM v3.0, and assesses organizational IG maturity along 22 key processes.[6] The IGPMM is a thorough tool for IG program assessments.

> You must inform and frame IG policy with principles, models, internal and external frameworks, best practices, and standards.

Interpreting the IGRM Diagram*

Outer Ring

Starting from the outside of the diagram, successful information management is about conceiving a complex set of interoperable processes and implementing the procedures and structural elements to put them into practice. It requires:

- An understanding of the business imperatives of the enterprise.
- Knowledge of the appropriate tools and infrastructure for managing information.

- Sensitivity to the legal and regulatory obligations with which the enterprise must comply.

For any piece of information you hope to manage, the primary stakeholder is the business user of that information [emphasis added]. We use the term "business" broadly; the same ideas apply to end users of information in organizations whose ultimate goal might not be to generate a profit.

Once the business value is established, you must also understand the legal duty attached to a piece of information. The term "legal" should also be read broadly to refer to a wide range of legal and regulatory constraints and obligations, from e-discovery and government regulation, to contractual obligations such as payment card industry requirements.

Finally, IT organizations must manage the information accordingly, ensuring privacy and security as well as appropriate retention as dictated by both business and legal or regulatory requirements.

> The business user is the primary stakeholder of managed information.

Center

In the center of the diagram is a workflow or life cycle diagram. We include this component in the diagram to illustrate the fact that *information management is important at all stages of the information life cycle—from its creation through its ultimate disposition*. This part of the diagram, once further developed, along with other secondary-level diagrams, will outline concrete, actionable steps that organizations can take in implementing information management programs.

> Information management is important at all stages of the life cycle.

Even the most primitive business creates information in the course of daily operations, and IT departments spring up to manage the logistics; indeed, one of the biggest challenges in modern organizations is trying to stop individuals from excess storing and securing of information. Legal stakeholders can usually mandate the preservation of what is most critical, though often at great cost. However, it takes the coordinated effort of all three groups to defensibly dispose of a piece of information that has outlived its usefulness and retain what *is* useful in a way that enables accessibility and usability for the business user.

> Legal stakeholders can usually mandate the preservation of what is most critical, though often at great cost.

How the IGRM Complements the Generally Accepted Recordkeeping Principles[†]

The IGRM supports ARMA International's Principles by identifying the cross-functional groups of key information governance stakeholders and by depicting their intersecting objectives for the organization. This illustration of the relationship among duty, value, and the information asset demonstrates cooperation among stakeholder groups to achieve the desired level of maturity of effective information governance.

Effective IG requires a continuous and comprehensive focus. The IGRM will be used by proactive organizations as an introspective lens to facilitate visualization and discussion about how best to apply the Principles. The IGRM puts into sharp focus the Principles and provides essential context for the maturity model.

> The IGRM was developed by the EDRM Project to foster communication among stakeholders and adoption of IG. It complements ARMA's GAR Principles.

Best Practices Considerations

IG best practices should also be considered in policy formulation. Best practices in IG are evolving and expanding, and those that apply to organizational scenarios may vary. A best practices review should be

conducted, customized for each particular organization.

In Chapter 5, we provided a list of 21 IG best practices with some detail. The IG world is maturing, and additional best practices will evolve and develop. The 21 best practices, summarized below, are fairly generic and widely applicable. Bear in mind that the best practices most applicable to your IG program are those developed within your industry, or, especially, within your organization:

1. Executive sponsorship is crucial.
2. Establish a cross-functional IG council or steering committee.
3. Create a formal IG Program Charter for guidance.
4. Develop an overall Organizational Strategy for the IG Program.[7]
5. IG is not a project but rather an ongoing program.[8]
6. Using an IG framework or maturity model is helpful in assessing and guiding IG programs.
7. Business processes must be redesigned when implementing new technologies to streamline operations and maximize impact.
8. Leverage analytics to improve decision making and possibly find new value.
9. Focus data governance efforts heavily on data quality.
10. Creating standardized metadata terms should be part of an IG effort that enables faster, more complete, and more accurate searches and retrieval of records.
11. Defensible deletion of data debris and information that no longer has value is critical in the era of Big Data and increased compliance regulations.
12. IG policies must be developed before enabling technologies are deployed to assist in enforcement.
13. To provide comprehensive e-document security throughout an e-document's, life cycle documents must be secured upon creation using highly sophisticated technologies, such as encryption and information rights management (IRM) technology.
14. A records retention schedule and legal hold notification (LHN) process are two foundational elements of a fundamental IG program.
15. An information risk mitigation plan is a critical part of the IG planning process.[9]
16. Proper metrics are required to measure the conformance and performance of your IG program.
17. IG programs must be audited for effectiveness.
18. Business processes must be redesigned to improve and optimize the management and security of information and especially the most critical of information, electronic records, before implementing enabling technologies.
19. Personal archiving of e-mail messages should be disallowed.
20. Destructive retention of e-mail helps to reduce storage costs and legal risk while improving “findability” of critical records.
21. Some digital information assets must be preserved permanently as part of *an organization's documentary heritage*.

Standards Considerations

There are two general types of standards: de jure and de facto. De jure (“the law”) standards are those published by recognized standards-setting bodies, such as the International Organization for Standardization (ISO), American National Standards Institute (ANSI), National Institute of Standards and Technology (NIST) —this is what most people refer to it as (they do not know what the acronym stands for), British Standards Institute (BSI), Standards Council of Canada, and Standards Australia. Standards promulgated by authorities such as these have the formal status of standards.

De facto (“the fact”) standards are not formal standards but are regarded by many as if they were. They may arise though popular use (e.g. MS Windows at the business desktop in the 2001–2010 decade) or may be published by other bodies, such as the US National Archives and Records Administration (NARA) or Department of Defense (DoD) for the US military sector. They may also be published by formal standards-

setting bodies without having the formal status of a "standard" (such as some technical reports published by ISO).[10]

Benefits and Risks of Standards

Some benefits of developing and promoting standards are:

- *Quality assurance support*. If a product meets a standard, you can be confident of a certain level of quality.
- *Interoperability support*. Some standards are detailed and mature enough to allow for system interoperability between different vendor platforms.
- *Implementation frameworks and certification checklists*. These help to provide guides for projects and programs to ensure all necessary steps are taken.
- *Cost reduction*, due to supporting uniformity of systems. Users have lower maintenance requirements and training and support costs when systems are more uniform.
- *International consensus*. Standards can represent "best practice" recommendations based on global experiences.[11]

Some *downside* considerations are:

- *Possible decreased flexibility* in development or implementation. Standards can, at times, act as a constraint when they are tied to older technologies or methods, which can reduce innovation.
- *"Standards confusion"* from competing and overlapping standards. For instance, an ISO standard may be theory-based and use different terminology, whereas regional or national standards are more specific, applicable, and understandable than broad international ones.
- *Real-world shortcomings due to theoretical basis*. Standards often are guides based on theory rather than practice.
- *Changing and updating requires cost and maintenance*. There are costs to developing, maintaining, and publishing standards.[12]

Key Standards Relevant to IG Efforts

Next we introduce and discuss some established standards that should be researched and considered as a foundation for developing IG policy.

Risk Management

ISO 31000:2009 is a broad, industry-agnostic (not specific to vertical markets) risk management standard. It states "principles and generic guidelines" of risk management that can be applied to not only IG but also to a wide range of organizational activities and processes throughout the life of an organization.[13] It provides a structured framework within which to develop and implement risk management strategies and programs. ISO 31000 defines a **risk management framework** as a set of two basic components that "support and sustain risk management throughout an organization."[14] The stated components are: foundations, which are high level and include risk management policy, objectives, and executive edicts; and organizational arrangements, which are more specific and actionable including strategic plans, roles and responsibilities, allocated budget, and business processes that are directed toward managing an organization's risk.

Additional risk management standards may be relevant to your organization's IG policy development efforts, depending on your focus, scope, corporate culture, and demands of your IG program executive sponsor.

ISO/TR 18128:2014 is a risk assessment standard for records processes and systems.[15] It can be used to assist organizations in assessing risks to records processes and systems so they can ensure records continue to meet identified business needs as long as required. It presents a method for analyzing and documenting risks related to records processes and their effects. The standard can be used to guide records risk assessments in all types and sizes of organizations.

ISO 31000 is a broad risk management standard that applies to all types of businesses. ISO 18128 is a risk assessment standard for records processes.

Information Security and Governance

ISO/IEC 27001:2013 is an information security management system (ISMS standard that provides guidance in the development of security controls to safeguard information assets. Like ISO 31000, the standard is applicable to all types of organizations, irrespective of vertical industry.[16] It "specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving a documented information security management system within the context of the organization's overall business risks." *ISO/IEC 27001* is flexible enough to be applied to a variety of activities and processes when evaluating and managing information security risks, requirements, and objectives, and compliance with applicable legal and regulatory requirements. *This includes use of the standards guidance by internal and external auditors as well as internal and external stakeholders (including customers and potential customers*.

ISO/IEC 27002:2013, "Information Technology—Security Techniques—Code of Practice for Information Security"[17]:

> establishes guidelines and general principles for initiating, implementing, maintaining, and improving information security management in an organization and is identical to the previous published standard, ISO 17799. The objectives outlined provide general guidance on the commonly accepted goals of information security management. ISO/IEC 27002:2013 contains best practices of control objectives and controls in the following areas of information security management:
>
> - Security policy
> - Organization of information security
> - Asset management
> - Human resources security
> - Physical and environmental security
> - Communications and operations management
> - Access control
> - Information systems acquisition, development, and maintenance
> - Information security incident management
> - Business continuity management
> - Compliance
>
> The control objectives and controls in ISO/IEC 27002:2013 are intended to be implemented to meet the requirements identified by a risk assessment. ISO/IEC 27002:2013 is intended as a common basis and practical guideline for developing organizational security standards and effective security management practices, and to help build confidence in interorganizational activities.

ISO/IEC 27001 and ISO/IEC 27002 are information security management systems (not a common acronym) standards that provide guidance in the development of security controls.

ISO/IEC 38500:2008 is an international standard that provides high-level principles and guidance for senior executives and directors, and those advising them, for the effective and efficient use of IT.[18] Based primarily on AS8015, the Australian IT governance standard "applies to the governance of management processes" that are performed at the IT service level, but the guidance assists executives in monitoring IT and ethically discharging their duties with respect to legal and regulatory compliance of IT activities.

The ISO 38500 standard comprises three main sections:

1. Scope, Application and Objectives
2. Framework for Good Corporate Governance of IT
3. Guidance for Corporate Governance of IT

It is largely derived from AS 8015, the guiding principles of which were:

- Establish responsibilities
- Plan to best support the organization
- Acquire validly
- Ensure performance when required
- Ensure conformance with rules
- Ensure respect for human factors

The standard also has relationships with other major ISO standards, and embraces the same methods and approaches. It is certain to have a major impact upon the IT governance landscape.[19]

ISO 38500 is an international standard that provides high-level principles and guidance for senior executives and directors responsible for IT governance.

Records and E-Records Management

ISO 15489–1:2016 is the international standard for RM. It identifies the elements of RM and provides a framework and high-level overview of RM core principles. RM is defined as the "field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use, and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records."[20]

The second part of the standard, ISO 15489–2:2001, contains the technical specifications and a methodology for implementing the standard, originally based on early standards work in Australia (**Design and Implementation of Recordkeeping Systems—DIRKS**). (Note: Although still actively used in Australian states, the National Archives of Australia has not recommended use of DIRKS by Australian national agencies since 2007 and has removed DIRKS from its website.[21])

The ISO 15489 standard makes little mention of electronic records, as it is written to address all kinds of records; nonetheless it was widely viewed as the definitive framework of what RM means.

ISO 15489 is the international RM standard.

In 2008, the International Council on Archives (ICA) formed a multinational team of experts to develop "Principles and Functional Requirements for Records in Electronic Office Environments," commonly referred to as ICA-Req.[22] The project was cosponsored by the Australasian Digital Recordkeeping Initiative (ADRI), which was undertaken by the Council of Australasian Archives and Records Authorities, which "comprises the heads of the government archives authorities of the Commonwealth of Australia, New Zealand, and each of the Australian States and Territories."[23] The National Archives of Australia presented a training and guidance manual to assist in implementing the principles at the 2012 International Congress on Archives Congress in Brisbane, Australia.

In Module 1 of ICA-Req, principles are presented in a high-level overview; Module 2 contains specifications for electronic document and records management systems (EDRMS) that are "globally harmonized"; and Module 3 contains a requirements set and "implementation advice for managing records in business systems."[24] Module 3 recognizes that digital record keeping does not have to be limited to the EDRMS paradigm—the insight that has now been picked up by "Modular Requirements for Records Systems" (MoReq2010, the European standard released in 2011).[25]

Parts 1 to 3 of ISO 16175 were fully adopted in 2010–2011 based on the ICA-Req standard. The standard may be purchased at www.ISO.org, and additional information on the Australian initiative may be found at www.adri.gov.au/.

ISO 16175 is guidance, not a standard that can be tested and certified against. This is the criticism by advocates of testable, certifiable standards like U.S. DoD 5015.2 and the European standard, MoReq2010 for e-records management.

> The ICA-Req standard was adopted as ISO 16175. It does not contain a testing regime for certification.

In November 2011, ISO issued new standards for ERM, the first two in the ISO 30300 series, which are based on a *managerial* point of view and targeted at a management-level audience rather than at records managers or technical staff:

- *ISO 30300:2011*, "Information and Documentation—Management Systems for Records—Fundamentals and Vocabulary"
- *ISO 30301:2011*, "Information and Documentation—Management Systems for Records—Requirements"

The standards apply to **management systems for records** (MSR), a term that, as of this printing, is not typically used to refer to ERM or RM application [RMA] software in the United States or Europe and is not commonly found in ERM research or literature.

The ISO 30300 series is a systematic approach to the creation and management of records that is "*aligned with organizational objectives and strategies*."[26]

ISO 30300 MSR, Fundamentals and Vocabulary, explains the rationale behind the creation of an MSR and the guiding principles for its successful implementation. It provides the terminology that ensures that it is compatible with other management systems standards.

ISO 30301 MSR, Requirements, specifies the requirements necessary to develop a records policy. It also sets objectives and targets for an organization to implement systemic improvements. This is achieved through designing records processes and systems; estimating the appropriate allocation of resources; and establishing benchmarks to monitor, measure, and evaluate outcomes. These steps help to ensure that corrective action can be taken and continuous improvements are built into the system in order to support an organization in achieving its mandate, mission, strategy, and goals.[27]

Additional standards related to digital records include:

- *ISO 13008:2012*, Information and documentation, Digital records conversion and migration process;
- *ISO/TR 13028:2010*, Information and documentation, Implementation guidelines for digitization of records.

Major National and Regional ERM Standards

For great detail on national and regional standards related to ERM, see the book *Managing Electronic Records: Methods, Best Practices, and Technologies* (John Wiley & Sons, 2013) by Robert F. Smallwood. Following is a short summary.

United States E-Records Standard

The US Department of Defense 5015.2, *Design Criteria Standard for Electronic Records Management Software Applications*, standard was established in 1997 and was endorsed by the leading archival authority, the US National Archives and Records Administration (NARA), in the past. A dated standard, reportedly being updated, it no longer has the impact that it did a decade ago. In fact, NARA did not require adherence to it in a major RM software RFP in the 2015–2016 timeframe. The DoD doesn't even adhere to it. It requires a central repository, which is a dated approach (managing records in-place has become *de rigueur*), and it has little relevance to RM requirements in other market sectors such as health care and finance, whereas the European ERM standard MoReq2010 does.

There is a testing regime that certifies software vendors on DoD 5015.2 that is administered by JITC. JITC "builds test case procedures, writes detailed and summary final reports on 5015.2-certified products, and performs on-site inspection of software."[28] The DoD standard was built for the defense sector, and logically

"reflects its government and archives roots."

> The US DoD 5015.2-STD has been the most influential worldwide since it was first introduced in 1997, although its influence has faded. It best suits military applications.

Since its endorsement by NARA, the standard has been the key requirement for ERM system vendors to meet, not only in US public sector bids, but also in the commercial sector.

The 5015.2 standard has since been updated and expanded, in 2002 and 2007, to include requirements for metadata, e-signatures, and Privacy and Freedom of Information Act requirements, and was scheduled for update by 2013, although that process did not begin until the 2017–2018 timeframe.

> The 5015.2 standard has been updated to include specifications such as those for e-signatures and FOI requirements. Yet as of 2019 it was still sorely out of date.

Canadian Standards and Legal Considerations for Electronic Records Management*

The National Standards of Canada for electronic records management are: (1) *Electronic Records as Documentary Evidence* CAN/CGSB-72.34–2017 ("72.34"), published in March 2017. Updates include:

- A section on new technologies that incorporates risk assessment and provides guidance on cloud computing, social media, and mobile devices
- Informative annexes on sources, metadata, preservation formats, and new technologies
- An information technology (IT) system management guide that details key aspects of backup and system recovery, security and protection, transmission, and audit trails in the context of electronic records as evidence
- The electronic image clauses formerly contained in the standard CAN/CGSB-72.11-1993 Microfilm and Electronic Images as Documentary Evidence (that is, Part III and Part IV Section 3).[29]

The Canada Revenue Agency has adopted these standards as applicable to records concerning taxation.[30]

Standard 72.34 deals with these topics: (1) management authorization and accountability; (2) documentation of procedures used to manage records; (3) "reliability testing" of electronic records according to existing legal rules; (4) the procedures manual and the chief records officer; (5) readiness to produce (the "prime directive"); (6) records recorded and stored in accordance with "the usual and ordinary course of business" and "system integrity," being key phrases from the Evidence Acts in Canada; (7) retention and disposal of electronic records; (8) backup and records system recovery; and (9) security and protection. From these standards practitioners have derived many specific tests for auditing, establishing, and revising electronic records management systems.[31]

The "prime directive" of these standards states: "An organization shall always be prepared to produce its records as evidence."[32] *The duty to establish the "prime directive" falls upon senior management*:[33]

Standard 5.4.3 Senior management, the organization's own internal law-making authority, proclaims throughout the organization the integrity of the organization's records system (and, therefore, the integrity of its electronic records) by establishing and declaring:

a. The system's role in the usual and ordinary course of business;
b. The circumstances under which its records are made; and
c. Its prime directive for all RMS [records management system] purposes, i.e. an organization shall always be prepared to produce its records as evidence. This dominant principle applies to all of the organization's business records, including electronic, optical, original paper source records, microfilm, and other records of equivalent form and content.

Being the "dominant principle" of an organization's electronic records management system, the duty to maintain compliance with the "prime directive" should fall upon its senior management.

Legal Considerations

Because an electronic record is completely dependent upon its ERM system for everything, compliance with these National Standards and their "prime directive" should be part of the determination of the "admissibility" (acceptability) of evidence and of electronic discovery in court proceedings (litigation) and in regulatory tribunal proceedings.[34]

There are 14 legal jurisdictions in Canada: 10 provinces, 3 territories, and the federal jurisdiction of the Government of Canada. Each has an Evidence Act (the Civil Code in the province of Quebec[35]), which applies to legal proceedings within its legislative jurisdiction. For example, criminal law and patents and copyrights are within federal legislative jurisdiction, and most civil litigation comes within provincial legislative jurisdiction.[36]

The admissibility of records as evidence is determined under the "business record" provisions of the Evidence Acts.[37] They require proof that a record was made "in the usual and ordinary course of business," and of "the circumstances of the making of the record." In addition, to obtain admissibility for electronic records, most of the Evidence Acts contain electronic record provisions, which state that an electronic record is admissible as evidence on proof of the "integrity of the electronic record system in which the data was recorded or stored."[38] This is the "system integrity" test for the admissibility of electronic records. The word "integrity" has yet to be defined by the courts.[39]

However, by way of sections such as the following, the electronic record provisions of the Evidence Acts make reference to the use of standards such as the National Standards of Canada:

> For the purpose of determining under any rule of law whether an electronic record is admissible, evidence may be presented in respect of any standard, procedure, usage or practice on how electronic records are to be recorded or stored, having regard to the type of business or endeavour that used, recorded, or stored the electronic record and the nature and purpose of the electronic record.[40]

UK and European Standards

In the United Kingdom, The National Archives (TNA) (formerly the Public Record Office, or PRO) "has published two sets of functional requirements to promote the development of the electronic records management software market (1999 and 2002)." It ran a program to evaluate products against the 2002 requirements.[41] Initially these requirements were established in collaboration with the central government, and they later were utilized by the public sector in general, and also in other nations. The National Archives 2002 requirements remain somewhat relevant, although no additional development has been underway for years. It is clear that the second version of Model Requirements for Management of Electronic Records, MoReq2, largely supplanted the UK standard, and subsequently the newer MoReq2010 further supplants the UK standard.

MoReq2010

MoReq2010 "unbundled" some of the core requirements in MoReq2, and sets out functional requirements in modules. The approach seeks to permit the later creation of e-records software standards in various vertical industries, such as defense, health care, financial services, and legal services.

MoReq2010 is available free—all 500-plus pages of it (by comparison, the U.S. DoD 5015.2 standard is less than 120 pages long). For more information on MoReq2010, visit https://www.moreq.info/. The entire specification may be downloaded from https://www.moreq.info/specification.

In November 2010, the DLM Forum, a European Commission supported body, announced the availability of the final draft of the MoReq2010 specification for electronic records management systems (ERMS), following extensive public consultation. The final specification was published in mid-2011.[42]

The DLM Forum explains that "With the growing demand for [electronic] records management, across a broad spectrum of commercial, not-for-profit, and government organizations, MoReq2010 provides the first practical specification against which all organizations can take control of their corporate information. IT software and services vendors are also able to have their products tested and certified that they meet the MoReq2010 specification."[43]

MoReq2010 supersedes its predecessor MoReq2 and has the continued support and backing of the European Commission.

Australian ERM and Records Management Standards

Australia has adopted all three parts of ISO 16175 as its e-records management standard.[44] (For more detail on this standard, go to ISO.org.)

Australia has long led the introduction of highly automated electronic document management systems and records management standards. Following the approval and release of the AS 4390 standard in 1996, the international records management community began work on the development of an International standard. This work used AS 4390–1996 Records Management as its starting point.

Development of Australian Records Standards

In 2002 Standards Australia published a new Australian Standard on records management, AS ISO 15489, based on the ISO 15489 international records management standard. It differs only in its preface verbiage.[45] AS ISO 15489 carries through all these main components of AS 4390, but internationalizes the concepts and brings them up to date. The standards thereby codify Australian best practice but are also progressive in their recommendations.

Additional Relevant Australian Standards

The Australian Government Recordkeeping Metadata Standard Version 2.0 provides guidance on metadata elements and sub-elements for records management. It is a baseline tool that "describes information about records and the context in which they are captured and used in Australian Government agencies." This standard is intended to help Australian agencies "meet business, accountability and archival requirements in a systematic and consistent way by maintaining reliable, meaningful and accessible records." The standard is written in two parts, the first describing its purpose and features and the second outlining the specific metadata elements and subelements.[46]

The Australian Government Locator Service, AGLS, is published as AS 5044–2010, the metadata standard to help find and exchange information online. It updates the 2002 version and includes changes made by the Dublin Core Metadata Initiative (DCMI).

Another standard, AS 5090:2003, "Work Process Analysis for Recordkeeping," complements AS ISO 15489, and provides guidance on understanding business processes and workflow, so that recordkeeping requirements may be determined.[47]

> The ISO 30300 series of e-records standards are written for a managerial audience and encourage ERM that is aligned to organizational objectives.

Long-Term Digital Preservation

Although many organizations shuffle dealing with digital preservation issues to the back burner, **long-term digital preservation** (LTDP) is a key area in which IG policy should be applied. LTDP methods, best practices, and standards should be applied to preserve an organization's historical and **vital records (**those without which it cannot operate or restart operations) and to maintain its corporate or organizational memory. The key standards that apply to LTDP are listed next.

The official standard format for preserving electronic documents is PDF/A-1, based on PDF 1.4, originally developed by Adobe. *ISO 19005–1:2005,* "Document Management—Electronic Document File Format for Long-Term Preservation—Part 1: Use of PDF 1.4 (PDF/A-1)," is the published specification for using PDF 1.4 for LTDP, which is applicable to e-documents that may contain not only text characters but also graphics (either raster or vector).[48]

> LTDP is a key area to which IG policy should be applied.

ISO 14721:2012, "Space Data and Information Transfer Systems—Open Archival Information Systems—Reference Model (OAIS)," is applicable to LTDP.[49] ISO 14271 "specifies a reference model for an open archival information system (OAIS). The purpose of ISO 14721 is to establish a system for archiving information, both digitalized and physical, with an organizational scheme composed of people who accept the responsibility to preserve information and make it available to a designated community."[50] The fragility of digital storage media combined with ongoing and sometimes rapid changes in computer software and hardware poses a

fundamental challenge to ensuring access to trustworthy and reliable digital content over time. Eventually, every digital repository committed to long-term preservation of digital content must have a strategy to mitigate computer technology obsolescence. Toward this end, the Consultative Committee for Space Data Systems developed the OAIS reference model to support formal standards for the long-term preservation of space science data and information assets. OAIS was not designed as an implementation model.

OAIS is the lingua franca of digital preservation as the international digital preservation community has embraced it as the framework for viable and technologically sustainable digital preservation repositories. *An LTDP strategy that is OAIS compliant offers the best means available today for preserving the digital heritage of all organizations, private and public.* (See Chapter 17.)

ISO TR 18492 (2005), "Long-Term Preservation of Electronic Document Based Information," provides practical methodological guidance for the long-term preservation and retrieval of authentic electronic document-based information, when the retention period exceeds the expected life of the technology (hardware and software) used to create and maintain the information assets. ISO 18492 takes note of the role of ISO 15489 but does not cover processes for the capture, classification, and disposition of authentic electronic document-based information.

> An LTDP strategy that is OAIS compliant (based on ISO 14721) offers the best means available today for preserving the digital heritage of all organizations.

ISO 16363:2012, "Space Data and Information Transfer Systems—Audit and Certification of Trustworthy Digital Repositories," "defines a recommended practice for assessing the trustworthiness of digital repositories. It is applicable to the entire range of digital repositories."[51] It is an audit and certification standard organized into three broad categories: Organization Infrastructure, Digital Object Management, and Technical Infrastructure and Security Risk Management. *ISO 16363 represents the gold standard of audit and certification for trustworthy digital repositories.* (See Chapter 17.)

Of note is that newer cloud-based approaches for digital preservation greatly simplify the process of adhering to technology-neutral standards and maintaining digital records in the cloud. This new breed of services supplier provides everything from document conversion to ongoing maintenance. Five or six copies of the records are stored in different parts of the globe using major cloud providers like Microsoft and Amazon. A checksum algorithm is used to periodically scan the stored digital records for any degradation or corruption, which can then be corrected using the undamaged copies.

> ISO 16363 represents the gold standard of audit and certification for trustworthy digital repositories.

Business Continuity Management

ISO 22301:2012, "Societal Security—Business Continuity Management Systems—Requirements," spells out the requirements for creating and implementing a standardized approach to business continuity management (BCM, also known as disaster recovery [DR]), in the event an organization is hit with a disaster or major business interruption.[52] The guidelines can be applied to any organization regardless of vertical industry or size. The specification includes the "requirements to plan, establish, implement, operate, monitor, review, maintain and continually improve a documented management system to protect against, reduce the likelihood of occurrence, prepare for, respond to, and recover from disruptive incidents when they arise."

> ISO 22301 spells out requirements for creating and implementing a standardized approach to BCM.

The UK business continuity standard, BS25999-2, which heavily influenced the newer ISO standard, was withdrawn when ISO 22301 was released.[53] The business rationale is that, with the increasing globalization of business, ISO 22301 will allow and support more consistency worldwide not only in business continuity planning and practices but will also promote common terms and help to embed various ISO management systems standards within organizations. U.S.-based ANSI, Standards Australia, Standards Singapore, and other standards bodies also contributed to the development of ISO 22301.

Benefits of ISO 22301

- *Threat identification and assessment*. Discover, name, and evaluate potential serious threats to the viability of the business.
- *Threat and recovery planning*. Ensuring that the impact and resultant downtime and recovery from real threats that do become incidents is minimized.
- *Mission-critical process protection*. Identifying key processes and taking steps to ensure they continue to operate even during a business interruption.
- *Stakeholder confidence*. Shows prudent management planning and business resilience to internal and external stakeholders, including employees, business units, customers, and suppliers.[54]

Making Your Best Practices and Standards Selections to Inform Your IG Framework

You must take into account your organization's corporate culture, appetite for risk, management style, and organizational goals when determining which best practices and standards should receive priority in your IG framework. However, you must step through your business rationale in discussions with your cross-functional IG team and fully document the reasons for your approach. Then you must present this approach and your draft IG framework to your key stakeholders and be able to defend your determinations while allowing for input and adjustments. Perhaps you have overlooked some key factors that your larger stakeholder group uncovers, and their input should be folded into a final draft of your IG framework.

You must take into account your organization's corporate culture, management style, and organizational goals when determining which best practices and standards should be selected for your IG framework.

Next, you are ready to begin developing IG policies that apply to various aspects of information use and management in specific terms. You must detail the policies you expect employees to follow when handling information on various information delivery platforms (e.g. e-mail, blogs, social media, mobile computing, cloud computing).

It is helpful at this stage to collect and review all your current policies that apply and to gather some examples of published IG policies, particularly from peer organizations and competitors (where possible). Of note: *You should not just adopt another organization's polices* and believe that you are done with policy making. Rather, you must enter into a deliberative process, using your IG framework for guiding principles and considering the views and needs of your cross-functional IG team. Of paramount importance is to be sure to incorporate the alignment of your organizational goals and business objectives when crafting policy.

With each policy area, be sure that you have considered the input of your stakeholders, so that they will be more willing to comply with the new policies and so that the policies do not run counter to their business needs and required business processes. Otherwise, stakeholders will skirt, avoid, or halfheartedly follow the new IG policies, and the IG program risks failure.

Once you have finalized your policies, be sure to obtain necessary approvals from your executive sponsor and key senior managers.

Roles and Responsibilities

Policies will do nothing without people to advocate, support, and enforce them. So *clear lines of authority and accountability must be drawn*, and responsibilities must be assigned.

You may find it useful to develp a **responsibility assignment matrix**, also known as a **RACI matrix**, which delineates the parties who are responsible, accountable, consulted, and informed.

Overall IG program responsibility resides at the executive sponsor level, but beneath that, an IG program manager or program lead—perhaps even a formal Chief Information Governance Officer (CIGO)—should drive team members toward milestones and business objectives and should shoulder the responsibility for day-to-day program activities, including implementing and monitoring key IG policy tasks. These tasks should be approved by executive stakeholders and assigned as appropriate to an employee's functional area of expertise. For instance, the IG team member from legal may be assigned the responsibility for researching and

determining legal requirements for retention of business records, perhaps working in conjunction with the IG team member from RM, who can provide additional input based on interviews with representatives from business units and additional RM research into best practices. However, it is important that the IG program team is cross-trained to improve communications and effectiveness. Essentially, key stakeholders must be able to understand the various viewpoints and "speak each other's language."

> Lines of authority, accountability, and responsibility must be clearly drawn for the IG program to succeed. A RACI matrix can be a useful tool.

Program Communications and Training

Your IG program must contain a communications and training component, as a standard function. This is critical, as IG programs require a **change management** component. Your stakeholder audience must be made aware of the new policies and practices that are to be followed and how this new approach contributes toward the organization's goals and business objectives. Further, key concepts must be reinforced continually to drive cultural change at the core of the organization.

The first step in your communications plan is to identify and segment your stakeholder audiences and to customize or modify your message to the degree that is necessary to be effective. Communications to your IT team can have a more technical slant, and communications to your legal team can have some legal jargon and emphasize legal issues. The more forethought you put into crafting your communications strategy, the more effective it will be.

That is not to say that *all* messages must have several versions: some key concepts and goals should be emphasized in communications to all employees.

> Communications regarding your IG program should be consistent and clear and somewhat customized for various stakeholder groups.

How should you communicate? *The more ways you can get your IG message to your core stakeholder audiences, the more effective and lasting the message will be.* So posters, newsletters, e-mail, text messages, internal blog or intranet posts, and company meetings should all be a part of the communications mix. You can even make it fun, perhaps creating an IG program mascot, or gamifying IG training to encourage healthy competition.

Remember, the IG program requires not only training but *re*training, and the aim should be to create a compliance culture that is so prominent and expected that employees adopt the new practices and policies and integrate them into their daily activities. Ideally, employees will provide valuable input to help fine-tune and improve the IG program.

Training should take multiple avenues as well. Some can be classroom instruction, some online learning, and you may want to create a series of training videos. You may also want to deploy a privacy awareness training (PAT) or security awareness training (SAT) series, which is an effective way to reduce information risk on an ongoing basis. But the training effort must be consistent and ongoing to maintain high levels of IG effectiveness. Certainly, this means you will need to add to your new hire onboarding program for employees joining or transferring to your organization.

Program Controls, Monitoring, Auditing, and Enforcement

How do you know how well you are doing? You will need to develop metrics to determine the level of employee compliance, its impact on key operational areas, and progress made toward established business objectives. *Relevant and valid metrics can only be developed through stakeholder consultation*.

Testing and auditing the program provides an opportunity to give feedback to employees on how well they are doing and to recommend changes they may make. But having objective feedback on key metrics also will allow for your executive sponsor to see where progress has been made and where improvements need to focus.

Eventually, clear penalties for policy violations must be communicated to employees so they know the seriousness of the IG program and how important it is in helping the organization pursue its business goals

and accomplish stated business objectives.

CHAPTER SUMMARY: KEY POINTS

- You must inform and frame IG policy with internal and external frameworks, models, best practices, and standards.
- The business user is the primary stakeholder of managed information.
- Information management is important at all stages of the life cycle.
- Legal stakeholders usually can mandate the preservation of what is most critical, though often at great cost.
- The IGRM was developed by the EDRM Project to foster communication among stakeholders and adoption of IG. It complements ARMA's GAR Principles.
- ISO 31000 is a broad risk management standard that applies to all types of businesses. ISO 18128 is a risk assessment standard for records processes.
- ISO/IEC 27001 and ISO/IEC 27002 are ISMS standards that provide guidance in the development of security controls.
- ISO 15489 is the international RM standard.
- The ICA-Req standard was adopted as ISO 16175. It does not contain a testing regime for certification.
- The ISO 30300 series of e-records standards is written for a managerial audience and to encourage adherence to ERM that is aligned to organizational objectives.
- DoD 5015.2 is the US ERM standard although its use has waned; the European ERM standards is MoReq2010. Australia has adopted all three parts of ISO 16175 as its e-records management standard.
- LTDP is a key area to which IG policy should be applied.
- An LTDP strategy that is OAIS compliant (based on ISO 14721) offers the best means available today for preserving the digital heritage of all organizations.
- ISO 16363 represents the gold standard of audit and certification for trustworthy digital repositories.
- ISO 38500 is an international standard that provides high-level principles and guidance for senior executives and directors responsible for IT governance.
- ISO 22301 spells out requirements for creating and implementing a standardized approach to business continuity management.
- You must take into account your organization's corporate culture, management style, and organizational goals when determining which best practices and standards should be selected for your IG framework.
- Lines of authority, accountability, and responsibility must be clearly drawn for the IG program to succeed. An RACI matrix can be a useful tool.
- Communications regarding your IG program should be consistent and clear and somewhat customized for various stakeholder groups.
- IG program audits are an opportunity to improve training and compliance, not to punish employees.

Notes

1. TechTarget.com, "Generally Accepted Recordkeeping Principles," https://searchcompliance.techtarget.com/definition/Generally-Accepted-Recordkeeping-Principles (accessed December 3, 2018).
2. ARMA International, "Information Governance Maturity Model," https://www.arma.org/page/IGMaturityModel (accessed December 3, 2018.

3. Electronic Discovery, "IGRM v3.0 Update: Privacy & Security Officers As Stakeholders – Electronic Discovery," https://www.edrm.net/frameworks-and-standards/information-governance-reference-model/white-paper/ (accessed December 3, 2018).

4. EDRM, "Information Governance Reference Model (IGRM)," https://www.edrm.net/papers/igrm-it-viewpoint/ (accessed December 3, 2018).

5. Ibid.

6. CGOC.com, "CGOC: Information Governance Process Maturity Model," https://www.cgoc.com/resource/information-governance-process-maturity-model/ (accessed December 3, 2018).

7. https://www.infogovbasics.com/best-practices/by-industry/healthcare/ (accessed February 7, 2018).

8. Monica Crocker, e-mail to author, June 21, 2012.

9. Project Management Institute, *A Guide to the Project Management Body of Knowledge* (*PMBOK Guide*), 4th ed. (Newtown Square, PA, Project Management Institute, 2008), ANSI/PMI 99-001-2008, pp. 273–312.

10. Marc Fresko, e-mail to author, May 13, 2012.

11. Hofman, "The Use of Standards and Models," in Julie McLeod and Catherine Hare, eds., *Managing Electronic Records*, p. 34 and pp. 20–21 (London: Facet, 2005)

12. Ibid.

13. International Organization for Standardization, "ISO 31000:2009 Risk Management—Principles and Guidelines," www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43170 (accessed April 22, 2013).

14. Ibid.

15. International Organization for Standardization, ISO/TR 18128:2014, "Risk Assessment for Records Processes," https://www.iso.org/standard/61521.html (accessed December 3, 2018).

16. International Organization for Standardization, ISO/IEC 27001:2013, "Information Technology—Security Techniques—Information Security Management Systems—Requirements," https://www.iso.org/standard/54534.html (accessed December 3, 2018).

17. International Organization for Standardization, ISO/IEC 27002:2013, "Information Technology—Security Techniques—Code of Practice for Information Security Management," https://www.iso.org/standard/54533.html (accessed December 3, 2018).

18. International Organization for Standardization, ISO/IEC 38500:2008, www.iso.org/iso/catalogue_detail?csnumber=51639 (accessed March 12, 2013).

19. ISO 38500 IT Governance Standard, www.38500.org/ (accessed March 12, 2013).

20. International Organization for Standardization, *ISO 15489-1: 2001 Information and Documentation—Records Management. Part 1: General* (Geneva: ISO, 2001), section 3.16.

21. National Archives of Australia, www.naa.gov.au/records-management/publications/DIRKS-manual.aspx (accessed October 15, 2012).

22. International Council on Archives, "ICA-Req: Principles and Functional Requirements for Records in Electronic Office Environments: Guidelines and Training Material," November 29, 2011, www.ica.org/11696/activities-and-projects/icareq-principles-and-functional-requirements-for-records-in-electronic-office-environments-guidelines-and-training-material.html.

23. Council of Australasian Archives and Records Authorities, www.caara.org.au/ (accessed May 3, 2012).

24. Adrian Cunningham, blog post comment, May 11, 2011. http://thinkingrecords.co.uk/2011/05/06/how-moreq-2010-differs-from-previous-electronic-records-management-erm-system-specifications/.

25. Ibid.

26. "Relationship between the ISO 30300 Series of Standards and Other Products of ISO/TC 46/SC 11: Records Processes and Controls," White Paper, ISO TC46/SC11- Archives/Records Management (March 2012), www.iso30300.es/wp-content/uploads/2012/03/ISOTC46SC11_White_paper_relationship_30300_technical_standards12032012v6.pdf

27. Ibid.

28. Julie Gable, *Information Management Journal,* November 1, 2002, www.thefreelibrary.com/Everything+you+wanted+to+know+about+DoD+5015.2:+the+standard+is+not+a...-a095630076.

29. "Electronic Records as Documentary Evidence" www.publications.gc.ca/site/eng/9.839939/publication.html (accessed December 4, 2018).

30. The Canada Revenue Agency (CRA) informs the public of its policies and procedures by means, among others, of its *Information Circulars* (IC's), and *GST/HST Memoranda* (GST: goods and services tax; HST: harmonized sales tax, that is, the harmonization of federal and provincial sales taxes into one retail sales tax). In particular, see: *IC05-1*, dated June 2010, entitled, *Electronic Record Keeping*, paragraphs 24, 26, and 28. Note that use of the National Standard cited in paragraph 26, *Microfilm and Electronic Images as Documentary Evidence* CAN/CGSB-72.11-93 is mandatory for, "Imaging and microfilm (including microfiche) reproductions of books of original entry and source documents ..." Paragraph 24 recommends the use of the newer national standard, *Electronic Records as Documentary Evidence* CAN/CGSB-72.34-2005, "To ensure the reliability, integrity and authenticity of electronic records." However, if this newer standard is given the same treatment by CRA as the older standard, it will be made mandatory as well. And similar statements appear in the GST Memoranda, *Computerized Records* 500-1-2, *Books and Records* 500-1. IC05-1. *Electronic Record Keeping*, concludes with the note, "Most Canada Revenue Agency publications are available on the CRA website www.cra.gc.ca under the heading 'Forms and Publications.'" https://www.tpsgc-pwgsc.gc.ca/ongc-cgsb/programme-program/normes-standards/can-cgsb-72-34-eng.html (accessed December 4, 2018).

31. There are more than 200 specific compliance tests that can be applied to determine if the principles of 72.34 are being complied with. The analysts—a combined team of records management and legal expertise—analyze: (1) the nature of the business involved; (2) the uses and value of its records for its various functions; (3) the likelihood and risk of the various types of its records being the subject of legal proceedings, or of their being challenged by some regulating authority; and (4) the consequences of the unavailability of acceptable records—for example, the consequences of its records not being accepted in legal proceedings. Similarly, in regard to the older National Standard of Canada, 72.11, there is a comparable series of more than 50 tests that can be applied to determine the state of compliance with its principles.

32. *Electronic Records as Documentary Evidence* CAN/CGSB-72.34-2005 ("72.34"), clause 5.4.3 c) at p. 17; and *Microfilm and Electronic Images as Documentary Evidence* CAN/CGSB-72.11-93 ("72.11"), paragraph 4.1.2 at p. 2, *supra* note 49.

33. 72.34, Clause 5.4.3, ibid.

34. "Admissibility" refers to the procedure by which a presiding judge determines if a record or other proffered evidence is acceptable as evidence according the rules of evidence. "Electronic discovery" is the compulsory exchange of relevant records by the parties to legal proceedings prior to trial. As to the admissibility of records as evidence see: Ken Chasse, "The Admissibility of Electronic Business Records" (2010), 8 *Canadian Journal of Law and Technology* 105; and Ken Chasse, "Electronic Records for Evidence and Disclosure and Discovery" (2011) 57 The Criminal Law Quarterly 284. For the electronic discovery of records see: Ken Chasse, "Electronic Discovery—*Sedona Canada* is Inadequate on Records Management—Here's *Sedona Canada* in Amended Form," *Canadian Journal of Law and Technology* 9 (2011): 135; and Ken Chasse, "Electronic Discovery in the Criminal Court System," *Canadian Criminal Law Review* 14 (2010): 111. See also note 18 *infra*, and accompanying text.

35. For the province of Quebec, comparable provisions are contained in Articles 2831–2842, 2859–2862, 2869–2874 of Book 7 "Evidence" of the Civil Code of Quebec, S.Q. 1991, c. C-64, to be read in conjunction with, An Act to Establish a Legal Framework for Information Technology, R.S.Q. 2001, c. C-1.1, ss. 2, 5–8, and 68.

36. For the legislative jurisdiction of the federal and provincial governments in Canada, see The Constitution

Act, 1867 (UK) 30 & 31 Victoria, c. 3, s. 91 (federal), and s. 92 (provincial), www.canlii.org/en/ca/laws/stat/30—31-vict-c-3/latest/30—31-vict-c-3.html.

37. The two provinces of Alberta and Newfoundland and Labrador do not have business record provisions in their Evidence Acts. Therefore "admissibility" would be determined in those jurisdictions by way of the court decisions that define the applicable common law rules; such decisions as *Ares v. Venner* [1970] S.C.R. 608, 14 D.L.R. (3d) 4 (S.C.C.), and decisions that have applied it.

38. See for example, the Canada Evidence Act, R.S.C. 1985, c. C-5, ss. 31.1-31.8; Alberta Evidence Act, R.S.A. 2000, c. A-18, ss. 41.1-41.8; (Ontario) Evidence Act, R.S.O. 1990, c. E.23, s. 34.1; and the (Nova Scotia) Evidence Act, R.S.N.S. 1989, c. 154, ss. 23A-23G. The Evidence Acts of the two provinces of British Columbia and Newfoundland and Labrador do not contain electronic record provisions. However, because an electronic record is no better than the quality of the record system in which it is recorded or stored, its "integrity" (reliability, credibility) will have to be determined under the other provincial laws that determine the admissibility of records as evidence.

39. The electronic record provisions have been in the Evidence Acts in Canada since 2000. They have been applied to admit electronic records into evidence, but they have not yet received any detailed analysis by the courts.

40. This is the wording used in, for example, s. 41.6 of the Alberta Evidence Act, s. 34.1(8) of the (Ontario) Evidence Act; and s. 23F of the (Nova Scotia) Evidence Act, *supra* note 10. Section 31.5 of the Canada Evidence Act, *supra* note 58, uses the same wording, the only significant difference being that the word "document" is used instead of "record." For the province of Quebec, see sections 12 and 68 of, An Act to Establish a Legal Framework for Information Technology, R.S.Q., chapter C-1.1.

41. "Giving Value: Funding Priorities for UK Archives 2005–2010, a key new report launched by the National Council on Archives (NCA) in November 2005," www.nationalarchives.gov.uk/documents/standards_guidance.pdf (accessed October 15, 2012).

42. DLM Forum Foundation, *MoReq2010®: Modular Requirements for Records Systems—Volume 1: Core Services & Plug-in Modules*, 2011, https://www.moreq.info/ (accessed December 4, 2018), published in paper form as ISBN 978-92-79-18519-9 by the Publications Office of the European Communities, Luxembourg.

43. DLM Forum, Information Governance across Europe, www.dlmforum.eu/ (accessed December 4, 2018).

44. National Archives of Australia, "Australian and International Standards," 2018, www.naa.gov.au/information-management/information-governance/legislation-standards/ISO16175/index.aspx (accessed December 4, 2018).

45. Marc Fresko, e-mail to author, May 13, 2012.

46. National Archives of Australia, "Australian Government Recordkeeping Metadata Standard," 2012, www.naa.gov.au/records-management/publications/agrk-metadata-standard.aspx (accessed July 16, 2012).

47. National Archives of Australia, "Australian and International Standards," 2012, www.naa.gov.au/records-management/strategic-information/standards/ASISOstandards.aspx (accessed July 16, 2012).

48. International Organization for Standardization, ISO 19005-1:2005, "Document Management—Electronic Document File Format for Long-Term Preservation—Part 1: Use of PDF 1.4 (PDF/A-1)," www.iso.org/iso/catalogue_detail?csnumber=38920 (accessed July 23, 2012).

49. International Organization for Standardization, ISO 14721:2012, "Space Data and Information Transfer Systems Open Archival Information System—Reference Model," www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=57284 (accessed November 25, 2013).

50. Ibid.

51. International Organization for Standardization, ISO 16363:2012, "Space Data and Information Transfer Systems—Audit and Certification of Trustworthy Digital Repositories," www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510 (accessed July 23, 2012).

52. International Organization for Standardization, ISO 22301:2012 "Societal Security—Business Continuity Management Systems—Requirements," www.iso.org/iso/catalogue_detail?csnumber=50038 (accessed April 21, 2013).

53. International Organization for Standardization, "ISO Business Continuity Standard 22301 to Replace BS 25999-2," www.continuityforum.org/content/news/165318/iso-business-continuity-standard-22301-replace-bs-25999-2 (accessed April 21, 2013).

54. BSI, "ISO 22301 Business Continuity Management," www.bsigroup.com/en-GB/iso-22301-business-continuity (accessed April 21, 2013).

*This section is adapted with permission by EDRM.net, https://www.edrm.net/frameworks-and-standards/information-governance-reference-model (accessed December 3, 2018).

†This section is adapted with permission by EDRM.net, https://www.edrm.net/frameworks-and-standards/information-governance-reference-model/ (accessed December 3, 2018).

*This section was contributed by Ken Chasse J.D., LL.M., a records management attorney and consultant, and member of the Law Society of Upper Canada (Ontario) and of the Law Society of British Columbia, Canada.

PART THREE

Information Governance Key Impact Areas

CHAPTER 7
Information Governance for Business Units

When looking at the Information Governance Reference Model (previously presented), note that there are five key areas of impact: business units, legal, records and information management (RIM), information technology (IT), and privacy and security. In this section of the book we cover how each of these major areas is impacted by, and participates in, an IG program.

Business units generate profit, and this is where IG programs can have their greatest impact. Rather than focusing on "soft cost" justifications such as productivity increases or process improvements, hard-dollar savings and revenue generation can be achieved through successful IG programs targeted in business units.

Supporting functions, including legal, RIM, IT, and privacy and security *must work in a cross-functional, collaborative way* to reduce information risks and costs, leverage information as an asset, and support business units to increase profitability.

Start with Business Objective Alignment

IG program planning begins with developing business objectives for the program itself. Those business objectives must be aligned with and support the accomplishment of overall organizational business objectives. This alignment is key to winning executive support and funding for an IG program. If executives can see a clearly aligned path and that the IG program is synchronized with other organizational initiatives, then they are more likely to back the IG program.

If, for instance, one corporate objective is:

- Grow the company by making acquisitions and folding them into existing operations and systems.

Then we know that existing operations and systems must be standardized so that the acquired companies can be more easily incorporated into operations. To do this, we can form some objectives for the IG program, such as:

1. Improve and standardize document and records search and retrieval functions.
2. Develop or update access rights policies.
3. Train all employees and new employees on new search and access capabilities and restrictions.

Some key metrics that may be used to measure progress may be:

1. Implement a standardized functional taxonomy for classifying documents and records within 180 days.
2. Implement standardized metadata schema for organizing documents and records within 180 days.
3. Implement auto-categorization analytics software by year end.
4. Train employees and newly acquired business units on search tools to improve productivity and access restrictions to improve security by year-end.

These IG program objectives and metrics all align to support the higher-level corporate objectives, and should win executive support, as well as support from key stakeholder groups.

> **IG program objectives must be aligned with organizational business objectives to gain executive and stakeholder support.**

If, for instance, other corporate objectives are:

- Reduce information risk and the likelihood and impact of breaches.
- Preserve and protect brand equity.

Then we know that cybersecurity measures must be implemented and employee training must take place. To do this, we can form some objectives for the IG program, such as:

1. Reduce information risk by implementing new security measures.
2. Reduce information risk by training employees on new and emerging cybersecurity threats.

Some key metrics that may be used to measure progress may be:

1. Conduct an assessment of IT security practices using the CIS Top 20 (more detail is in Chapter 11) within 90 days.
2. Implement new cybersecurity recommendations within 180 days.
3. Train 100 percent of home office knowledge workers with new security awareness training (SAT) program by year-end.

In both of the above examples, the IG program objectives flow up to support the accomplishment of organizational business objectives, and time-delimited metrics are in place to measure progress.

Which Business Units Are the Best Candidates to Pilot an IG Program?

When evaluating possible starting places for an IG program, look for business units that:

1. *Have a high litigation profile*. Since IG programs reduce e-discovery collection and review costs by better organizing

information, and reducing the information footprint by disposing of information that has met its lifecycle requirements, business units that have a high level of litigation can show significant benefits by implementing IG programs.

2. *Have excessive compliance risks and costs*. When employees have difficulty producing documentation in a timely manner for regulators, or cannot find the documentation, and the organization is fined or sanctioned excessively, those business units are good candidates for IG program pilots. IG programs can have a significant impact by improving search and retrieval functions as a result of standardizing taxonomy and metadata approaches.
3. *Have opportunities to monetize or leverage information value*. By applying infonomics principles and formulas, information can be measured and monetized to add new value to the organization.

What Is Infonomics?

Following is an excerpt from Doug Laney's groundbreaking book Infonomics *(Bibliomotion/Taylor & Francis, 2018). Used with permission.*

Infonomics is the theory, study, and discipline of asserting economic significance to information. It provides the framework for businesses to monetize, manage, and measure information as an actual asset. Infonomics endeavors to apply both economic and asset management principles and practices to the valuation, handling, and deployment of information assets.

As a business, information, or information technology (IT) leader, chances are that you regularly talk about information as one of your most valuable assets. Do you value or manage our organization's information like an actual asset? Consider your company's well-honed supply chain and asset management practices for physical assets or your financial management and reporting discipline. Do you have similar accounting and asset management practices in place for your “information assets”? Most organizations do not.

When considering how to put information to work for your organization, it's essential to go beyond thinking and talking about information as an asset to actually valuing and treating it as one. The discipline of infonomics provides organizations with a foundation and methods for quantifying information asset value and formal information asset management practices.

Infonomics posits that information should be considered a new asset class in that it has measurable economic value and other properties that qualify it to be accounted for and administered as any other recognized type of asset—and that there are significant strategic, operational, and financial reasons for doing so. Infonomics provides the framework businesses and governments need to value information, manage it, and wield it as a real asset.

> Infonomics provides the framework for businesses to monetize, manage, and measure information as an actual asset.

How to Begin an IG Program

Your organization needs a starting point, one that is measured and defined so that you can measure and track your progress from that point. So, typically an IG assessment must take place. That assessment will provide information as to where problems lie, and smoke out some issues that might not be readily apparent. An IG assessment will measure, in objective terms, the level of maturity of existing IG efforts, and lay out a roadmap with metrics or **key performance indicators** (KPIs) for IG improvement.

Several assessment methodologies are available. Perhaps the most comprehensive one is the **IG Process Maturity Model** (IGPMM) from the Compliance, Governance and Oversight Council (CGOC). The IGPMM measures the maturity of 22 specific processes for the five key impact areas depicted in the IG Reference Model: lines of business, legal, RIM, IT, and privacy and security. The IGPMM heavily emphasizes the role of legal functions, and privacy and security, and shows that RIM plays a smaller but important role in the overall implementation of IG programs.

The model was created in 2012, and updated in 2017. Processes were revised and new ones added in line with marketplace and technology developments. New processes include data security, including cloud security considerations; data quality; and privacy and data protection processes that consider the impact of the European Union General Data Protection Regulation (GDPR).[1]

> The most IG comprehensive assessment tool is the IG Process Maturity Model from CGOC.

Other assessment tools and methodologies also exist. It is important to note that *all* assessment tools have some inherent bias, depending on the agenda of the organization that created them. For instance, the Information Governance Maturity Model (IGMM) from ARMA International is based on the Generally Accepted Recordkeeping Principles®, so *it is best suited not for holistic IG assessments, but for assessments of the maturity of recordkeeping processes*. It is possible to utilize this model to evaluate RIM functions, while also utilizing the IGPMM from CGOC for broader and more relevant IG processes. Other assessment tools have been developed for specific industries such as healthcare (for details on IG programs in this vertical market, refer to *Information Governance for Healthcare Professionals* (HIMSS/CRC Press, 2018), by Robert F. Smallwood.

Business Considerations for an IG Program

By Barclay T. Blair

The business case for **information governance** (IG) programs has historically been difficult to justify. It is hard to apply a strict, short-term return on investment (ROI) calculation. A lot of time, effort, and expense is involved before true economic benefits can be realized. Therefore, a commitment to the long view and an understanding of the many areas where an organization will improve

as a result of a successful IG program are needed. But the bottom line is that reducing exposure to business risk, improving the quality and security of data and e-documents, cutting out unneeded stored information, and streamlining information technology (IT) development while focusing on business results add up to better organizational health and viability and, ultimately, an improved bottom line.

Let us take a step back and examine the major issues affecting information costing and calculating the real cost of holding information, consider Big Data and e-discovery ramifications, and introduce some new concepts that may help frame information costing issues differently for business managers. Getting a good handle on the true cost of information is essential to governing it properly, shifting resources to higher-value information, and discarding information that has no discernible business value and carries inherent, avoidable risks.

Changing Information Environment

The information environment is changing. Data volumes are growing, but **unstructured information** (such as e-mail, word processing documents, social media posts) is growing faster than our ability to manage it. Some unstructured information has more structure than others, containing some identifiable metadata (e.g. e-mail messages all have a header, subject line, time/date stamp, and message body). Some call this *semistructured* information, but for the purposes of this book, we use the term "unstructured information" to include semistructured information as well.

> The problem of unstructured IG is growing faster than the problem of data volume itself.

The volume of unstructured information is growing dramatically. Analysts estimate that, from 2017 to 2025, the amount of worldwide data will grow tenfold, to 163ZB (1 zettabyte = 1 trillion gigabytes).[2] However, the volume of *unstructured information* will actually grow approximately 50 percent faster than structured data. Analysts also estimate that fully 90 percent of unstructured information will require formal governance and management by 2020. In other words, the problem of unstructured IG is growing faster than the problem of data volume itself.

What makes unstructured information so challenging? There are several factors, including:

- *Horizontal versus vertical*. Unstructured information is typically not clearly attached to a department or a business function. Unlike the vertical focus of an enterprise resource planning (ERP) database, for example, an e-mail system serves multiple business functions—from employee communication to filing with regulators—for all parts of the business. Unstructured information is much more horizontal, making it difficult to develop and apply business rules.
- *Formality*. The tools and applications used to create unstructured information often engender informality and the sharing of opinions that can be problematic in litigation, investigations, and audits—as has been repeatedly demonstrated in front-page stories over the past decade. This problem is not likely to get any easier as social media technologies and mobile devices become more common in the enterprise.
- *Management location*. Unstructured information does not have a single, obvious home. Although e-mail systems rely on central messaging servers, e-mail is just as likely to be found on a file share, mobile device, or laptop hard drive. This makes the application of management rules more difficult than the application of the same rules in structured systems, where there is a close marriage between the application and the database.
- *"Ownership" issues*. Employees do not think that they "own" data in an accounts receivable system like they "own" their e-mail or documents stored on their hard drive. Although such information generally has a single owner (i.e., the organization itself), this nonownership mindset can make the imposition of management rules for unstructured information more challenging than for structured data.
- *Classification*. The business purpose of a database is generally determined prior to its design. Unlike structured information, the business purpose of unstructured information is difficult to infer from the application that created or stores the information. A word processing file stored in a collaboration environment could be a multimillion-dollar contract or a lunch menu. As such, classification of unstructured content is more complex and expensive than structured information.

Taken together, these factors reveal a simple truth: *Managing unstructured information is a separate and distinct discipline from managing databases*. It requires different methods and tools. Moreover, determining the costs and benefits of owning and managing unstructured information is a unique—but critical—challenge.

The governance of unstructured information creates enormous complexity and risk for business managers to consider while making it difficult for organizations to generate *real value* from all this information. Despite the looming crisis, most organizations have limited ability to quantify the real cost of owning and managing unstructured information. Determining the total cost of owning unstructured information is an essential precursor to managing and monetizing that information while cutting information costs—key steps in driving profit for the enterprise.

> Storing things is cheap ... I've tended to take the attitude, "Don't throw electronic things away."
>
> —*Data scientist quoted in Anne Eisenberg, "What 23 Years of E-Mail May Say About You,"* New York Times, April 7, 2012

> The company spent $900,000 to produce an amount of data that would consume less than one-quarter of the available capacity of an ordinary DVD.
>
> —*Nicholas M. Pace and Laura Zakaras, "Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery," RAND Institute for Civil Justice, 2012*

Calculating Information Costs

We are not very good at figuring out what information costs—*truly* costs. Many organizations act as if storage is an infinitely renewable resource and the only cost of information. But, somehow, enterprise storage spending rises each year and IT support costs rise, even as the root commodity (disk drives) grows ever cheaper and denser. Obviously, they are not considering labor and overhead costs incurred with managing information, and the additional knowledge worker time wasted sifting through mountains of information to find what they need.

Some of this myopic focus on disk storage cost is simple ignorance. The executive who concludes that a terabyte costs less than a nice meal at a restaurant after browsing storage drives on the shelves of a favorite big-box retailer on the weekend is of little help.

Rising information storage costs cannot be dismissed. Each year the billions that organizations worldwide spend on storage grows, even though the cost of a hard drive is less than 1 percent of what it was about a decade ago. We have treated storage as a resource that has no cost to the organization outside of the initial capital outlay and basic operational costs. This is shortsighted and outdated.

Some of the reason that managers and executives have difficulty comprehending the true cost of information is old-fashioned miscommunication. IT departments do not see (or pay for) the full cost of e-discovery and litigation. Even when IT "partners" with litigators, what IT learn rarely drives strategic IT decisions. Conversely, law departments (and outside firms) rarely own and pay for the IT consequences of their litigation strategies. It is as if when the litigation fire needs to be put out, nobody calculates the cost of gas for the fire trucks.

But calculating the cost of information—especially information that does not sit neatly in the rows and columns of enterprise database "systems of record"—is complex. It is more art than science. And it is more politics than art. There is no Aristotelian Golden Mean for information.

The true cost of mismanaging information is much more profound than simply calculating storage unit costs. It is the cost of *opportunity* lost—the lost benefit of information that is disorganized, created and then forgotten, cast aside, and left to rot. It is the cost of *information that cannot be brought to market*. Organizations that realize this, and invest in managing and leveraging their unstructured information, will be the winners of the next decade.

Most organizations own vast pools of information that is effectively "dark": They do not know what it is, where it is, who is responsible for managing it, or whether it is an asset or a liability. It is not classified, indexed, or managed according to the organization's own policies. It sits in shared drives, mobile devices, abandoned content systems, single-purpose cloud repositories, legacy systems, and outdated archives.

And when the light is finally flicked on for the first time by an intensive hunt for information during e-discovery, this dark information can turn out to be a liability. An e-mail message about "paying off fat people who are a little afraid of some silly lung problem" might seem innocent—until it is placed in front of a jury as evidence that a drug company did not care that its diet drug was allegedly killing people.[3]

The importance of understanding the total cost of owning unstructured information is growing. We are at the beginning of a "seismic economic shift" in the information landscape, one that promises to not only "reinvent society," (according to an MIT data scientist) but also to create "the new oil ... a new asset class touching all aspects of society."[4]

> Smart leaders across industries will see using big data for what it is: a management revolution.
>
> —*Andrew McAfee and Erik Brynjolfsson, "Big Data: The Management Revolution,"* Harvard Business Review (October 2012)

Big Data Opportunities and Challenges

We are entering the epoch of Big Data—an era of Internet-scale enterprise infrastructure, powerful analytical tools, and massive data sets from which we can potentially wring profound new insights about business, society, and ourselves. It is an epoch that, according to the consulting firm McKinsey, promises to save the European Union public sector billions of euros, increase retailer margins by 60 percent, reduce US national healthcare spending by 8 percent, while creating hundreds of thousands of jobs.[5] Sounds great, right?

However, the early days of this epoch are unfolding in almost total ignorance of the true cost of information. In the near nirvana contemplated by some Big Data proponents, *all data is good, and more data is better*. Yet it would be an exaggeration to say that there is no awareness of potential Big Data downsides. A recent study by the Pew Research Center was positive overall but did note concerns about privacy, social control, misinformation, civil rights abuses, and the possibility of simply being overwhelmed by the deluge of information.[6]

But the real-world burdens of managing, protecting, searching, classifying, retaining, producing, and migrating unstructured information are foreign to many Big Data cheerleaders. This may be because the Big Data hype cycle[7] is not yet in the "trough of disillusionment" where the reality of corporate culture and complex legal requirements sets in. But set in it will, and when it does, the demand for intelligent analysis of costs and benefits will be high.

IG professionals must be ready for these new challenges and opportunities—ready with new models for thinking about unstructured information. Models that calculate the *risks* of keeping too much of the wrong information as well as the *benefits* of clean, reliable, and accessible pools of the right information. Models that drive desirable behavior in the enterprise, and position organizations to succeed on the "next frontier for innovation, competition, and productivity."[8]

IG professionals must be ready with new models that calculate the risks of storing too much of the wrong information and also the benefits of clean, reliable, accessible information.

Full Cost Accounting for Information

It is difficult for organizations to make educated decisions about unstructured information without knowing its full cost. Models like total cost of ownership (TCO) and ROI are designed for this purpose and have much in common with **full cost accounting** (FCA) models. FCA seeks to create a complete picture of costs that includes past, future, direct, and indirect costs rather than direct cash outlays alone.

FCA has been used for many purposes, including the decidedly earthbound task of determining what it costs to take out the garbage and the loftier task of calculating how much the International Space Station really costs. A closely related concept often called triple bottom line has gained traction in the world of environmental accounting, positing that organizations must take into account societal and environmental costs as well as monetary costs.

The US Environmental Protection Agency promotes the use of FCA for municipal waste management, and several states have adopted laws requiring its use. It is fascinating—and no accident—that this accounting model has been widely used to calculate the full cost of managing an unwanted by-product of modern life. The analogy to outdated, duplicate, and unmanaged unstructured information is clear.

Organizations can learn from accounting models used by cities to calculate the total cost of managing municipal waste and apply them to the IG problem.

Applying the principles of FCA to information can increase cost transparency and drive better management decisions. In municipal garbage systems where citizens do not see a separate bill for taking out the garbage, it is more difficult to get new spending on waste management approved.[9] Without visibility into the true cost, how can citizens—or CEOs—make informed decisions?

Responsible, innovative managers and executives should investigate FCA models for calculating the total cost of owning unstructured information. Consider costs such as:

- *General and administrative costs,* such as cost of IT operations and personnel, facilities, and technical support.
- *Productivity gains or losses* related to the information.
- *Legal and e-discovery* costs associated with the information and information systems.
- *Indirect costs,* such as the accounting, billing, clerical support, contract management, insurance, payroll, purchasing, and so on.
- *Up-front costs,* such as the acquisition of the system, integration and configuration, and training. This should include the depreciation of capital outlays.
- *Future costs,* such as maintenance, migration, and decommissioning of information systems. Future outlays should be amortized.

Calculating the Cost of Owning Unstructured Information

Any system designed to calculate the cost or benefit of a business strategy is inherently political. That is, it is an *argument* designed to convince an *audience*. Well-known models like TCO and ROI are primarily decision tools designed to help organizations predict the economic consequences of a decision. While there are certainly objective truths about the information environment, human decision making is a complex and imperfect process. There are plenty of excellent guides on how to create a standard TCO or ROI. That is not our purpose here. Rather, we want to inspire creative thinking about how to calculate the cost of owning unstructured information and help organizations minimize the risk—and maximize the value—of unstructured information.

Any economic model for calculating the cost of unstructured information depends on reliable facts. But facts can be hard to come by. A client recently went in search of an accurate number for the annual cost per terabyte of Tier 1 storage in her company. The company's storage environment was completely outsourced, leading her to believe that the number would be transparent and easy to find. However, after days spent poring over the massive contract, she was no closer to the truth. Although there was a line item for storage costs, the true costs were buried in “complexity fees” and other opaque terms.

Organizations need tools that help them establish facts about their unstructured information environment. The business case for better management depends on these facts. Look for tools that can help you:

- *Find unstructured information wherever it resides* across the enterprise, including e-mail systems, shared network drives, legacy content management systems, and archives.
- Enable fast and intuitive access to *basic metrics*, such as size, date of last access, and file type.
- Provide *sophisticated analysis* of the nature of the content itself to drive classification and information life cycle decisions.
- Deliver visibility into the environment through *dashboards* that are easy to for nonspecialists to configure and use.

Identifying and building consensus on the sources of cost for unstructured information is critical to any TCO or ROI calculation. It is critical that all stakeholders agree on these sources, or they will not incorporate the output of the calculation in their strategy and planning.

Sources of Cost

Unstructured information is ubiquitous. It is typically not the product of a single-purpose business application. It often has no clearly defined owner. It is endlessly duplicated and transmitted across the organization. Determining where and how unstructured information generates cost is difficult.

However, doing so *is* possible. Our research shows that at least 10 key factors drive the total cost of owning unstructured information. These 10 factors identify where organizations typically spend money throughout the life cycle of managing unstructured information. These factors are listed in Figure 7.1, along with examples of elements that typically *increase* cost ("Cost Drivers," on the left side) and elements that typically *reduce* costs ("Cost Reducers," on the right side).

1. *E-discovery:* Finding, processing, and producing information to support lawsuits, investigations, and audits. Unstructured information is typically the most common target in e-discovery, and a poorly managed information environment can add millions of dollars in cost to large lawsuits. Simply reviewing a gigabyte of information for litigation can cost $14,000.[10]

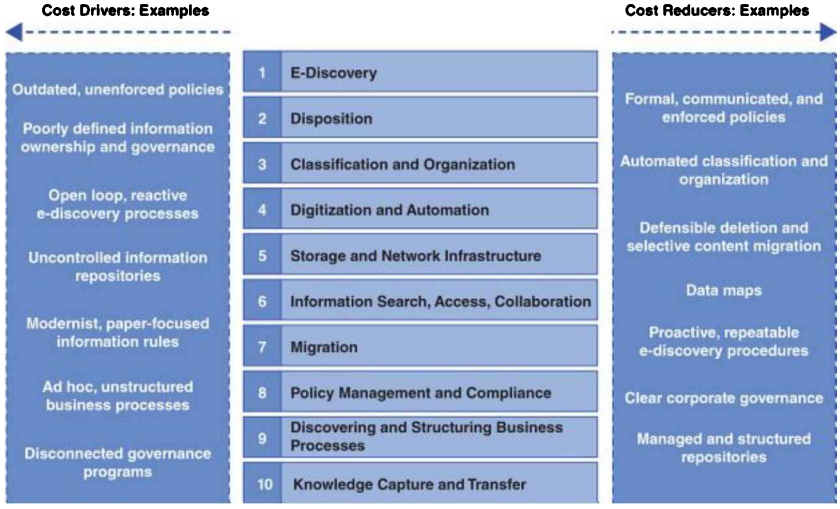


Figure 7.1 Key Factors Driving Cost

Source: Barclay T. Blair.

2. *Disposition:* Getting rid of information that no longer has value because it is duplicate, out of date, or has no value to the business. In poorly managed information environments, separating the wheat from the chaff can cost large organizations millions of dollars. For enterprises with frequent litigation, the risk of throwing away the wrong piece of information only increases risk and cost. Better management and smart IG tools drive costs down.
3. *Classification and organization:* Keeping unstructured information organized so that employees can use it. It also is necessary so management rules supporting privacy, privilege, confidentiality, retention, and other requirements can be applied.
4. *Digitization and automation*. Many business processes continue to be a combination of digital, automated steps and paper-based, manual steps. Automating and digitizing these processes requires investment but also can drive significant returns. For example, studies have shown that automating accounts payable "can reduce invoice processing costs by 90 percent."[11]
5. *Storage and network infrastructure:* The cost of the devices, networks, software, and labor required to store unstructured information. Although the cost of the baseline commodity (i.e. a gigabyte of storage space) continues to fall, for most organizations overall volume growth and complexity means that storage budgets go up each year. For example, between 2000 and 2010, organizations more than doubled the amount they spent on storage-related software even though the cost of raw hard drive space dropped by almost 100 times.[12]
6. *Information search, access, and collaboration:* The cost of hardware, software, and services designed to ensure that information is available to those who need it, when they need it. This typically includes enterprise content management systems, enterprise search, case management, and the infrastructure necessary to support employee access and use of these systems.
7. *Migration:* The cost of moving unstructured information from outdated systems to current systems. In poorly managed information environments, the cost of migration can be very high—so high that some organizations maintain legacy systems long after they are no longer supported by the vendor just to avoid (more likely, simply to *defer*) the migration cost and complexity.
8. *Policy management and compliance:* The cost of developing, implementing, enforcing, and maintaining IG policies on

unstructured information. Good policies, consistently enforced, will drive down the total cost of owning unstructured information.

9. *Discovering and structuring business processes:* The cost of identifying, improving, and systematizing or "routinizing" business processes that are currently ad hoc and disorganized. Typical examples include contract management and accounts receivable as well as revenue-related activities, such as sales and customer support. Moving from informal, e-mail, and document-based processes to fixed workflows drives down cost.
10. *Knowledge capture and transfer:* The cost of capturing critical business knowledge held at the department and employee level and putting that information in a form that enables other employees and parts of the organization to benefit from it. Examples include intranets and their more contemporary cousins such as wikis, blogs, and enterprise social media platforms.

The Path to Information Value

At the peak of World War II, 70,000 people went to work every day at the Brooklyn Navy Yard. The site was once America's premier shipbuilding facility, building the steam-powered *Ohio* in 1820 and the aircraft carrier *USS Independence* in the 1950s. But the site fell apart after it was decommissioned in the 1960s. Today, the "Admiral's Row" of Second Empire–style mansions once occupied by naval officers is an extraordinary sight, with gnarled oak trees pushing through the rotting mansard roofs.[13]

> Seventy percent of managers and executives say data are "extremely important" for creating competitive advantage. "The key, of course, is knowing which data matter, who within a company needs them, and finding ways to get that data into users' hands."
>
> *—The Economist Intelligence Unit, "Levelling the Playing Field: How Companies Use Data to Create Advantage" (January 2011)*

However, after decades of decay, the Navy Yard is being reborn as the home of hundreds of businesses—from major movie studios to artisanal whisky makers—taking advantage of abundant space and a desirable location. There were three phases in the yard's rebirth:

1. *Clean*. Survey the site to determine what had value and what did not. Dispose of toxic waste and rotting buildings, and modernize the infrastructure.
2. *Build and maintain*. Implement a plan to continuously improve, upgrade, and maintain the facility.
3. *Monetize*. Lease the space.

Most organizations face a similar problem. However, our Navy yards are the vast piles of unstructured information that were created with little thought to how and when the pile might go away. They are records management programs built for a different era —like an automobile with a metal dashboard, six ashtrays, and no seat belts. Our Navy yards are information environments no longer fit for purpose in the Big Data era, overwhelmed by volume and complexity.

We are doing a bad job at managing information. McKinsey estimates that in some circumstances, companies are using up to 80% of their infrastructure to store *duplicate* data.[14] Nearly half of respondents in a survey ViaLumina recently conducted said that at least 50% of the information in their organization is duplicate, outdated, or unnecessary.[15] We can do better.

Phase 1. Clean

We should put the Navy Yard's blueprint to work, by first identifying our piles of rotting unstructured information. Duplicate information. Information that has not been accessed in years. Information that no longer supports a business process and has little value. Information that we have no legal obligation to keep. The economics of such "defensible deletion" projects can be compelling simply on the basis of recovering the storage space and thus *reallocating capital that would have been spent on the annual storage purchase*.

Step 2. Build and Maintain

Cleaning up the Navy Yard is only the first step. We cannot repeat the past mistakes. We avoid this by building and maintaining an IG program that establishes our information constitution (why), laws (what), and regulations (how). We need a corporate governance, compliance, and audit plan that gives the program teeth, and a technology infrastructure that makes it real. It must be a defensible program to ensure we comply with the law and manage regulatory risk.

> Key steps in driving information value are: (1) clean; (2) build and maintain; and (3) monetize.

Phase 3. Monetize

IG is a means to an end, and that end is value creation. IG also mitigates risk and drives down cost. But extracting value is the key. Although monetization and value creation often are associated with structured data, new tools and techniques create exciting new opportunities for value creation from unstructured information.

For example, what if an organization could use sophisticated analytics on the e-mail account of their top salesperson (the more years of e-mail the better), look for markers of success, then train and hire salespeople based on that template? What is the pattern of a salesperson's communications with customers and prospects in her territory? What is the substance of the communications? What is the tone? When do successful salespeople communicate? How are the patterns different between successful deals and failed deals? What knowledge and insight resides in the thousands of messages and gigabytes of content? The tools and techniques of Big

Data applied to e-mail can bring powerful business insights. However, we have to know what questions to ask. According to Computerworld, "the hardest part of using big data is trying to get business people to sit down and define what they want out of the huge amount of unstructured and semi-structured data that is available to enterprises these days."[16]

The analytics challenges of Big Data create opportunities. For example, McKinsey predicts that demand for "deep analytical talent in the United States could be 50 to 60 percent greater than its projected supply by 2018." A chief reason for this gap is that "this type of talent is difficult to produce, taking years of training in the case of someone with intrinsic mathematical abilities." However, the more profound opportunity is for the "1.5 million extra additional managers and analysts in the United States who can ask the right questions and consume the results of the analysis of big data effectively."[17]

Table 7.1 Key Steps in the IG Process

Source: Barclay T. Blair.

Phase 1. Clean	Phase 2. Build and Maintain	Phase 3. Monetize
Information inventory Defensible deletion Records retention and legal hold	IG policies and procedures Corporate governance, compliance and audit Technology	Create value through information, e.g. drive sales and improve customer satisfaction Business insights Increase margins

Some companies are using analytics to set prices. For example, the largest distributor of heating oil in the United States sets prices on the fly, based on commodity prices and customer retention risks.[18] In a case that caught the attention of morning news shows, with breathless headlines like "Are Mac Users Paying More?" an online travel company revealed that "Mac users are 40 percent more likely to book four or five-star hotels ... compared to PC users."[19] Despite the headlines, the company was not charging Mac users more. Rather, computer brand was a variable used to determine which products were highlighted.

The path to information value is not necessarily linear. Different parts of your business may achieve maturity at different rates, driven by the unique risks and opportunities of the information they possess.

Challenging the Culture

The best models for calculating the total cost of owning unstructured are those that information professionals can use to challenge and change organizational culture. Much of the unstructured information that represents the greatest cost and risk to organizations is created, communicated, and managed directly by employees—that is, by human beings. As such, better IG relies in part on improving the way those human beings use and manage information.

New Information Models

The information calorie and information cap-and-trade models, explored next, are two new models designed to help with the challenge of governing information.

> There's not a person in a business anywhere who gets up in the morning and says, "Gee, I want to race into the office to follow some regulation." On the other hand, if you say, "There's an upside potential here, you're going to make money," people do get up early and do drive hard around the possibility of finding themselves winners on this.
>
> —*Dan Etsy, environmental policy professor at Yale University, quoted in Richard Conniff, "The Political History of Cap and Trade,"* Smithsonian Magazine (August 2009)

Consider a cap-and-trade system for information. Do not limit the creation and storage of *useful* information—that defeats the purpose of investing in IT in the first place. Rather, design a cap-and-trade system that controls the amount of *information pollution* and rewards innovation and management discipline.

While there is no objective "right amount" of information for every organization or department, we can certainly do better than "as much as you want, junk or not." After all, "nearly all sectors in the US economy had at least an average of 200 terabytes of stored data ... and many sectors had more than 1 petabyte in mean stored data per company."[20] Moreover, up to 50% of that information is easily identifiable as data pollution.[21] So, we have a reasonable starting point.

Here are some tips for creating an information cap-and-trade system:

- *Baseline the desired amount* of information per system, department, and/or type of user. How much information do you currently have? How much has value? How much should you have? These are not easy questions to answer, but even rough calculations can make a big difference.
- Create information volume targets or quotas, and *allocate them by business unit,* system, or user. This is the "cap" part of the system.
- Calculate the fully loaded cost of a *unit of information*, and adopt it as a baseline metric for the "trade" part of the system. Consider whether annual e-discovery costs can be allocated to this unit in a reasonable way.
- Create an internal accounting system for tracking and *trading information units*, or credits within the organization. Innovative departments will be rewarded, laggards will be motivated.
- Get *creative* in what the credits can purchase. New revenue-generating software? Headcount?

Future State: What Will the IG-Enabled Organization Look Like?

When an organization is IG enabled, or "IG mature"—meaning IG is "baked in" or infused into operations throughout the enterprise and coordinated on an organization-wide level—it will look significantly different from most organizations today. Not only will the organization routinely execute business processes with inherent privacy and security considerations, but also it will have a solid handle on the total cost of information; not only will it have shifted resources to capitalize on the opportunities of Big Data; not only will it be managing the deluge in a systematic, business-oriented way by cutting out data debris and leveraging information value; it will also look significantly different in key operational areas including legal, privacy and security, and IT.

The organization will have an embedded, robust, and ongoing security awareness training (SAT) program, and will adhere to standards such as ISO 27001 and guidance from the Cloud Security Alliance to keep data secure. It will also deploy and utilize advanced tools such as data loss prevention (DLP) and information rights management (IRM). Concerning privacy matters, the organization will have an ongoing **privacy awareness training** (PAT) program, and be able to routinely meet the requirements of new regulations, such as GDPR, or the California Consumer Privacy Act.

In legal matters, the mature IG-enabled organization will be better suited to address litigation in a more efficient way through a standardized legal hold notification (LHN) process. Legal risk is reduced through improved IG, which will manage information privacy in accordance with applicable laws and regulations. During litigation, your legal team will be able to sort through information more rapidly and efficiently, improving your legal posture, cutting e-discovery costs, and allowing for attorney time to be focused on strategy and to zero in on key issues. This means attorneys should have the tools to be more effective. Adherence to retention schedules means that records and documents can be discarded at the earliest possible time, which reduces the chances that some information could pose a legal risk, while also reducing the storage footprint. Hard costs can be saved by eliminating the approximately 69% of stored information that no longer has business value. That cost savings may be the primary rationale for the initial IG program effort. By leveraging advanced technologies such as artificial intelligence (AI), analytics, and forms of these, such as predictive coding, the organization can reduce the costs of e-discovery and better utilize attorney time.

RIM functions will operate with more efficiency and in compliance with laws and regulations. Appropriate retention periods will be applied and enforced, and authentic, original copies of business records will be readily identifiable, so that managers are using current and accurate information on which to base their decisions. Over the long term, valuable information from projects, product development, marketing programs, and strategic initiatives will be retained in corporate memory, reducing the impact of turnover and providing distilled information and knowledge to contribute to a **knowledge management** (KM) program. KM programs can facilitate innovation in organizations, as a knowledge base is built, retained, expanded, and leveraged.

In your IT operations, a focus on how IT can contribute to business objectives will bring about a new perspective. Using more of a business lens to view IT projects will help IT to contribute toward the achievement of business objectives. IT will be working more closely with privacy, security, legal, RIM, risk, and other business units, which should help these groups to have their needs and issues better addressed by IT solutions. Having a standardized data governance program in place means cleaning up corrupted or duplicated data and providing users with clean, accurate data as a basis for line-of-business software applications and for decision support in **business intelligence** (BI) and analytics applications. Better data is the basis for improved insights, which can be gained by leveraging analytics, and will improve management decision-making capabilities and help to provide better customer service, which can impact customer retention. It costs a lot more to gain a new customer than to retain an existing one, and with better data quality, the opportunities to cross-sell and upsell customers will be improved. *This can provide a sustainable competitive advantage.* Standardizing the use of business terms will facilitate improved communications between IT and other business units, which should lead to improved software applications that address user needs. Adhering to information life cycle management principles will help the organization to apply the proper level of IT resources to its high-value information while decreasing costs by managing information of declining value appropriately. IT effectiveness and efficiency will be improved by using IT frameworks and standards, such as COBIT 2019 and ISO/IEC 38500:2008, the international standard that provides high-level principles and guidance for senior executives and directors, and those advising them, for the effective and efficient governance of IT.[22] Implementing a master data management (MDM) program will help larger organizations with complex IT operations to ensure that they are working with consistent, accurate, and up-to-date data from a single source. Improved database security through data masking, database activity monitoring, database auditing, and other tools will help guard the organization's critical databases against the risk of rogue attacks by hackers. Deploying document life cycle security tools such as data loss prevention and information rights management will help secure your confidential information assets and keep them from prying eyes. This helps to secure the organization's competitive position and protect its valuable intellectual property.

By securing your electronic documents and data, not only within the organization but also for mobile use, and by monitoring and complying with applicable privacy laws, your confidential information assets will be safeguarded, your brand will be better protected, and your employees will be able to be productive without sacrificing the security of your information assets.

Moving Forward

We are not very good at figuring out what unstructured information costs. The Big Data deluge is upon us. If we hope to manage—and, more important, to monetize—this deluge, we must form cross-functional teams and challenge the way our organizations think about unstructured information. The first and most important step is developing the ability to convincingly calculate what unstructured information really costs and then to discover ways we can recoup those costs and drive value. These are foundational skills for information professionals in the new era of Big Data and infonomics. In this era, information is currency—but a currency that has value only when IG professionals drive innovation and management rigor in the unstructured information environment. Shaping up unstructured information by standardizing and inserting metadata (by using file analysis/content analysis tools) can help an organization harvest value from the majority of its information, which has largely been untapped.

CHAPTER SUMMARY: KEY POINTS

- IG program objectives must be aligned with organizational business objectives to gain executive and stakeholder support.
- Infonomics provides the framework for businesses to monetize, manage, and measure information as an actual asset.
- The most comprehensive assessment tool is the IG Process Maturity Model from CGOC.
- The business case for IG programs has historically been difficult to justify.
- It takes a commitment to the long view to develop a successful IG program.
- The problem of unstructured IG is growing faster than the problem of data volume itself.
- IG professionals must be ready with new models that calculate the risks of storing too much of the wrong information and also the benefits of clean, reliable, accessible information.
- Key steps in driving information value are: (1) clean; (2) build and maintain; and (3) monetize.
- The information calorie approach and information cap-and-trade are two new models for assisting in IG.
- Legal risk is reduced through improved IG, and legal costs are reduced.
- Leveraging newer technologies like predictive coding can improve the efficiency of legal teams.
- Adherence to retention schedules means that records and documents can be discarded at the earliest possible time, which reduces costs by eliminating unneeded information that no longer has business value.
- RIM functions will operate with more efficiency and in compliance with laws and regulations under a successful IG program.
- A compliant RIM program helps to build the organization's corporate memory of essential "lessons learned," which can foster a KM program.
- KM programs can facilitate innovation in organizations.
- Focusing on business impact and customizing your IG approach to meet business objectives are key best practices for IG in the IT department.
- Effective data governance can yield bottom-line benefits derived from new insights, especially with the use of BI software.
- IT governance seeks to align business objectives with IT strategy to deliver business value.
- Using IT frameworks like COBIT 2019 can improve the ability of senior management to monitor IT value and processes.
- Identifying sensitive information in your databases and implementing database security best practices help reduce organizational risk and the cost of compliance.
- By securing your electronic documents and data, your information assets will be safeguarded and your organization can more easily comply with privacy laws and regulations.
- We are not very good at figuring out what unstructured information costs. To thrive in the era of Big Data requires challenging the way we think about the cost of managing unstructured information.

Notes

1. "2017 CGOC Information Governance Process Maturity Model," https://www.cgoc.com/updated-ig-process-maturity-model-reflects-todays-data-realities-2/ (accessed December 19, 2018).
2. Nick Ismail, "The Value of Data: Forecast to Grow 10-fold by 2025," *Information Age*, April 5, 2017, https://www.information-age.com/data-forecast-grow-10-fold-2025-123465538/.
3. Richard B. Schmidt, "The Cyber Suit: How Computers Aided Lawyers in Diet-Pill Case," *Wall Street Journal*, October 8, 1999, http://webreprints.djreprints.com/000000000000000000012559001.html.
4. Nick Bilton, "At Davos, Discussions of a Global Data Deluge," *New York Times*, January 25, 2012, http://bits.blogs.nytimes.com/2012/01/25/at-davos-discussions-of-a-global-data-deluge/; Alex Pentland, quoted by Edge.org in "Reinventing Society in the Wake of Big Data," August 30, 2012, www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data; World Economic Forum, "Personal Data: The Emergence of a New Asset Class" (January 2011), http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf.
5. James Manyika et al., "Big Data: The Next Frontier for Innovation, Competitions, and Productivity," McKinsey Global Institute (May 2011), https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.
6. Janna Quitney Anderson and Lee Ranie, "Future of the Internet: Big Data," Pew Internet and American Life Project, July 20, 2012. www.pewinternet.org/2011/12/13/future-of-the-internet-role-of-the-web-and-new-media-in-the-public-sector/.
7. Louis Columbus, "Roundup of Big Data Forecasts and Market Estimates, 2012," *Forbes*, August 16, 2012,

https://www.forbes.com/sites/louiscolumbus/2012/08/16/roundup-of-big-data-forecasts-and-market-estimates-2012/#1c8022903cdf.

8. McKinsey Global Institute, "Big Data: The Next Frontier for Innovation, Competitions, and Productivity" (May 2011).

9. U.S. EPA, "Making Solid Waste Decisions with Full Cost Accounting," n.d., https://nepis.epa.gov/Exe/ZyNET.exe/9100MNXP.TXT?ZyActionD=ZyDocument&Client=EPA&Index=1995+Thru+1999&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack (accessed December 19, 2018).

10. Nicholas M. Pace and Laura Zakaras, "Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery," RAND Institute for Civil Justice, 2012, https://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1208.pdf (accessed December 19, 2018).

11. Accounts Payable Network, "A Detailed Guide to Imaging and Workflow ROI," 2010.

12. Various sources. See, for example: Barclay T. Blair, "Today's PowerPoint Slide: The Origins of Information Governance by the Numbers," October 28, 2010, http://barclaytblair.com/origins-of-information-governance-powerpoint/.

13. Brooklyn Navy Yard Development Corporation, "The History of Brooklyn Navy Yard," http://www.brooklynnavyyard.org/history.html (accessed December 19, 2018).

14. Manyika et al., "Big Data."

15. Barclay Blair and Barry Murphy, "Defining Information Governance: Theory or Action? Results of the 2011 Information Governance Survey," *eDiscovery Journal* (September 2011). eDJ.com.

16. Jaikumar Vijayan, "Finding the Business Value in Big Data Is a Big Problem," *Computerworld*, September 12, 2012, www.computerworld.com/s/article/9231224/Finding_the_business_value_in_big_data_is_a_big_problem.

17. Manyika et al., "Big Data."

18. Economist Intelligence Unit, "Levelling the Playing Field: How Companies Use Data to Create Advantage" (January 2011), https://eiuperspectives.economist.com/technology-innovation/levelling-playing-field.

19. Genevieve Shaw Brown, "Mac Users May See Pricier Options on Orbitz," *ABC Good Morning America*, June 25, 2012, http://abcnews.go.com/Travel/mac-users-higher-hotel-prices-orbitz/story?id=16650014#.UDlkVBqe7oV.

20. Manyika et al., "Big Data."

21. Blair and Murphy, "Defining Information Governance."

22. International Organization for Standardization, ISO/IEC 8500:2008, Corporate Governance of Information Technology, https://www.iso.org/standard/51639.html (accessed December 19, 2018).

CHAPTER 8
Information Governance and Legal Functions

Robert Smallwood with Randy Kahn, Esq., and Barry Murphy

Perhaps the key functional area that **information governance** (IG) impacts most is legal functions, since legal requirements are paramount. Failure to meet them can literally put an organization out of business or land executives in prison. Privacy, security, records management, information technology (IT), and business management functions are important—very important—but the most significant aspect of all of these functions relates to legality and regulatory compliance.

Key legal processes include electronic discovery (**e-discovery**) readiness and associated business processes, information and record retention policies, the **legal hold notification** (LHN) process, and legally **defensible disposition** practices.

Some newer technologies have become viable to assist organizations in implementing their IG efforts, namely, **predictive coding** and **technology-assisted review** (TAR; also known as **computer-assisted review**). In this chapter we cover the 2006 and 2015 changes to the Federal Rules of Civil Procedure (FRCP), explore the need for leveraging IT in IG efforts aimed at defensible disposition, the intersection between IG processes and legal functions, policy implications, and some key enabling technologies.

Introduction to E-Discovery: The Revised 2006 and 2015 Federal Rules of Civil Procedure Changed Everything

Since 1938, the Federal Rules of Civil Procedure "have governed the discovery of evidence in lawsuits and other civil cases."[1] In law, **discovery** is an early phase of civil litigation where plaintiffs and defendants investigate and exchange evidence and testimony to better understand the facts of a case and to make early determinations of the strength of arguments on either side. Each side must produce evidence requested by the opposition or show the court why it is unreasonable to produce the information. As the Supreme Court stated in an early case under the Federal rules, "Mutual knowledge of all the relevant facts gathered by both parties is essential to proper litigation" (*Hickman v. Taylor*, 329 U.S. 495, 507 (1947)).

The FRCP apply to US district courts, which are the trial courts of the federal court system. The district courts have jurisdiction (within limits set by Congress and the Constitution) to hear nearly all categories of federal cases, including civil and criminal matters.[2]

The FRCP were substantially amended in 2006, and further modified in 2015, specifically to address issues pertaining to the preservation and discovery of

electronic records in the litigation process.[3] These changes were a long time coming, reflecting the lag between the state of technology and the courts' ability to catch up to the realities of electronically generated and stored information.

> Legal functions are the most important area of IG impact.

After years of applying traditional paper-based discovery rules to e-discovery, amendments to the FRCP were made to accommodate the modern practice of discovery of **electronically stored information** (ESI). *ESI is any information that is created or stored in electronic format*. The goal of the 2006 FRCP amendments was to recognize the importance of ESI and to respond to the increasingly prohibitive costs of document review and protection of privileged documents. These amendments reinforced the importance of IG policies, processes, and controls in the handling of ESI.[4] Organizations must produce requested ESI reasonably quickly, and failure to do so, or failure to do so within the prescribed time frame, can result in sanctions. This requirement dictates that organizations put in place IG policies and procedures to be able to produce ESI accurately and in a timely fashion.[5]

> ESI is any information that is created or stored in electronic format.

All types of litigation are covered under the FRCP, and all types of e-documents—most especially e-mail—are included, which can be created, accessed, or stored in a wide variety of methods, and on a wide variety of devices beyond hard drives. The FRCP apply to ESI held on all types of storage and communications devices: thumb drives, CD/DVDs, smartphones, tablets, personal digital assistants (PDAs), personal computers, servers, zip drives, floppy disks, backup tapes, and other storage media. ESI content can include information from e-mail, reports, blogs, social media posts (e.g., Twitter posts), voicemails, wikis, Web sites (internal and external), word processing documents, and spreadsheets, and includes the **metadata** associated with the content itself, which provides descriptive information.[6]

Under the FRCP amendments, corporations must proactively manage the e-discovery process to avoid sanctions, unfavorable rulings, and a loss of public trust. Corporations must be prepared for early discussions on e-discovery with all departments. Topics should include the form of production of ESI and the methods for preservation of information. Records management and IT departments must have made available all relevant ESI for attorney review.[7]

The goal of the 2006 FRCP amendments was to recognize the importance of ESI and to respond to the increasingly prohibitive costs of document review and protection of privileged documents.

This new era of ESI preservation and production demands the need for cross-functional collaboration: Records management, IT, and legal teams particularly need to work closely together. Legal teams, with assistance and input of records management staff, must identify relevant ESI, and IT teams must be mindful of preserving and protecting the ESI to maintain its legal integrity and prove its authenticity.

The further revisions made in 2015 emphasized the importance of "proportionality" in litigation, as well as setting a more uniform standard for when courts may impose "curative measures," including sanctions, in discovery.

Big Data Impact

Now throw in the Big Data effect: the average employee creates roughly one gigabyte of data annually (and growing), and data volumes are expected to increase over the next decade not 10-fold, or even 20-fold, but as much as 40 to 50 times what it is today![8] This underscores the fact that organizations must meet legal requirements while paring down the mountain of data debris they are holding to reduce costs and potential liabilities hidden in that monstrous amount of information. There are also costs associated with **dark data**—useless data, such as old log files, that takes up space and continues to grow and needs to be cleaned up.

Some data is important and relevant, but distinctions must be made by IG policy to classify, prioritize, and schedule data for disposition *and to dispose of the majority of it in a systematic, legally defensible way*. If organizations do not accomplish these critical IG tasks they will be overburdened with storage and data handling costs and will be unable to meet legal obligations.

According to a survey by the Compliance, Governance, and Oversight Council (CGOC), approximately 25% of information stored in organizations has real business value, while 5% must be kept as business records and about 1% is retained due to a litigation hold.[9] "*This means that [about] 69 percent of information in most companies has no business, legal, or regulatory value*. Companies that are able to [identify and] dispose of this debris return more profit to shareholders, can use more of their IT budgets for strategic investments, and can avoid excess expense in legal and regulatory response" (emphasis added).

The amended FRCP reinforce the importance of IG. Only about 25% of business information has real value, and 5% are business records.

If organizations are not able to draw clear distinctions between that roughly 30 percent of "high-value" business data, records, and that which is on legal hold, their IT department are tasked with the impossible job of managing all data as if it is high value. This "overmanaging" of information is a significant waste of IT resources.[10]

More Details on the Revised FRCP Rules

Here we present a synopsis of the key points in FRCP rules that apply to e-discovery.

FRCP 1—Scope and Purpose. This rule is simple and clear; its aim is to have all rules construed by both courts and parties in litigation "to secure the just, speedy, and inexpensive determination of every action."[11] Your discovery effort and responses must be executed in a timely manner.

FRCP 16—Pretrial Conferences; Scheduling; Management. This rule provides guidelines for preparing for and managing the e-discovery process; the court expects IT and network literacy on both sides, so that pretrial conferences regarding discoverable evidence are productive.

FRCP 26—Duty to Disclose; General Provisions Governing Discovery. This rule protects litigants from costly and burdensome discovery requests, given certain guidelines.

FRCP 26(a)(1)(C). Requires that you make initial disclosures no later than 14 days after the Rule 26(f) meet and confer, unless an objection or another time is set by stipulation or court order. If you have an objection, now is the time to voice it.

Rule 26(b)(1). The rule as amended in 2015 expressly states that the scope of discovery must be proportional to the needs of the case, taking into account the importance of the issues at stake, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving issues, and whether the burden or expense of discovery outweighs its likely benefit. The rule allows parties responding to discovery to argue that in certain cases that discovery is unreasonably burdensome.

Rule 26(b)(2)(B). Introduced the concept of *not reasonably accessible* ESI. The concept of *not reasonably accessible paper* had not existed. This rule provides procedures for shifting the cost of accessing not reasonably accessible ESI to the requesting party.

Rule 26(b)(5)(B). Gives courts a clear procedure for settling claims when you hand over ESI to the requesting party that you shouldn't have.

Rule 26(f). This is the meet and confer rule. This rule requires all parties to meet within 99 days of the lawsuit's filing and at least 21 days before a scheduled conference.

Rule 26(g). Requires an attorney to sign every e-discovery request, response, or objection.

FRCP 33—Interrogatories to Parties. This rule provides a definition of business e-records that are discoverable and the right of opposing parties to request and access them.

FRCP 34—Producing Documents, Electronically Stored Information, and Tangible Things, or Entering onto Land, for Inspection and Other Purposes. In disputes over document production, this rule outlines ways to resolve and move forward. Specifically, FRCP 34(b) addresses the format for requests and requires that e-records be accessible without undue difficulty (i.e. the records must be organized and identified). The requesting party chooses the preferred format, which are usually native files (which also should contain metadata). The key point is that electronic files must be accessible, readable, and in a standard format.

FRCP 37—Failure to Make Disclosures or to Cooperate; Sanctions. As amended in 2015, Rule 37(e) spells out a more uniform test for courts finding that curative measures or sanctions are appropriate to impose based on a party's failure to preserve ESI. If ESI should have been preserved but is lost due to a party's failure to take reasonable steps to preserve it, and it cannot be replaced, then upon a finding of prejudice a court may order measures to cure the prejudice. In the case of intentional misconduct, this may even include case-ending sanctions. This rule underscores the need for a legally defensible document management program under the umbrella of clear IG policies.

> The Big Data trend underscores the need for defensible deletion of data debris.

Landmark E-Discovery Case: Zubulake v. UBS Warburg

A landmark case in e-discovery arose from the opinions rendered in *Zubulake v. UBS Warburg*, an employment discrimination case where the plaintiff, Laura Zubulake, sought access to e-mail messages involving or naming her. Although UBS produced over 100 pages of evidence, it was shown that employees intentionally deleted some relevant e-mail messages.[12] The plaintiffs requested copies of e-mail from backup tapes, and the defendants refused to provide them, claiming it would be too expensive and burdensome to do so.

The judge ruled that UBS had not taken proper care in preserving the e-mail

evidence, and the judge ordered an **adverse inference** (assumption that the evidence was damaging) instruction against UBS. Ultimately, the jury awarded Zubulake over $29 million in total compensatory and punitive damages. “The court looked at the proportionality test of Rule 26(b)(2) of the Federal Rules of Civil Procedure and applied it to the electronic communication at issue. Any electronic data that is as accessible as other documentation should have traditional discovery rules applied.”[13] Although Zubulake's award was later overturned on appeal, it is clear the stakes are huge in e-discovery and preservation of ESI.

In the landmark case *Zubulake v. UBS Warburg*, the defendants were severely punished by an adverse inference for deleting key e-mails and not producing copies on backup tapes.

E-Discovery Techniques

Current e-discovery techniques include online review, e-mail message archive review, and cyberforensics. Any and all other methods of seeking or searching for ESI may be employed in e-discovery. Expect capabilities for searching, retrieving, and translating ESI to improve, expanding the types of ESI that are discoverable. Consider this potential when evaluating and developing ESI management practices and policies.[14]

SEVEN STEPS OF THE E-DISCOVERY PROCESS

In the e-discovery process, you must perform certain functions for identifying and preserving electronically stored (ESI), and meet requirements regarding conditions such as relevancy and privilege. Typically, you follow this e-discovery process:

1. Create and retain ESI according to an enforceable electronic records retention policy and electronic records management (ERM) program. Enforce the policy, and monitor compliance with it and the ERM program.
2. Identify the relevant ESI, preserve any so it cannot be altered or destroyed, and collect all ESI for further review.
3. Process and filter the ESI to remove the excess and duplicates. You reduce costs by reducing the volume of ESI that moves to the next stage in the e-discovery process.
4. Review and analyze the filtered ESI for privilege because privileged ESI is not discoverable, unless some exception kicks in.
5. Produce the remaining ESI, after filtering out what's irrelevant, duplicated, or privileged. Producing ESI in native format is common.
6. Clawback [find and return] the ESI that you disclosed to the opposing party that you should have filtered out, but did not. Clawback is not unusual, but you have to work at getting clawback approved, and the court may deny it.
7. Present at trial if your case hasn't settled. Judges have little to no patience with lawyers who appear before them not understanding e-discovery and the ESI of their clients or the opposing side.

Source: Linda Volonino and Ian Redpath, e-*Discovery For Dummies* (Hoboken, NJ: John Wiley & Sons, 2010), https://www.dummies.com/business/e-discovery-for-dummies-cheat-sheet/ (accessed December 5, 2018). Used with permission.

E-Discovery Reference Model

The **E-Discovery Reference Model** is a visual planning tool created by EDRM.net to assist in identifying and clarifying the stages of the e-discovery process. Figure 8.1 is the graphic depiction with accompanying detail on the process steps.

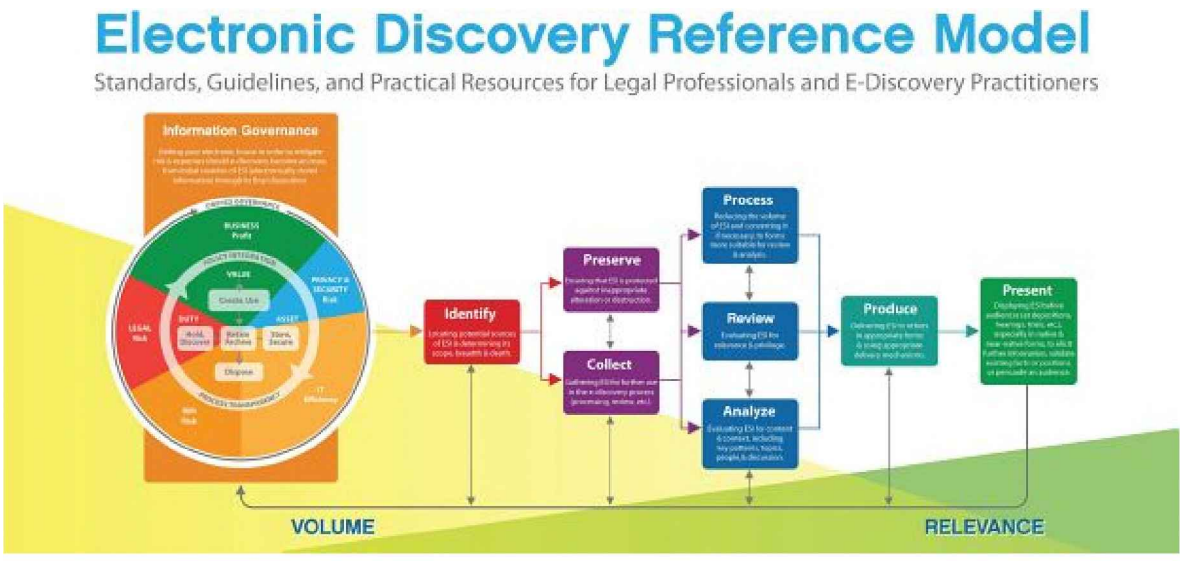


Figure 8.1 Electronic Discovery Reference Model

Source: EDRM (https://www.edrm.net/frameworks-and-standards/edrm-model/).

Information governance. Getting your electronic house in order to mitigate risk and expenses should e-discovery become an issue, from initial creation of electronically stored information through its final disposition

Identification. Locating potential sources of ESI and determining their scope, breadth, and depth

Preservation. Ensuring that ESI is protected against inappropriate alteration or destruction

Collection. Gathering ESI for further use in the e-discovery process (processing, review, etc.)

Processing. Reducing the volume of ESI and converting it, if necessary, to forms more suitable for review and analysis

Review. Evaluating ESI for relevance and privilege

Analysis. Evaluating ESI for content and context, including key patterns, topics, people, and discussion

Production. Delivering ESI to others in appropriate forms, and using appropriate delivery mechanisms

Presentation. Displaying ESI before audiences (at depositions, hearings, trials, etc.), especially in native and near-native forms, to elicit further information, validate existing facts or positions, or persuade an audience[15]

The Electronic Discovery Reference Model can assist organizations in focusing and segmenting their efforts when planning e-discovery initiatives.

The E-Discovery Reference Model is in a planning tool that presents key e-discovery process steps.

Guidelines for E-Discovery Planning

1. *Implement an IG program*. The highest impact area to focus are your legal processes, particularly e-discovery. From risk assessment to processes, communications, training, controls, and auditing, fully implement IG to improve and measure compliance capabilities.
2. *Inventory your ESI*. File scanning and e-mail archiving software can assist you. You also will want to observe files and data flows by doing a walk-through beginning with centralized servers in the computer room and moving out into business areas. Then, using a prepared inventory form, you should interview users to find out more detail. Be sure to inventory ESI based on computer systems or applications, and diagram it out.
3. *Create and implement a comprehensive records retention policy, and also include an e-mail retention policy and retention schedules for major ESI areas*. This is required since all things are potentially discoverable. You must devise a comprehensive retention and disposition policy that is legally defensible. So, for instance, if your policy is to destroy all e-mail messages that do not have a legal hold (or are expected to) after 120 days, and apply that policy uniformly, you will be able to defend the practice in court. Also, implementing the retention policy reduces your storage burden and costs while cutting the risk of liability that might be buried in obscure e-mail messages.
4. *As an extension of your retention policy, implement a legal hold policy that is enforceable, auditable, and legally defensible. Be sure to include all potentially discoverable ESI*. We discuss legal holds in more depth later in this chapter, but be sure to cast a wide net when developing retention policies so that you include all relevant electronic records, such as e-mail, e-documents and scanned documents, storage discs, and backup tapes.
5. *Leverage technology*. Bolster your e-discovery planning and execution efforts by deploying enabling technologies, such as e-mail archiving, advanced enterprise search, TAR, and predictive coding.
6. *Develop and execute your e-discovery plan*. You may want to begin from this point forward with new cases, and bear in mind that starting small and piloting is usually the best course of action.

Implementing IG, inventorying ESI, and leveraging technology to implement records retention and LHN policies are key steps in e-discovery planning.

The Intersection of IG and E-Discovery

By Barry Murphy

Effective IG programs can alleviate e-discovery headaches by reducing the amount of information to process and review, allowing legal teams to get to the facts of a case quickly and efficiently, and can even result in better case outcomes. Table 8.1 shows the impact of IG on e-discovery, by function.

Table 8.1 IG Impact on E-Discovery

Source: Barry Murphy.

Impact	Function
Cost reduction	Reduce downstream costs of processing and review by defensibly disposing of data according to corporate retention policies Reduce cost of collection by centralizing collection interface to save time Keep review costs down by prioritizing documents and assigning to the right level associates (better resource utilization) Reduce cost of review by culling information with advanced analytics
Risk management	Reduce risk of sanctions by managing the process of LHN and the collection and preservation of potentially responsive information
Better litigation win rates	Optimize decision making (e.g. settling cases that can't be won) quickly with advanced analytics that prioritize hot documents Quickly find the necessary information to win cases with advanced searches and prioritized review
Strategic planning for matters based on merit	Determine the merits of a matter quickly and decide if it is a winnable case Quickly route prioritized documents to the right reviewers via advanced analytics (e.g., clustering)
Strategic planning for matters based on cost	Quickly determine how much litigation will cost via early access to amount of potentially responsive information and prioritized review to make decisions based on the economics of the matter (e.g. settle for less than the cost of litigation)
Litigation budget optimization	Minimize litigation budget by only pursuing winnable cases Minimize litigation budget by utilizing the lowest cost resources possible while putting high-cost resource on only the necessary documents

Legal Hold Process

The legal hold process is a foundational element of IG.[16] The way the legal hold

process is supposed to work is that a formal system of polices, processes, and controls is put in place to notify key employees of a civil lawsuit (or impending one) and the set of documents that must put on legal hold. These documents, e-mail messages, and other relevant ESI must be preserved in place and no longer edited or altered so that they may be reviewed by attorneys during the discovery phase of the litigation. But, in practice, this is not always what takes place. In fact, *the opposite can take place*—employees can quickly edit or even delete relevant e-documents that may raise questions or even implicate them. This is possible only if proper IG controls are not in place, monitored, enforced, and audited.

Many organizations start with Legal Hold Notification (LHN). LHN management as a very discrete IG project. *LHN management is arguably the absolute minimum an organization should be doing* in order to meet the guidelines provided by court rules, common law, and case law precedent. It is worth noting, though, that the expectation is that organizations should connect the notification process to the actual collection and preservation of information in the long term.

> LHN management is the absolute minimum an organization should implement to meet the guidelines, rules, and precedents.

How to Kick-Start Legal Hold Notification

Implementing an LHN program attacks some of the lower-hanging fruit within an organization's overall IG position. *This part of the e-discovery life cycle must not be outsourced.* Retained counsel provides input, but the mechanics of LHN are managed and owned by internal corporate resources.

In preparing for a LHN implementation project, it is important to first lose the perception that LHN tools are expensive and difficult to deploy. It is true that some of these tools cost considerably more than others and can be complex to deploy; however, that is because the tools in question go far beyond simple LHN and reach into enterprise systems and also handle data mapping, collection, and work flow processes. Other options include Web-based hosted solutions, custom-developed solutions, or processes using tools already in the toolbox (e.g. e-mail, spreadsheets, word processing).

The most effective approach involves three basic steps:

1. Define requirements.
2. Define the ideal process.
3. Select the technology.

Defining both LHN requirements and processes should include input from key stakeholders—at a minimum—in legal, records management, and IT. Be sure to take into consideration the organization's litigation profile, corporate culture, and available resources as part of the requirements and process defining exercise.

Managing steps 1 and 2 thoroughly makes tool selection easier because defining requirements and processes creates the confidence of knowing exactly what the tool must accomplish.

IG and E-Discovery Readiness

Having a solid IG underpinning means that your organization will be better prepared to respond and execute key tasks when litigation and the e-discovery process proceed. Your policies will have supporting business processes, and clear lines of responsibility and accountability are drawn. The policies must be reviewed and fine-tuned periodically, and business processes must be streamlined and continue to aim for improvement over time.

In order for legal hold or **defensible deletion** (discussed in detail in the next section—disposing of unneeded data, e-documents, and reports based on set policy) projects to deliver the promised benefit to e-discovery, it is important to avoid the very real roadblocks that exist in most organization. To get the light to turn green at the intersection of e-discovery and IG, it is critical to:

- *Establish a culture that both values information and recognizes the risks inherent in it*. Every organization must evolve its culture from one of keeping everything to one of information compliance. This kind of change requires high-level executive support. It also requires constant training of employees about how to create, classify, and store information. While this advice may seem trite, many managers in leading organizations say that without this kind of culture change, IG projects tend to be dead on arrival.
- *Create a truly cross-functional IG team*. Culture change is not easy, but it can be even harder if the organization does not bring all stakeholders together when setting requirements for IG. Stakeholders include: legal; security and ethics; IT; records management; internal audit; corporate governance; human resources; compliance; and business units and employees. That is a lot of stakeholders. In organizations that are successfully launching and executing IG projects, many have dedicated IG teams. Some of those IG teams are the next generation of records management departments while others are newly formed. The stakeholders can be categorized into three areas: legal/risk, IT, and the business. The IG team can bring those areas together to ensure that any projects meet requirements of all stakeholders.
- *Use e-discovery as an IG proof of concept*. Targeted programs like e-discovery, compliance, and archiving have history of return on investment (ROI) and an ability to get budget. These projects are also challenging, but more straightforward to implement and can address subsets of information in early phases (e.g. only those information assets that are reasonable to account for). The lessons learned from these targeted projects can then be applied to other IG initiatives.
- *Measure ROI on more than just cost savings*. Yes, one of the primary benefits of addressing e-discovery via IG is cost reduction, but it is wise to begin

measuring all e-discovery initiatives on how they impact the life cycle of legal matters. The efficiencies gained in collecting information, for example, have benefits that go way beyond reduced cost; the IT time not wasted on reactive collection is more time available for innovative projects that drive revenue for companies. And a better litigation win rate will make any legal team happier.

IG serves as the underpinning for efficient e-discovery processes.

Building on Legal Hold Programs to Launch Defensible Disposition

By Barry Murphy

Defensible deletion programs can build on legal hold programs, because legal hold management is a necessary first step before defensibly deleting anything. The standard is "reasonable effort" rather than "perfection." Third-party consultants or auditors can support the diligence and reasonableness of these efforts.

Next, prioritize what information to delete and what information the organization is capably able to delete in a defensible manner. *Very few organizations are deleting information across all systems*. It can be overly daunting to try to apply deletion to all enterprise information. Choosing the most important information sources—e-mail, for example—and attacking those first may make for a reasonable and tenable approach. *For most organizations, e-mail is the most common information source to begin deleting*. Why e-mail? It is fairly easy for companies to put systematic rules on e-mail because the technology is already available to manage e-mail in a sophisticated manner. Because e-mail is such a critical data system, e-mail providers and e-mail archiving providers early on provided for systematic deletion or application of retention rules. However, in non–e-mail systems, the retention and deletion features are less sophisticated; therefore, organizations do not systematically delete across all systems.

Once e-mail is under control, the organization can begin to apply lessons learned to other information sources and eventually have better IG policies and processes that treat information consistently based on content rather than on the repository.

For most organizations, e-mail is the most common information source to begin deleting according to established retention policies.

Destructive Retention of E-Mail

A **destructive retention program** is an approach to e-mail archiving where e-mail messages are retained for a limited time (say, 90 days), followed by the

permanent manual or automatic deletion of the messages from the organization network, so long as there is no litigation hold or the e-mail has not been declared a record.

E-mail retention periods can vary from 90 days to as long as seven years:

- Osterman Research reports that "nearly one-quarter of companies delete e-mail after 90 days."
- Heavily regulated industries, including energy, technology, communications, and real estate, favor archiving for one year or more, according to Fulbright and Jaworski research.
- The most common e-mail retention period traditionally has been seven years; however, some organizations are taking a hard-line approach and stating that e-mails will be kept for only 90 days or six months, unless it is declared as a record, classified, and identified with a classification/retention category and tagged or moved to a repository where the integrity of the record is protected (i.e. the record cannot be altered and an audit trail on the history of the record's usage is maintained).

Destructive retention of e-mail is a method whereby e-mail messages are retained for a limited period and then destroyed.

Newer Technologies That Can Assist in E-Discovery

Few newer technologies are viable for speeding the document review process and improving the ability to be responsive to court-mandated requests. Here we introduce predictive coding and technology-assisted review (also known as computer-assisted review), the most significant of new technology developments that can assist in e-discovery.

Predictive Coding

During the **early case assessment** (ECA) phase of e-discovery, **predictive coding** is a "court-endorsed process"[17] utilized to perform document review. It uses human expertise and IT to facilitate analysis and sorting of documents. Predictive coding software leverages human analysis when experts review a subset of documents to "teach" the software what to look for, so it can apply this logic to the full set of documents,[18] making the sorting and culling process faster and more accurate than solely using human review or automated review.

Predictive coding uses a blend of several technologies that work in concert:[19] software that performs **machine learning** (a type of **artificial intelligence** software that "learns" and improves its accuracy, fostered by guidance from human input and progressive ingestion of data sets—in this case documents);[20]**workflow** software, which routes the documents through a series of work steps to be

processed; and **text analytics** software, used to perform functions such as searching for keywords (e.g., "asbestos" in a case involving asbestos exposure). Then, the next step is using **keyword search** capabilities, or *concepts* using **pattern search** or **meaning-based** search, and sifting through and sorting documents into basic groups using **filtering** technologies, based on document content, and **sampling** a portion of documents to find patterns and to review the accuracy of filtering and keyword search functions.

The goal of using predictive coding technology is to reduce the total group of documents a legal team needs to review manually (viewing and analyzing them one by one) by finding that gross set of documents that is most likely to be relevant or **responsive** (in legalese) to the case at hand. It does this by automating, speeding up, and improving the accuracy of the document review process to locate and "digitally categorize" documents that are responsive to a discovery request.[21] Predictive coding, when deployed properly, also reduces billable attorney and paralegal time and therefore the costs of ECA. Faster and more accurate completion of ECA can provide valuable time for legal teams to develop insights and strategies, improving their odds for success. Skeptics claim that the technology is not yet mature enough to render more accurate results than human review.

> Predictive coding software leverages human analysis when experts review a subset of documents to "teach" the software what to look for, so it can apply this logic to the full set of documents.

The first state court ruling allowing the use of predictive coding technology instead of human review to cull through approximately two million documents to "execute a first-pass review" was made in April 2012 by a Virginia state judge.[22] This was the first time a judge was asked to grant permission without the two opposing sides first coming to an agreement. The case, *Global Aerospace, Inc., et al. v. Landow Aviation, LP, et al.*, stemmed from an accident at Dulles Jet Center.

> In an exhaustive 156-page memorandum, which included dozens of pages of legal analysis, the defendants made their case for the reliability, cost-effectiveness, and legal merits of predictive coding. At the core of the memo was the argument that predictive coding "is capable of locating upwards of seventy-five percent of the potentially relevant documents and can be effectively implemented at a fraction of the cost and in a fraction of the time of linear review and keyword searching."[23]

This was the first big legal win for predictive coding use in e-discovery.

Basic Components of Predictive Coding

Here is a summary of the main foundational components of predictive coding.

- *Human review*. Human review is used to determine which types of document content will be legally responsive based on a case expert's review of a sampling of documents. These sample documents are fed into the system to provide a seed set of examples.[24]
- *Text analytics*. This involves the ability to apply "keyword-agnostic" (through a thesaurus capability based on contextual meaning, not just keywords) to locate responsive documents and build create seed document sets.
- *Workflow*. Software is used to route e-documents through the processing steps automatically to improve statistical reliability and streamlined processing.
- *Machine learning*. The software "learns" what it is looking for and improves its capabilities along the way through multiple, iterative passes.
- *Sampling*. Sampling is best applied if it is integrated so that testing for accuracy is an ongoing process. This improves statistical reliability and therefore defensibility of the process in court.

Predictive Coding Is the Engine; Humans Are the Fuel

Predictive coding sounds wonderful, but it does not replace the expertise of an attorney; it merely helps leverage that knowledge and speed the review process. It "takes all the documents related to an issue, ranks and tags them so that a human reviewer can look over the documents to confirm relevance." So it cannot work without human input to let the software know what documents to keep and which ones to discard, but it is an emerging technology tool that will play an increasingly important role in e-discovery.[25]

Technology-Assisted Review

TAR, also known as computer-assisted review, is *not* predictive coding. TAR includes aspects of the nonlinear review process, such as culling, clustering and de-duplication, but it does not meet the requirements for comprehensive predictive coding.

Many technologies can help in making incremental reductions in e-discovery costs. *Only fully integrated predictive coding, however, can completely transform the economics of e-discovery.*

Mechanisms of Technology-Assisted Review

There are three main mechanisms, or methods, for using technology to make legal review faster, less costly, and generally smarter.[26]

1. *Rules driven*. "I know what I am looking for and how to profile it." In this scenario, a case team creates a set of criteria, or rules, for document review and builds what is essentially a coding manual. The rules are fed into the tool for execution on the document set. For example, one rule might be to "redact for privilege any time XYZ term appears and add the term 'redacted' where the

data was removed." This rule-driven approach requires iteration to truly be effective. The case team will likely have rules changes and improvements as the case goes on and more is learned about strategy and merit. This approach assumes that the case team knows the document set well and can apply very specific rules to the corpus in a reasonable fashion.

2. *Facet driven*. "I let the system show me the profile groups first." In this scenario, a tool analyzes documents for potential items of interest or groups potentially similar items together so that reviewers can begin applying decisions. Reviewers typically utilize visual analytics that guide them through the process and take them to prioritized documents. This mechanism can also be called present and direct.
3. *Propagation based*. "I start making decisions and the system looks for similar-related items." This type of TAR is about passing along, or propagating, what is known based on a sample set of documents to the rest of the documents in a corpus. In the market, this is often referred to as predictive coding because the system predicts whether documents will be responsive or privileged based on how other documents were coded by the review team. Propagation-based TAR comes in different flavors, but all involve an element of machine learning. In some scenarios, a review team will have access to a seed set of documents that the team codes and then feeds into the system. The system then mimics the action of the review team as it codes the remainder of the corpus. In other scenarios, there is not a seed set; rather, the systems give reviewers random documents for coding and then create a model for relevance and nonrelevance. It is important to note that propagation-based TAR goes beyond simple mimicry; it is about creating a linguistic mathematical model for what relevance looks like.

These TAR mechanisms are not mutually exclusive. In fact, combining the mechanisms can help overcome the limitations of individual approaches. *For example, if a document corpus is not rich (e.g. does not have a high enough percentage of relevant documents), it can be hard to create a seed set that will be a good training set for the propagation-based system.* However, it is possible to use facet-based TAR—for example, concept searching—to more quickly find the documents that are relevant so as to create a model for relevance that the propagation-based system can leverage.[27]

It is important to be aware that these approaches require more than just technology. It is critical to have the right people in place to support the technology and the workflow required to conduct TAR. Organizations looking to exercise these mechanisms of TAR will need:

- *Experts in the right tools and information retrieval*. Software is an important part of TAR. The team executing TAR will need someone that can program the tool set with the rules necessary for the system to intelligently mark documents. Furthermore, information retrieval is a science unto itself, blending linguistics, statistics, and computer science. Anyone practicing TAR

will need the right team of experts to ensure a defensible and measurable process.

- *Legal review team*. While much of the chatter around TAR centers on its ability to cut lawyers out of the review process, the reality is that the legal review team will become more important than ever. The quality and consistency of the decisions this team makes will determine the effectiveness that any tool can have in applying those decisions to a document set.
- *Auditor*. Much of the defensibility and acceptability of TAR mechanisms will rely on the statistics behind how certain the organization can be that the output of the TAR system matches the input specification. Accurate measures of performance are important not only at the end of the TAR process, but also throughout the process in order to understand where efforts need to be focused in the next cycle or iteration. Anyone involved in setting or performing measurements should be trained in statistics.

For an organization to use a propagated approach, in addition to people it may need a "seed" set of known documents. Some systems use random samples to create seed sets while others enable users to supply small sets from the early case investigations. These documents are reviewed by the legal review team and marked as relevant, privileged, and so on. Then, the solution can learn from the seed set and apply what it learns to a larger collection of documents. Often this seed set is not available, or the seed set does not have enough positive data to be statistically useful.

Professionals using TAR state that the practice has value, but it requires a sophisticated team of users (with expertise in information retrieval, statistics, and law) who understand the potential limitations and danger of false confidence that can arise from improper use. For example, using a propagation-based approach with a seed set of documents can have issues when less than 10% of the seed set documents are positive for relevance. In contrast, rules driven and other systems can result in false negative decisions when based on narrow custodian example sets.

However TAR approaches and tools are used, they will only be effective if usage is anchored in a thought out, methodically sound process. This requires a definition of what to look for, searching for items that meet that definition, measuring results, and then refining those results on the basis of the measured results. Such an end-to-end plan will help to decide what methods and tools should be used in a given case.[28]

Defensible Disposal: The Only Real Way to Manage Terabytes and Petabytes

By Randy Kahn, Esq.

Records and information management (RIM) is not working. At least, it is *not working well*. Information growth and management complexity has meant that

the old records retention rules and the ways businesses apply them are no longer able to address the life cycle of information. So the mountains of information grow and grow and grow, often unfettered.

Too much data has outlived its usefulness, and no one seems to know how or is willing to get rid of it. While most organizations need to right-size their information footprint by cleaning out the digital data debris, they are stymied by the complexity and enormity of the challenge.

Growth of Information

According to *International Data Corporation* (IDC, from now until 2020, the digital universe is expected by expand to more than 14 times its current size.[29] One exabyte is the data equivalent of about 50,000 years of DVD movies running continuously. With about 1,800 exabytes of new data created in 2011, 2840 exabytes in 2012, and a predicted 6,120 exabytes in 2014, the volumes are truly staggering. While the data footprint grows significantly each year, that says nothing of what has already been created and stored.

Contrary to what many say (especially hardware salespeople) storage is *not* cheap. In fact, it becomes quite expensive when you add up not only the hardware costs but also maintenance, air conditioning, and space overhead, and the highly skilled labor needed to keep it running. Many large companies spend tens if not hundreds of millions of dollars per year just to store data. This is money that could go straight to the bottom line if the unneeded data could be discarded. When you consider that most organizations' information footprints are growing at between 20 and 50% per year and the cost of storage is declining by a few percentage points per year, in real terms they are spending way more this year than last to simply house information.

Volumes Now Impact Effectiveness

The law of diminishing returns applies to information growth. Assuming information is an asset, at some point when there is so much data, its value starts to decline. That is not because the intrinsic value goes down (although many would argue there is a lot of idle chatter in the various communications technologies). Rather the decline is related to the inability to expeditiously find or have access to needed business information. According the Council of Information Auto-Classification "Information Explosion" Survey, there is now so much information that nearly 50% of companies need to re-create business records to run their business and protect their legal interests because they cannot find the original retained record.[30] It is a poor business practice to spend resources to retain information and then, when it cannot be found, to spend more to reconstitute it.

There is increasing regulatory pressure, enforcement, and public scrutiny on all of an organization's data storage activities. Record sanctions and fines, new regulations, and stunning court decisions have converged to mandate heightened controls and accountability from government regulators, industry and standards

groups as well as the public. When combined with the volume of data, information privacy, security, protection of trade secrets, and records compliance become complex and critical, high-risk business issues that only executive management can truly fix. However, executives typical view records and information management (RIM) as a low-importance cost center activity, which means that the real problem does not get solved.

In most companies, there is no clear path to classify electronic records, to formally manage official records, or to ensure the ultimate destruction of these records. Vast stores of legacy data are unclassified, and most data is never touched again shortly after creation. Further, traditional records retention rules are too voluminous, too complex, and too granular and do not work well with the technology needed to manage records.

Finally, it is clear that employees can no longer be expected to pull the oars to cut through the information ocean, let alone boil it down into meaningful chunks of good information. Increasingly, technology has to play a more central role in managing information. Better use of technology will create business value by reducing risk, driving improvements in productivity, and facilitating the exploitation and protection of ungoverned corporate knowledge.

How Did This Happen?

Over the past several years, organizations have come to realize that the exposure posed by uncontrolled data growth requires emergency, reactive action, as seemingly no other viable approach exists. Faced with massive amounts of unknown unstructured data, many organizations have chosen to adopt a risk-averse save-everything policy. This approach has brought with it immediate repercussions:

- Inability to quickly locate needed business content buried in ill-managed file systems.
- Sharply increased storage costs, with some companies refusing to allocate any more storage to the business. The user reaction, out of necessity, is to store data wherever they can find a place for it. (Do *not* buy the argument that storage is cheap—everyone is spending more on storing unnecessary data, even if the per-gigabyte media cost has gone down.)
- Soaring litigation and discovery costs, as organizations have lost track of what is where, who owns it, and how to collect, sort, and process it.
- Buried intellectual property, trade secrets, personally identifiable information, and regulated content is subject to leakage and unauthorized deletion, and are a clear target for opposing counsel—or anyone who can access it.
- Lack of centralized policies and systems for the storage of records, which results in hard-to-manage record sites spread throughout the organization.
- The lack of a clear strategy for managing records that have long-term, rather than short-term, business, legal, and research value.

Information Glut in Organizations

- 71% of organizations surveyed have no idea of the content in their stored data.
- 58% of organizations are keeping information indefinitely.
- 79% of organizations say too much time and effort is spent manually searching and disposing of information.
- 58% of organizations still rely on employees to decide how to apply corporate policies.[31]

What Is Defensible Disposition, and How Will It Help?

A solution to the unmitigated data sprawl is to defensibly dispose of the business content that no longer has business or legal value to the organization. In the old days of records management, it was clear that courts and regulators alike understood that records came into being and eventually were destroyed in the ordinary course of business. It is good business practice to destroy unneeded content, provided that the rules on which those decisions are made consider legal requirements and business needs. Today, however, the good business practice of cleaning house of old records has somehow become taboo for some businesses. Now it needs to start again.

An understanding of how technology can help defensibly dispose and how methodology and process help an organization achieve a thinner information footprint is critical for all companies overrun with outdated records that do not know where to start to address the issue. While no single approach is right for every organization, records and legal teams need to take an informed approach, looking at corporate culture, risk tolerance, and litigation profile.

A defensible disposition framework is an ecosystem of technology, policies, procedures, and management controls designed to ensure that records are created, managed, and disposed at the end of their life cycle.

New Technologies—New Information Custodians

Responsibility for records management and IG has changed dramatically over time. In the past, the responsibility rested primarily with the records manager. However, the nature of electronic information is such that its governance today requires the participation of IT, which frequently has custody, control, or access to such data, along with guidance from the legal department. As a result, IT personnel with no real connection or ownership of the data may be responsible for the accuracy and completeness of the business-critical information being managed. See the problem?

For many organizations advances in technology mixed with an explosive growth of data forced a reevaluation of core records management processes. Many organizations have deployed archiving, litigation, and e-discovery point solutions with the intent of providing record retention compliance and responsiveness to litigation. Such systems may be tactically useful but fail to strategically address the

heart of the matter: too much information, poorly managed over years and years—if not decades.

> A defensible disposition framework is an ecosystem of technology, policies, procedures, and management controls designed to ensure that records are created, managed, and disposed at the end of their life cycle.

A better approach is for organizations to move away from a reactive keep-everything strategy to a proactive strategy that allows the reasonable and reliable identification and deletion of records when retention requirements are reached, absent a preservation obligation. Companies develop retention schedules and processes precisely for this reason; it is not misguided to apply them.

Why Users Cannot, Will Not—and Should Not—Make the Hard Choices

Employees usually are not sufficiently trained on records management principles and methods and have little incentive (or downside) to properly manage or dispose of records. Further, many companies today see that requiring users to properly declare or manage records places an undue burden on them. The employees not only do not provide a reasonable solution to the huge data pile (which for some companies may be petabytes of data) but contribute to its growth by using more unsanctioned technologies and parking company information in unsanctioned locations. This is how the digital landfill continues to grow.

Most organizations have programs that address paper records, but these same organizations commonly fail to develop similar programs for electronic records and other digital content.

> A better approach is for organizations to move away from a reactive keep-everything strategy to a proactive strategy of defensible deletion.

Technology Is Essential to Manage Digital Records Properly

Having it all—but not being able to find it—is like *not* having it at all.

While the content of a paper document is obvious, viewing the content of an electronic document depends on software and hardware. Further, the content of electronic storage media cannot be easily accessed without some clue as to its structure and format. Consequently, *the proper indexing of digital content is fundamental to its utility*. Without an index, retrieving electronic content is expensive and time consuming, if it can be retrieved at all.

Search tools have become more robust, but they do not provide a panacea for finding electronic records when needed because there is too much information

spread out across way too many information parking lots. Without **taxonomies** and common business terminology, accessing the one needed business record may be akin to finding the needle in a stadium-size haystack.

Technological advances can help solve the challenges corporations face and address the issues and burdens for legal, compliance. and information governance. When faced with hundreds of terabytes to petabytes of information, no amount of user intervention will begin to make sense of the information tsunami.

Auto-Classification and Analytics Technologies

Increasingly companies are turning to new analytics and classification technologies that can analyze information faster, better, and cheaper. These technologies should be considered essential for helping with defensible disposition, but do not make the mistake of underestimating their expense or complexity.

Machine learning technologies mean that software can "learn" and improve at the tasks of clustering files and assigning information (e.g. records, documents) to different preselected topical categories based on a statistical analysis of the data characteristics. In essence, classification technology evaluates a set of data with known classification mappings and attempts to map newly encountered data within the existing classifications. This type of technology should be on the list of considerations when approaching defensible disposition in large, uncontrolled data environments.

Can Technology Classify Information?

What is clear is that IT is better and faster than people in classifying information. Period.

Increasingly studies and court decisions make clear that, when appropriate, companies should not fear using enabling technologies to help manage information.

For example, in the recent *Da Silva Moore v. Publicis Groupe* case, Judge Andrew Peck stated:

> Computer-assisted review appears to be better than the available alternatives, and thus should be used in appropriate cases. While this Court recognizes that computer-assisted review is not perfect, the Federal Rules of Civil Procedure do not require perfection.... Counsel no longer have to worry about being the "first" or "guinea pig" for judicial acceptance of computer assisted review.
>
> This work presents evidence supporting the contrary position: that a technology-assisted process, in which only a small fraction of the document collection is ever examined by humans, can yield higher recall and/or precision than an exhaustive manual review process, in which the entire document collection is examined and coded by humans.[32]

Moving Ahead by Cleaning Up the Past

Organizations can improve disposition and IG programs with a systemized, repeatable, and defensible approach that enables them to retain and dispose of all data types in compliance with the business and statutory rules governing the business's operations.

Generally, an organization is under no legal obligation to retain every piece of information it generates in the course of its business. Its records management process is there to clean up the information junk in a consistent, reasonable way. That said, what should companies do if they have not been following disposal rules, so information has piled up and continues unabated? They need to clean up old data. *But how?*

Manual intervention (by employees) will likely not work, due to the sheer volumes of data involved. Executives will not and should not have employees abdicate their regular jobs in favor of classifying and disposing of hundreds of millions of old stored files. (Many companies have billions of old files.) *This buildup necessitates leveraging technology, specifically, technologies that can discern the meaning of stored unstructured content, in a variety of formats, regardless of where it is stored.*

Organizations can improve disposition and IG programs with a systemized, repeatable, and defensible approach.

Here is a starting point: most likely, file shares, legacy e-mail systems, and other large repositories will prove the most target-rich environments, while better-managed document management, records management, or archival systems will be in less need of remediation. A good time to undertake a cleanup exercise is when litigation will not prevent action or when migrating to a new IT platform. (Trying to conduct a comprehensive, document-level inventory and disposition is neither reasonable nor practical. In most cases, it will create limited results and even further frustration.)

Technology choices should be able to withstand legal challenges in court. Sophisticated technologies available today should also look beyond mere keyword searches (as their defensibility may be called into question) and should look to advanced techniques such as automatic text classification (auto-classification), concept search, contextual analysis, and automated clustering. While technology is imperfect, it is better than what employees can do and will never be able to accomplish—to manage terabytes of stored information and clean up big piles of dead data.

Defensibility Is the Desired End State; Perfection Is Not

Defensible disposition is a way to take on huge piles of information without personally cracking each one open and evaluating it. Perhaps it is, in essence, operationalizing a retention schedule that is no longer viable in the electronic age.

Defensible disposition is a must because most big companies have hundreds of millions or billions of files, which makes their individualized management all but impossible.

As the list of eight steps to defensible disposition makes clear, different chunks of data will require different diligence and analysis levels. If you have 100,000 backup tapes from 20 years ago, minimal or cursory review may be required before the whole lot of tapes can be comfortably discarded. If, however, you have an active shared drive with records and information that is needed for ongoing litigation, there will need to be deeper analysis with analytics and/or classification technologies that have become much more powerful and useful. In other words, the facts surrounding the information will help inform if the information can be properly disposed with minimal analysis or if it requires deep diligence.

Kahn's Eight Essential Steps to Defensible Disposition

1. Define a reasonable diligence process to assess the business needs and legal requirements for continued information retention and/or preservation, based on the information at issue.
2. Select a practical information assessment and/or classification approach, given information volumes, available resources, and risk profile.
3. Develop and document the essential aspects of the disposition program to ensure quality, efficacy, repeatability, auditability, and integrity.
4. Develop a mechanism to modify, alter, or terminate components of the disposition process when required for business or legal reasons.
5. Assess content for eligibility for disposition, based on business need, record retention requirements, and/or legal preservation obligations.
6. Test, validate, and refine as necessary the efficacy of content assessment and disposition capability methods with actual data until desired results have been attained.
7. Apply disposition methodology to content as necessary, understanding that some content can be disposed with sufficient diligence without classification.
8. On an ongoing basis, verify and document the efficacy and results of the disposition program and modify and/or augment the process as necessary.

Business Case Around Defensible Disposition

What is clear is that defensible disposition can have significant ROI impact to a company's financial picture. This author has clients for whom we have built the defensible disposition business case, which saves them tens of millions of dollars on a net basis but also makes them a more efficient business, reduces litigation cost and risks, mitigates the information security and privacy risk profiles, and makes their work force more productive, etc.

However, remember auto-classification technology is neither simple nor

inexpensive, so be realistic and conservative when building the business case. Often it is easiest to simply use only hardware storage cost savings to make the case because it is a hard number and provides a conservative approach to justifying the activities. Then you can add on the additional benefits, which are more difficult to calculate, and also the intangible benefits of giving your employees a cleaner information stack to search and base decisions on.

Defensible Disposition Summary

Defensible disposition is a way to bring your records management program into today's business reality—information growth makes management at the record level all but impossible. Defensible disposition should be about taking simplified retention rules and applying them to both structured and unstructured content with the least amount of human involvement possible. While it can be a daunting challenge, it is also an opportunity to establish and promote operational excellence through better IG and to significantly enhance an organization's business performance and competitive advantage.

CHAPTER SUMMARY: KEY POINTS

- Legal functions are the most important area of IG impact.
- IG serves as the underpinning for efficient e-discovery processes.
- ESI is any information that is created or stored in electronic format.
- The goal of the FRCP amendments is to recognize the importance of ESI and to respond to the increasingly prohibitive costs of document review and protection of privileged documents.
- The further revisions made in 2015 emphasized the importance of "proportionality" in litigation, as well as setting a more uniform standard for when courts may impose "curative measures," including sanctions, in discovery.
- The amended FRCP reinforce the importance of IG. Only about 25% of business information has real value, and 5% are business records.
- The Big Data trend underscores the need for defensible deletion of data debris.
- In the landmark case *Zubulake v. UBS Warburg*, the defendants were severely punished by an adverse inference for deleting key e-mails and not producing copies on backup tapes.
- The E-Discovery Reference Model is a planning tool that depicts key e-discovery process steps.
- Implementing IG, inventorying ESI, and leveraging technology to implement records retention and LHN policies are key steps in e-discovery planning.

- LHN management is the absolute minimum an organization should implement to meet the guidelines, rules, and precedents.
- Predictive coding software leverages human analysis when experts review a subset of documents to "teach" the software what to look for, so it can apply this logic to the full set of documents.
- Many technologies assist in making incremental reductions in e-discovery costs, but only fully integrated predictive coding is able to completely transform the economics of e-discovery.
- TAR, also known as computer-assisted review, speeds the review process by leveraging IT tools.
- In TAR, there are three main ways to use technology to make legal review faster, less costly, and generally smarter: rules driven, facet driven, and propagation based.
- It is important to have the right people in place to support the technology and the workflow required to conduct TAR.
- A defensible disposition framework is an ecosystem of technology, policies, procedures, and management controls designed to ensure that records are created, managed, and disposed of at the end of their life cycle.
- A better approach is for organizations to move away from a reactive "keep-everything" strategy to a *proactive strategy* of defensible deletion.
- Organizations can improve disposition and IG programs with a systemized, repeatable, and defensible approach.

Notes

1. Linda Volonino and Ian Redpath, *e-Discovery For Dummies* (Hoboken, NJ: John Wiley & Sons, 2010), 9.
2. "New Federal Rules of Civil Procedure," http://www.uscourts.gov/FederalCourts/Understanding-theFederalCourts/DistrictCourts.aspx (accessed November 26, 2013).
3. Ibid.
4. Ibid.
5. Volonino and Redpath, *e-Discovery For Dummies*, p. 13.
6. Ibid., p. 11.
7. "New Federal Rules of Civil Procedure."
8. "The Digital Universe Decade—Are You Ready?" IDC iView (May 2010).

https://www.cio.com.au/campaign/370004?content=%2Fwhitepaper%2F370012%2Fthe-digital-universe-decade-are-you-ready%2F%3Ftype%3Dother%26arg%3D0%26location%3Dfeatured_list.

9. Deidra Paknad, "Defensible Disposal: You Can't Keep All Your Data Forever," July 17, 2012, www.forbes.com/sites/ciocentral/2012/07/17/defensible-disposal-you-cant-keep-all-your-data-forever/.

10. Sunil Soares, *Selling Information Governance to the Business* (Ketchum, ID: MC Press Online, 2011), 229.

11. All quotations from the FRCP are from Volonino and Redpath, *e-Discovery For Dummies*, www.dummies.com/how-to/content/ediscovery-for-dummies-cheat-sheet.html (accessed May 22, 2013).

12. Linda Volonino and Ian Redpath, *e-Discovery For Dummies* (Hoboken, NJ: John Wiley & Sons, 2010), p. 11.

13. Case Briefs, LLC, "*Zubulake v. UBS Warburg LLC*," http://www.casebriefs.com/blog/law/civil-procedure/civil-procedure-keyed-to-friedenthal/pretrial-devices-of-obtaining-information-depositions-and-discovery-civil-procedure-keyed-to-friedenthal-civil-procedure-law/zubulake-v-ubs-warburg-llc/2/ (accessed May 21, 2013).

14. Amy Girst, "E-discovery for Lawyers," IMERGE Consulting Report, 2008.

15. ECM^2, "15-Minute Guide to eDiscovery and Early Case Assessment," www.emc.com/collateral/15-min-guide/h9781-15-min-guide-ediscovery-eca-gde.pdf (accessed May 21, 2013).

16. Barry Murphy, telephone interview with author, April 12, 2013.

17. Recommind, "What Is Predictive Coding?" www.recommind.com/predictive-coding (accessed May 7, 2013).

18. Michael LoPresti, "What Is Predictive Coding?: Including eDiscovery Applications," KMWorld, January 14, 2013, www.kmworld.com/Articles/Editorial/What-Is-.../What-is-Predictive-Coding-Including-eDiscovery-Applications-87108.aspx.

19. "Predictive Coding," TechTarget.com, August 31, 2012, http://searchcompliance.techtarget.com/definition/predictive-coding (accessed May 7, 2013).

20. "Machine Learning," TechTarget.com http://whatis.techtarget.com/definition/machine-learning (accessed May 7, 2013).

21. "Predictive Coding."

22. LoPresti, "What Is Predictive Coding?"

23. Ibid.

24. Recommind, "What Does Predictive Coding Require?" www.recommind.com/predictive-coding (accessed May 24, 2013).

25. Ibid.

26. Barry Murphy, e-mail to author, May 10, 2013.

27. Ibid.

28. Ibid.

29. "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf (accessed November 26, 2013).

30. Council of Information Auto-Classification, "Information Explosion" survey, http://infoautoclassification.org/survey.php (accessed November 26, 2013).

31. Ibid.

32. Maura R. Grossman and Gordon V. Cormack, "Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review." http://delve.us/downloads/Technology-Assisted-Review-In-Ediscovery.pdf (accessed November 26, 2013).

CHAPTER 9
Information Governance and Records and Information Management Functions *

Records and information management (RIM) is a key impact area of **information governance** (IG)—so much so that in the records management (RM) space, IG is often thought of as synonymous with or a simple superset of RM. But IG is much more than that. We will delve into the details of RM here—a sort of crash course on how to identify and inventory records, conduct the necessary legal research, develop retention and disposition schedules, and more. Also, we identify the relationship and impact of IG on the RM function in an organization in this chapter.

The International Organization for Standardization (ISO) defines (business) *records* as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business."[1] It further defines RM as "[the] field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use, and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records."[2]

RIM extends beyond RM (although the terms are often used interchangeably) to include information—that is, information such as e-mail, electronic documents, and reports. For this reason, RIM professionals must expand their reach and responsibilities to include policies for retention and disposition of all legally discoverable forms of information, RIM professionals today generally know that "everything is discoverable" and that includes e-mail, voicemail, social media posts, mobile data and documents held on portable devices, cloud storage and applications, and other enterprise data and information.

Electronic records management (ERM) has moved to the forefront of business issues with the increasing automation of business processes and the vast growth in the volume of electronic information that organizations create. These factors, coupled with expanded and tightened reporting laws and compliance regulations—most especially the EU **General Data Protection Regulation** (GDPR) and **California Consumer Privacy Act** (CCPA)—have made ERM essential for most enterprises, especially highly regulated and public ones.

ERM follows generally the same principles as traditional paper-based records management: There are **classification** and **taxonomy** needs to group and organize the records; and there are **retention** and **disposition** schedules to govern the length of time a record is kept and its ultimate disposition, which is usually destruction but can also include transfer (e.g., U.S. federal agency records sent to the national Archives and Records Administration for archiving and safekeeping) or long-term archiving. Yet e-records must be handled differently, and they contain more detailed data about their contents and characteristics, known as **metadata**. (For more detail on these topics see Appendix A.)

E-records are also subject to changes in **information technology** (IT) like file formats and protocols that may make them difficult to retrieve and view and therefore render them obsolete. These issues can be addressed through a sound ERM program that includes **long-term digital preservation** (LTDP) methods and technologies for digital records needed to be maintained 10 years or more.

ERM is primarily the organization, management, control, monitoring, and auditing of formal business records that exist in electronic form. But automated ERM systems also track paper-based and other physical records. So ERM goes beyond simply managing electronic records; it is *the management of electronic records and the electronic management of nonelectronic records (e.g. paper, CD/DVDs, magnetic tape, audio-visual, and other physical records)*.

> E-records management has become much more critical to enterprises with increased compliance legislation and massively increasing volumes of electronic information.

Most electronic records, or e-records, originally had an equivalent in paper form, such as memos (now e-mail), accounting documents (e.g. purchase orders, invoices), personnel documents (e.g. job applications, resumes, tax documents), contractual documents, line-of-business documents (e.g. loan applications, insurance claim forms, health records), and required regulatory documents (e.g. material safety data sheets). Before e-document and e-record software began to mature in the 1990s, many of these documents were first archived to microfilm or microform/microfiche.

> ERM follows the same basic principles as paper-based records management.

Not all documents rise to the level of being declared a formal business record that needs to be retained; that definition depends on the specific regulatory and legal requirements imposed on the organization and the internal definitions and

requirements the organization imposes on itself, through internal IG measures and business policies. IG *is control of information to meet legal, regulatory, business and risk demands*. In short, IG is security, control, and optimization of information.

ERM is a component of enterprise content management (ECM), just as document management, Web content management, digital asset management, enterprise report management, workflow, and several other technology sets are components. ECM encompasses *all* an organization's unstructured digital content, which means it excludes structured data (i.e. databases). ECM includes the vast majority—typically 80 to 90%—of an organization's overall information that must be governed and managed. Structured information held in databases makes up the remainder; however, due to its structured and consistent nature it is more easily managed.

> ERM includes the management of electronic and nonelectronic records, such as paper and other physical records.

ERM extends ECM to provide control and to manage records through their life cycle—from creation to destruction. ERM is used to complete the life cycle management of information, documents, and records.

ERM adds the functionality to complete the management of information and records by applying business rules to manage the maintenance, preservation, and disposition of records. Both ERM and ECM systems aid in locating and managing the records and information needed to conduct business efficiently, to comply with legal and regulatory requirements, and to effectively destroy (paper) and delete (digital) records that have met their retention policy time frame requirement, freeing up valuable physical and digital space and eliminating records that could be a liability if kept.

In the last few years, the term ***content services*** has been used to supplant and expand the definition of ECM, mostly to include cloud-based platforms that offer Software-as-a-Service tools to manage content. This renaming effort has been led by Gartner, and many see it as a necessary recharacterization of a market that has evolved.

Records Management Business Rationale

Historically, highly regulated industries, such as banking, energy, and pharmaceuticals, have had the greatest need to implement RM programs, due to their compliance and reporting requirements. However, over the past decade or so, increased regulation and changes to legal statutes and rules have made RM a business necessity for nearly every enterprise (beyond very small businesses).

Notable industry drivers fueling the growth of RM programs include:

- *Increased government oversight and industry regulation*. Government regulations that require enhanced reporting and accountability were early business drivers that fueled the implementation of formal RM programs. This is true at the federal and state or provincial level. In the United States, the Sarbanes-Oxley Act of 2002 (SOX) created and enhanced standards of financial reporting and transparency for the boards and executive management of public corporations and accounting firms. It also addressed auditor independence and corporate governance concerns. SOX imposes fines or imprisonment penalties for noncompliance and requires that senior officers sign off on the veracity of financial statements. It states clearly that pertinent business records cannot be destroyed during litigation or compliance investigations. Since SOX was enacted, Japan, Australia, Germany, France, and India also have adopted stricter “SOX-like” governance and financial reporting standards. Newer legislation, such as the EU GDPR, and global privacy concerns have further driven the need for updated and expanded RM programs. This has given the RM profession a boost in visibility and energy.
- *Changes in legal procedures and requirements during civil litigation*. In 2006, the need to amend the US Federal Rules of Civil Procedure (FRCP) to contain specific rules for handling electronically generated evidence was addressed. The changes included processes and requirements for legal discovery of **electronically stored information** (ESI) during civil litigation. *Today, e-mail is the leading form of evidence requested in civil trials*. The changes to the US FRCP in 2006 and its update in 2015 had a pervasive impact on American enterprises and required them to gain control over their ESI and implement formal RM and electronic discovery (e-discovery) programs to meet new requirements. Although they have generally been ahead of the United States in their development and maturity of RM practices, Canadian, British, and Australian law is closely tracking that of the United States in legal discovery. The United States is a more litigious society, so this is not unexpected.
- *IG awareness*. IG, in sum, is the set of rules, policies, and business processes used to manage and control the totality of an organization's information. Monitoring technologies are required to enforce and audit IG compliance. Beginning with SOX in 2002 and continuing with the massive US FRCP changes in 2006 and 2015, along with the introduction of the far-reaching GDPR legislation, enterprises have been forced to become more IG aware and have spurred ramped up efforts to control, manage, and secure their information. *A significant component of any IG program is implementing an RM program that specifies the retention periods and disposition (e.g. destruction, transfer, archive) of formal business records*. This program, for instance, allows enterprises to destroy records once their required retention period (based on external regulations, legal requirements, and internal IG policies)

has been met and allows them to legally destroy records with no negative impact or lingering liability. This practice, if consistently implemented, allows for the elimination of information that has low value to make room for new information that has higher business value.

- *Business continuity concerns*. In the face of real disasters, such as the 9/11 terrorist attacks, Hurricane Katrina, and Superstorm Sandy, executives now realize that disaster recovery and business resumption is something they must plan and prepare for. Disasters really happen, and businesses that are not well prepared really go under. The focus is on **vital records,** which are those most mission-critical records that are needed to resume operations in the event of a disaster, and managing those records is part of an overall RM program.

A number of factors provide the business rationale for ERM, including facilitating compliance, supporting IG, and providing backup capabilities in the event of a disaster.

Why Is Records Management So Challenging?

With these changes in the business environment and in regulatory, legal, and IG influences comes increased attention to RM as a driver for **corporate compliance**. For most organizations, a lack of defined policies and the enormous and growing volumes of documents (e.g. e-mail messages) make implementing a formal RM program challenging and costly. Some reasons for this include:

- *Changing and increasing regulations*. Just when records and compliance managers have sorted through the compliance requirements of federal regulations, new ones at the state or provincial level are created or tightened down, and even those from other countries, such as GDPR, have had an impact.
- *Maturing IG requirements within the organization*. As senior managers become increasingly aware of the value of IG programs—the rules, policies, and processes that control and manage information—they promulgate more reporting and auditing requirements for the management of formal business records.
- *Managing multiple retention and disposition schedules*. Depending on the type of record, retention requirements vary, and they may vary for the same type of record based on state and federal regulations. Further, internal IG policies may extend retention periods and may fluctuate with management changes.[3]
- *Compliance costs and requirements with limited staff*. RM and compliance departments are notoriously understaffed, since they do not generate revenue. Departments responsible for executing and proving compliance with new and increasing regulatory requirements must do so expediently, often with only skeletal staffs. This leads to expensive outsourcing solutions or staff increases. The cost of compliance must be balanced with the risk of maintaining a minimum level of compliance.
- *Changing information delivery platforms*. With cloud computing, mobile computing, Web 2.0, social media, and other changes to information delivery and storage platforms, records and compliance managers must stay apprised of the latest IT trends and provide records on multiple platforms all while maintaining the security and integrity of organizational records.
- *Security concerns*. Protecting and preserving corporate records is of paramount importance, yet users must have reasonable access to official records to conduct everyday business. “Organizations are struggling to balance the need to provide accessibility to critical corporate information with the need to protect the integrity of corporate records.”[4]
- *Dependence on the IT department or provider*. Since tracking and auditing use of formal business records requires IT, and records and compliance departments typically are understaffed, those departments must rely on assistance from the IT department or outsourced IT provider—which often does not have the same perspective and priorities as the departments they serve.
- *User assistance and compliance*. Users often go their own way with regard to records, ignoring directives from records managers to stop storing shadow files of records on their desktop (for their own convenience) and inconsistently following directives to classify records as they are created. Getting users across a range of departments in the enterprise to adhere uniformly with records and compliance requirements is a daunting and unending task that requires constant attention and reinforcement. Increasingly, the solution is to automate the task of classification as much as possible using file analysis tools, often enabled with artificial intelligence (AI).

Implementing ERM is challenging because it requires user support and compliance, adherence to changing laws, and support for new information delivery platforms, such as mobile and cloud computing.

Benefits of Electronic Records Management

A number of business drivers and benefits combine to create a strong case for implementing an enterprise ERM program. Most are tactical, such as cost savings, time savings, and building space savings. *But some drivers can be thought of as strategic*, in that they proactively give the enterprise an advantage. One example may be the advantages gained in litigation by having more control and ready access to complete business records. This yields more accurate results and more time for corporate attorneys to develop strategies while the opposition is placing blunt legal holds on entire job functions and wading through reams of information, never knowing if it has found the complete set of records it needs. Another example is more complete and better information for managers to base decisions on. Further, applying the principles of infonomics may help organizations find new value or even to monetize information.

Implementing ERM represents a significant investment. *An investment in ERM is an investment in business process automation and yields document control, document integrity/trustworthiness, and security benefits*. The volume of records in organizations often exceeds employees' ability to manage them. ERM systems do for the information age what the assembly line did for the industrial age. The cost/benefit justification for ERM is sometimes difficult to determine, although there are real labor and cost savings. Also, many of the benefits are intangible or difficult to calculate but help to justify the capital investment. There are many ways in which an organization can gain significant business benefits with ERM.

More detail on business benefits is provided in Chapter 7, but hard, calculable benefits (when compared to storing paper files) include office space savings, office supplies savings, cutting wasted search time, and reduced office automation costs (e.g. fewer printers, copiers, and automated filing cabinets).

An investment in ERM is an investment in business process automation and yields document control, document integrity, and security benefits.

In addition, implementing ERM will provide the organization with:

- Improved capabilities for enforcing IG policy over business documents and records.
- Increased working confidence in making searches, which should improve decision making.
- Improved knowledge worker productivity.
- Reduced risk of compliance actions or legal consequences.
- Improved records security.
- Improved ability to demonstrate legally defensible RM practices.
- More professional work environment.

ERM benefits are both tangible and intangible or difficult to calculate.

Additional Intangible Benefits

The US Environmental Protection Agency (EPA), a pioneer and leader in e-records implementation in the federal sector, lists some additional benefits of implementing ERM:

1. *To control the creation and growth of records*. Despite decades of using various nonpaper storage media, the amount of paper in our offices continues to escalate. An effective records management program addresses both creation control (limits the generation of records or copies not required to operate the business) and records retention (a system for destroying useless records or retiring inactive records), thus stabilizing the growth of records in all formats.
2. *To assimilate new records management technologies*. A good records management program provides an organization with the capability to assimilate new technologies and take advantage of their many benefits. Investments in new computer systems don't solve filing problems unless current manual recordkeeping systems are analyzed (and occasionally, overhauled) before automation is applied.
3. *To safeguard vital information*. Every organization, public or private, needs a comprehensive program for protecting its vital records and information from catastrophe or disaster, because every organization is vulnerable to loss. Operated as part of the overall records management program, vital records programs preserve the integrity and confidentiality of the most important records and safeguard the vital information assets according to a "plan" to protect the records.
4. *To preserve the corporate memory*. An organization's files contain its institutional memory, an irreplaceable asset that is often overlooked. Every business day, you create the records that could become background data for future management decisions and planning. These records document the activities of the agency that future scholars may use to research the workings of the Environmental Protection Agency.

5. *To foster professionalism in running the business.* A business office with files askew, stacked on top of file cabinets and in boxes everywhere, creates a poor working environment. The perceptions of customers and the public, and "image" and "morale" of the staff, though hard to quantify in cost-benefit terms, may be among the best reasons to establish a good records management program.[5]

Thus, there are a variety of tangible and intangible benefits derived from ERM programs, and the business rationale that fits for your organization depends on its specific needs and business objectives.

> Improved professionalism, preserving corporate memory, and support for better decision making are key intangible benefits of ERM.

Inventorying E-Records

According to the US National Archives and Records Administration (NARA), "In **records management**, an *inventory* is a descriptive listing of each record series or system, together with an indication of location and other pertinent data. *It is not a list of each document or each folder but rather of each series or system*"[6] (emphasis added).

Conducting an inventory of electronic records is more challenging than performing a physical records inventory, but the purposes are the same: to ferret out RM problems and to use the inventory as the basis for developing the retention schedule. Some of the RM problems that may be uncovered

> include inadequate documentation of official actions, improper applications of recordkeeping technology, deficient filing systems and maintenance practices, poor management of nonrecord materials, insufficient identification of vital records, and inadequate records security practices. When completed, the inventory should include all offices, all records, and all nonrecord materials. An inventory that is incomplete or haphazard can only result in an inadequate schedule and loss of control over records.[7]

The first step in gaining control over an organization's records and implementing IG measures to control and manage them is to complete an inventory of all groupings of business records, including electronic records,[8] *at the system or file series level.*

The focus of this book is on e-records, and when it comes to e-records, NARA has a specific recommendation: inventory *at the computer systems level.* This differs from advice given by experts in the past.

The records inventory is the basis for developing a **records retention schedule** that spells out how long different types of records are to be held and how they will be archived or disposed of at the end of their life cycle. But first you must determine where business records reside, how they are stored, how many exist, and how they are used in the normal course of business.

> NARA recommends that electronic records are inventoried by information system, not by record series.

There are a few things to keep in mind when approaching the e-records inventorying process:

- Those who create and work with the records themselves are the best source of information about how the records are used. They are your most critical resource in the inventorying process.
- RM is something that everyone wants done but no one wants to do (although everyone will have an opinion on how to do it).
- The people working in business units are touchy about their records. It will take some work to get them to trust a new RM approach.[9]

These knowledge workers are your best resource and can be your greatest allies or worst enemies when it comes to gathering accurate inventory data; developing a workable file plan; and keeping the records declaration, retention, and disposition process operating efficiently. A sound RM program will keep the records inventory accurate and up to date.

RM Intersection with Data Privacy Management

By Teresa Schoch

The tsunamic rise in electronic information and increasing complexity in information management[10] has resulted in newly created or redefined roles tasked with creating order out of chaos. Information professionals attempt to control their domains in roles described as content management, knowledge management, e-discovery, data management, data security, records management, privacy management, and IG. Due to an increased focus on privacy rights in the EU, as well as in individual US states, potential enforcement actions have had an alarming impact on all of these roles, often

causing intraorganizational conflict as enforced silos inhibit compliance with expanding and complex privacy laws.

Other than prohibitively expensive e-discovery disasters or unexpected regulatory audits leading to fines, in the past, there has been no real accountability in the United States as to how an organization maintains, organizes, creates, and/or disposes of its information. However, data breaches of private information, as well as the hacking of business and trade secrets, have become commonplace. [11] Corporations have scrambled to determine the what, where, when, and ownership of information that has been compromised, often racing against the clock to notify law enforcement and impacted individuals in time to avoid financial damages. C-level executives have lost their jobs over their company's mishandling of breaches.[12] A court allowed a class action by credit card holders against Neiman Marcus,[13] and the FTC acted against Wyndham Resorts for failure to protect the data of its customers.[14] The potential for fines imposed by the FTC or state attorney generals based on state data-breach laws, damages in private lawsuits, or the untimely loss of highly placed executives increases the potential costs of a future breach. In addition, as reflected in the Neiman Marcus case, damage to a company's reputation due to the glaring scrutiny of its inadequate IG program serves to motivate others to remedy inadequate IG frameworks.[15]

Meanwhile, the EU has the right to fine US organizations collecting data on EU residents up to 20 million euros, or 4% of global sales, for the mishandling of personal data pursuant to the GDPR, which became effective on May 25, 2018. Records management is at the core of information management, since it is the gatekeeper to all information of ongoing value to the organization (a record is defined as information that has business value or meets regulatory or litigation requirements); it is even more obvious in privacy management now that private data being maintained is required to be held for a specified time in a specified manner. Personal data must be capable of easy access and any data maintained on a protected individual must be accurate. In addition, when private data no longer has business value, the risk of maintaining it becomes prohibitive and it must be deleted in a manner that ensures continued privacy protection. As records managers assess their own domains, they realize that many of the obligations created by new privacy laws can only be met if they understand the new laws' effects on how they manage personal data pursuant to laws that impact their organization. Some erroneously assumed that privacy managers/attorneys/directors would expand their roles by learning the RIM world and addressing the changes required by privacy laws, but instead, they refer the details of maintaining privacy-related records to their records staff. While their willingness to delegate the legal duties of access, scheduling, deletion, and reliability of personal data is laudable, it creates new dynamics and an increased level of responsibility within the RIM framework that might not be thoroughly understood by the organization.

Since the GDPR has taken effect, many corporate attorneys have instructed the RIM staff to reassess records retention schedules based on the GDPR. Overworked professionals from all domains have developed plans to meet compliance requirements, often attempting to make the law fit into how they have always handled their duties in the past. As an example, the RIM "big bucket" approach, utilized to create records retention schedules for global records containing personal data of EU residents, could lead to fines in the millions. When an EU country requires employment records' retention of 30 years, while another country requires disposal of the same record types three months after termination of employment, a default to a 30-year schedule for all EU employment-related data is simply an unsound practice. Likewise, deletion of all EU employment data three months after employment termination would leave an organization open to an inability to meet the legal obligations of other jurisdictions, and to the inability to defend the organization in the event of litigation. In this instance, each country needs to be addressed individually. If there is a legitimate basis for maintaining personal data (e.g. potential litigation relating to employment), the data can be maintained under GDPR solely for that purpose, even if there is a privacy-related requirement of a shorter retention period for that specific data. In these "conflict of laws" situations, the data maintained for the interim retention period based on legitimate business interests requires heightened security as well as restricted access. *Retention schedules relating to records containing personal data have their own rules, often involving a conflict of laws, that require a new data-scheduling framework within the RIM environment.*

In the RIM domain, managing information that contains personal data is an example where less is more. Less information makes it easier and faster to retrieve relevant information (in this case, personal data), costs less to maintain, and limits liability to those whose information is deleted as soon as it no longer has business value. Until recently, the decision to "keep it all" was based on an assessment of return on investment that considered the risks worth taking compared to the cost of ensuring compliance through the creation of a long-term IG roadmap. The lack of calculated routine disposition was defended as a strategic decision to maintain data for marketing or business planning using increasingly sophisticated analytical software.

However, attempting to meet GDPR requirements while maintaining large data pools or warehouses of information that have not been identified, much less classified (the unknown unknown), creates an extremely difficult environment for compliance. For companies that do business with European residents, enforcing defensible disposition has become a critical mission. While scheduling records disposition has become more complex under GDPR, meeting a defensibility standard relating to disposition has become easier.

Generally Accepted Recordkeeping Principles®

It may be useful to use a model or framework to guide your records inventorying efforts. Such frameworks could be the DIRKS (Designing and Implementing Recordkeeping Systems) used in Australia or the Generally Accepted

Recordkeeping Principles® (or "the Principles") that originated in the United States at **ARMA International**. The Principles are a *"framework for managing records in a way that supports an organization's immediate and future regulatory, legal, risk mitigation, environmental, and operational requirements."*[16] More detail can be found in Chapter 3.

Special attention should be given to creating an accountable, open inventorying process that can demonstrate integrity. The result of the inventory should help the organization adhere to records retention, disposition, availability, protection, and compliance aspects of the Principles.

> The Principles are guidelines for information management and governance of record creation, organization, security, maintenance, and other activities used to effectively support record keeping of an organization.

The Generally Accepted Recordkeeping Principles were created with the assistance of ARMA International and legal and IT professionals who reviewed and distilled global best practice resources. These included the international records management standard ISO15489–1 from the American National Standards Institute and court case law. The principles were vetted through a public call-for-comment process involving the professional records information management ... community.[17]

E-Records Inventory Challenges

If your organization has received a legal summons for e-records, and you do not have an accurate inventory, the organization is already in a compromising position: You do not know where the requested records might be, how many copies there might be, or the process and cost of producing them. Inventorying must be done sooner rather than later and proactively rather than reactively.

E-records present challenges beyond those of paper or microfilmed records due to their (electronic) nature:

1. You cannot see or touch them without searching online, as opposed to simply thumbing through a filing cabinet or scrolling through a roll of microfilm.
2. They are not sitting in a central file room but rather may be scattered about on servers, shared network drives, or on storage attached to mainframe or minicomputers.
3. They have metadata attached to them that may distinguish very similar-looking records.
4. Additional "shadow" copies of the e-records may exist, and it is difficult to determine the true or original copy.[18]

Records Inventory Purposes

The completed records inventory contributes toward the pursuit of an organization's IG objectives in a number of ways: It supports the ownership, management, and control of records; helps to organize and prepare for the discovery process in litigation; reduces exposure to business risk; and provides the foundation for a disaster recovery/business continuity plan.

Completing the records inventory offers at least eight additional benefits:

1. It identifies records ownership and sharing relationships, both internal and external.
2. It determines which records are physical, electronic, or a combination of both.
3. It provides the basis for retention and disposition schedule development.
4. It improves compliance capabilities.
5. It supports training objectives for those handling records.
6. It identifies vital and sensitive records needing added security and backup measures.
7. It assesses the state of records storage, its quality and appropriateness.
8. It supports the release of information for Freedom of Information Act (FOIA), Data Protection Act, and other mandated information release requirements for governmental agencies.[19]

With respect to e-records, the purpose of the records inventory should include the following objectives:

- Provide a survey of the existing electronic records situation.
- Locate and describe the organization's electronic record holdings.
- Identify obsolete electronic records.
- Determine storage needs for active and inactive electronic records.

- Identify vital and archival electronic records, indicating need for their ongoing care.
- Raise awareness within the organization of the importance of electronic records management.
- Lead to electronic recordkeeping improvements that increase efficiency.
- Lead to the development of a needs assessment for future actions.
- Provide the foundation of a written records management plan with a determination of priorities and stages of actions, assuring the continuing improvement of records management practices.[20]

The completed records inventory contributes toward the pursuit of an organization's IG objectives in a number of ways.

Records Inventorying Steps

NARA's guidance on how to approach a records inventory applies to both physical and e-records.

The steps in the records inventory process are:

1. *Define the inventory's goals*. While the main goal is gathering information for scheduling purposes, other goals may include preparing for conversion to other media, or identifying particular records management problems.
2. *Define the scope of the inventory;* it should include all records and other materials.
3. *Obtain top management's support*, preferably in the form of a directive, and keep management and staff informed at every stage of the inventory.
4. *Decide on the information to be collected* (the elements of the inventory). Materials should be located, described, and evaluated in terms of use.
5. *Prepare an inventory form*, or use an existing one.
6. *Decide who will conduct the inventory*, and train them properly.
7. *Learn where the agency's* [or business'] *files are located*, both physically and organizationally.
8. *Conduct the inventory*.
9. *Verify and analyze the results*.[21]

Goals of the Inventory Project

The goals of the inventorying project must be set and conveyed to all stakeholders. At a basic level, the primary goal can be simply to generate a complete inventory for compliance and reporting purposes. It may focus on a certain business area or functional group or on the enterprise as a whole. An enterprise approach requires segmenting the effort into smaller, logically sequenced work efforts, such as by business unit. *Perhaps the organization has a handle on its paper and microfilmed records but e-records have been growing exponentially and spiraling out of control, without good policy guidelines or IG controls*. So a complete inventory of records and e-records by system is needed, which may include e-records generated by application systems, residing in e-mail, created in office documents and spreadsheets, or other potential business records. This is a tactical approach that is limited in scope.

The goal of the inventorying process may be more ambitious: to lay the groundwork for the acquisition and implementation of an ERM system that will manage the retention, disposition, search, and retrieval of records. It requires more business process analysis and redesign, some rethinking of business classification schemes or file plans, and development of an enterprise-wide taxonomy. This redesign will allow for more sharing of information and records; faster, easier, and more complete retrievals; and a common language and approach for knowledge professionals across the enterprise to declare, capture, and retrieve business records.

The plan may be still much greater in scope and involve more challenging goals: That is, the inventorying of records may be the first step in the process of implementing an organization-wide IG program to manage and control information by rolling out ERM and IG systems and new processes; to improve litigation readiness and stand ready for e-discovery requests; and to demonstrate compliance adherence with business agility and confidence. Doing this involves an entire cultural shift in the organization and a long-term approach.

Whatever the business goals for the inventorying effort, they must be conveyed to all stakeholders, and that message must be reinforced periodically and consistently, and through multiple means. It must be clearly spelled out in communications and presented in meetings as the overarching goal that will help the organization meet its business objectives. The scope of the inventory must be appropriate for the business goals and objectives it targets.

Whatever the business goals for the inventorying effort are, they must be conveyed to all stakeholders, and that message must be reinforced periodically and consistently, and through multiple means.

Scoping the Inventory

"With senior-level support, the records manager must decide on the scope of the records inventory. A single inventory could not describe every electronic record in an organization; *an appropriate scope might enumerate the records of a single program or division, several functional series across divisions, or records that fall within a certain time frame*" [emphasis added].[22] Most organizations have not deployed an enterprise-wide records management system, which makes the e-records inventorying process arduous and time-consuming. It is not easy to find where all the electronic records reside—they are scattered all over the place and on different media. But impending (and inevitable) litigation and compliance demands require that it be done. And, again, sooner has been proven to be better than later. Since courts have ruled that if lawsuits have been filed against your competitors over a certain (industry-specific) issue, your organization should anticipate and prepare for litigation—which means conducting records inventories and placing a litigation hold on documents that might be relevant. Simply doing nothing and waiting on a subpoena is an avoidable business risk.

An appropriate scope might enumerate the records of a single program or division, several functional series across divisions, or records that fall within a certain time frame.

A methodical, step-by-step approach must be taken—it is the only way to accomplish the task. A plan that divides up the inventorying tasks into smaller, accomplishable pieces is the only one that will work. It has been said, "How do you eat an elephant?" And the answer is "One bite at a time." The inventorying process can be divided into segments, such as a business unit, division, or information system/application.

Management Support: Executive Sponsor

It is crucial to have management support to drive the inventory process to completion. There is no substitute for an executive sponsor. Asking employees to take time out for yet another survey or administrative task without having an executive sponsor will likely not work. Employees are more time-pressed than ever, and they will need a clear directive from above, along with an understanding of what role the inventorying process plays in achieving a business goal for the enterprise, if they are to take the time to properly participate and contribute meaningfully to the effort.

Information/Elements for Collection

During the inventory you should collect the following information at a minimum:

- What kind of record it is—*contracts, financial reports,* memoranda, and so on
- What department owns it
- What departments access it
- What application created the record (e-mail, MS Word, Acrobat PDF)
- Where it is stored, both physically (tape, server) and logically (network share, folder)
- Date created
- Date last changed
- Whether it is a vital record (mission-critical to the organization)
- Whether there are other forms of the record (for example, a document stored as a Word document, a PDF, and a paper copy) and which of them is considered the official record

Removable media should have a unique identifier and the inventory should include a list of records on the particular volume as well as the characteristics of the volume, for example, the brand, the recording format, the capacity and volume used, and the date of manufacture and date of last update.

Additional information not included in inventories of physical records must be collected in any inventory of e-records.

IT Network Diagram

Laying out the overall topology of the IT infrastructure in the form of a network diagram is an exercise that is helpful in understanding where to target efforts and to map information flows. Data mapping is a crucial early step in compliance efforts, such as GDPR. Creating this map of the IT infrastructure is a crucial step in inventorying e-records. It graphically depicts how and where computers are connected to each other and the software operating environments of various applications that are in use. This high-level diagram does not need to include every device; rather, it should indicate each *type* of device and how it is used.

The IT staff usually has a network diagram that can be used as a reference; perhaps after some simplification it can be put into use as the underpinning for inventorying e-records. It does not need great detail, such as where network bridges and routers are located, but it should show which applications are utilizing the cloud or hosted applications to store and/or process documents and records.

In diagramming the IT infrastructure for purposes of the inventory, it is easiest to start in the central computer room where any mainframe or other centralized servers are located and then follow the connections out into the departments and business unit areas, where there may be multiple shared servers and drives supported a network of desktop personal computers or workstations.

SharePoint is a prevalent document and RM portal platform, and many organizations have SharePoint servers to house and process e-documents and records. Some utilities and tools may be available to assist in the inventorying process on SharePoint systems. This process has been made easier with the introduction of cloud-based SharePoint services.

Mobile devices (e.g. tablets, smartphones, and other portable devices) that are processing documents and records should also be represented. And any e-records residing in cloud storage should also be included.

Creating a Records Inventory Survey Form

The record inventory survey form must suit its purpose. Do not collect data that is irrelevant, but, in conducting the survey, be sure to collect all the needed data elements. You can use a standard form, but some customization is recommended. The sample records survey form in Figure 9.1 is wide-ranging yet succinct and has been used successfully in practice.

Department Information

1. What is the reporting structure of the department?
2. Who is the department liaison for the records inventory?
3. Who is the IT or business analyst liaison?

Record Requirements

4. Are there any external agencies that impose guidelines, standards, or other requirements?
5. Are there specific legislative requirements for creating or maintaining records? Please provide a copy.
6. Is there a departmental records retention schedule?
7. What are the business considerations that drive recordkeeping? Regulatory requirements? Legal requirements?
8. Does the department have an existing records management policy? Guidelines? Procedures? Please provide a copy.
9. Does the department provide guidance to employees on what records are to be created?
10. How are policies, procedures, and guidance disseminated to the employees?
11. What is the current level of employees' awareness of their responsibilities for records management?
12. How are nonrecords managed?
13. What is the process for ensuring compliance with policies, procedures, and guidelines?
    When an employee changes jobs/roles or is terminated?
14. Does the department have a classification or file plans?
15. Are any records in the department confidential or sensitive?
16. What information security controls does the department have for confidential or sensitive records?
17. Does the department have records in sizes other than letter (8½ × 11)?

18. What is the cutoff date for the records?

    □ Fiscal Year □ Calendar Year □ Other

19. Have department vital records been identified?
20. Is there an existing business or disaster recovery policy?
21. Is the department subject to audits? Internal? External? Who conducts the audits?
22. Where and how are records stored? (Circle all that apply)

    Online? Near Line? Offline? On-site? Off-site? One location? Multiple locations?

23. How does the department ensure that records will remain accessible, readable, and useable throughout their scheduled retention period?

Technology and Tools

24. Are any tools used to track active records? Spreadsheets, word documents, databases, and so forth?
25. Are any tools used to track inactive records? Spreadsheets, word documents, databases, and so forth?
26. Does the department use imaging, document management, and so forth?

Disposition

27. Are there guidelines for destroying obsolete records?
28. What disposition methods are authorized or required?
29. How does disposition occur? Paper? Electronic? Other?
30. What extent does the department rely on each individual to destroy records? Paper? Electronic? Other?

Records Holds

31. What principles govern decisions for determining the scope of records that must be held or frozen for an audit or investigations?
32. How is the hold or freeze communicated to employees?
33. How are records placed on hold protected?

Figure 9.1 Records Inventory Survey Form

Source: Charmaine Brooks, IMERGE Consulting, e-mail to author, March 20, 2012.

If conducting the e-records portion of the inventory, the sample form may be somewhat modified, as shown in Figure 9.2.

Identifying Information

1. Name of system.
2. Program or legal authority for system.
3. System identification or control number.
4. Person responsible for administering the system. Include e-mail, office address, and phone contact info.
5. Date system put in service.
6. Business unit or agency supported by system.
7. Description of system (what does the application software do?).
8. Purpose of system.

System Inputs/Outputs

9. Primary sources of data inputs.
10. Major outputs of system (e.g., specific reports).

11. Informational content (all applicable): Description of data; applicability of data (people, places, things); geographic information; time span; update cycle; applications the system supports; how data are manipulated; key unit analysis for each file; public use or not?
12. Hardware configuration.
13. Software environment, including revision levels, operating system, database, and so forth.
14. Indices or any classification scheme/fi le plan that is in place?
15. Duplicate records? Location and volume of any other records containing the same information.

Record Requirements

16. Are there any external agencies that impose guidelines, standards, or other requirements?
17. Are there specific legislative requirements for creating or maintaining records? Please provide a copy.
18. Is there a departmental records retention schedule?
19. What are the business considerations that drive recordkeeping? Regulatory requirements? Legal requirements?
20. Does the department have an existing records management policy? Guidelines? Procedures? If so, please provide a copy.
21. How are nonrecords managed?
22. Are any records in the department confidential or sensitive? How are they indicated or set apart?
23. What information security controls does the department have for confidential or sensitive records?
24. What is the cutoff date for the records?
    □ Fiscal Year □ Calendar □ Year Other _____
25. Have department vital records been identified?
26. Is there an existing business or disaster recovery policy?
27. Is the department subject to audits? Internal? External? Who conducts the audits?
28. Where and how are records stored?
    Online? Near line? Offline? On-site? Off-site? One location? Multiple locations?
29. How does the department ensure that records will remain accessible, readable, and useable throughout their scheduled retention period?

Disposition

30. Are there guidelines for destroying obsolete records?
31. What disposition methods are authorized or required?
32. How does disposition occur? Are electronic deletions verified?
33. What extent does the department rely on each individual to destroy e-records?

Records Holds

34. What principles govern decisions for determining the scope of records that must be held or frozen for an audit or investigations?
35. How is the hold or freeze communicated to employees?
36. How are records placed on hold protected?

Figure 9.2 Electronic Records Inventory Survey Form

Source: Adapted from www.archives.gov/records-mgmt/faqs/inventories.html and Charmaine Brooks, IMERGE Consulting.

Who Should Conduct the Inventory?

Typically, a RM project team is formed to conduct the survey, often assisted by resources outside of the business units. These may be RM and IT staff members, business analysts, members of the legal staff, outside specialized consultants, or a combination of these groups. The greater the cross-section from the organization, the better, and the more expertise brought to bear on the project, the more likely it will be completed thoroughly and on time.

Critical to the effort is that those conducting the inventory are trained in the survey methods and analysis, so that when challenging issues arise, they will have the resources and know-how to continue the effort and get the job done.

Determine Where Records Are Located

The inventory process is, in fact, a surveying process, and it involves going physically out into the units where the records are created, used, and stored. Mapping out where the records are *geographically* is a basic necessity. Which buildings are they located in? Which office locations? Computer rooms?

Also, the inventory team must look *organizationally* at where the records reside (i.e., determine which departments and business units to target and prioritize in the survey process).

Conduct the Inventory

Several approaches can be taken to conduct the inventory, including four basic methods:

1. Distributing and collecting surveys
2. Conducting in-person interviews
3. Direct observation
4. Software tools

Creating and distributing a survey form is traditional and proven way to collect e-records inventory data. This is a relatively fast and inexpensive way to gather the inventory data. The challenge is getting the surveys completed and completed in a consistent fashion. This is where a strong executive sponsor can assist. The sponsor can make the survey a priority and tie it to business objectives, making the survey completion compulsory. The survey is a good tool, and it can be used to cover more ground in the data collection process. If following up with interviews, the survey form is a good starting point; responses can be verified and clarified, and more detail can be gathered.

> There are four ways to conduct the inventory: surveys, interviews, observation, and software tools. Combining these methods yields the best results.

Some issues may not be entirely clear initially, so following up with scheduled in-person interviews can dig deeper into the business processes where formal records are create and used. A good approach is to have users walk you through their typical day and how they access, use, and create records—but be sure to interview managers too, as managers and users have differing needs and uses for records.

You will need some direction to conduct formal observation, likely from IT staff or business analysts familiar with the record-keeping systems and associated business processes. They will need to show you where business documents and records are created and stored. If there is an existing ERM system or other automated search and retrieval tools available, you may use them to speed the inventorying process.

When observing and inventorying e-records, starting in the server room and working outward toward the end user is a logical approach. Begin by enumerating the e-records created by enterprise software applications (such as accounting, enterprise resource planning, or customer relationship management systems), and work your way to the departmental or business unit applications, on to shared network servers, then finally out to individual desktop and laptop PCs and other mobile devices. With today's smartphones, this can be a tricky area, due to the variety of platforms, operating systems, and capabilities. In a bring-your-own-device environment, records should not be stored on personal devices, but if they must be, they should be protected with technologies like encryption or information rights management.

Explore any potential software tools that may be available for facilitating the inventorying of e-records. Some software vendors that provide document management or e-record management solutions offer inventorying tools.

There are always going to be thorny areas when attempting to inventory e-records to determine what files series exist in the organization. Mobile devices and removable media may contain business records. These must be identified and isolated, and any records on these media must be recorded for the inventory. Particularly troublesome are thumb or flash drives, which are compact yet can store 20 gigabytes of data or more. If your IG measures call for excluding these types of media, the ports they use can be blocked on PCs, tablets, smartphones, and other mobile computing devices using **data loss prevention** (DLP) technology. A sound IG program will consider the proper use of removable media and the potential impact on your RM program.[23]

The best approach for conducting the inventory is to combine the available inventorying methods, where possible. Begin by observing, distribute surveys, collect and analyze them, and then target key personnel for follow-up interviews

and walk-throughs. Utilize whatever automated tools that are available along the way. This approach is the most complete. *Bear in mind that the focus is not on individual electronic files but rather, the file series level for physical records and the file series or system level for e-records (preferably the latter).*

Interviewing Programs/Service Staff

Interviews are a very good source of records inventory information. Talking with actual users will help the records lead or inventory team to better understand how documents and records are created and used in everyday operations. Users can also report *why* they are needed—an exercise that can uncover some obsolete or unnecessary processes and practices. This is helpful in determining where e-records reside and how they are grouped in records series or by system and ultimately, the proper length of their retention period and whether they should be archived or destroyed at the end of their useful life.

Since interviewing is a time-intensive task, it is crucial that some time is spent in determining the key people to interview: Interviews not only take your time but others' as well, and the surest way to lose momentum on an inventorying project is to have stakeholders believe you are wasting their time.

You need to interview representatives from all functional areas and levels of the program or service, including:

- Managers
- Supervisors
- Professional/technical staff
- Clerical/support staff

The people who work with the records can best describe to you their use. They will likely know where the records came from, if copies exist, who needs the records, any computer systems that are used, how long the records are needed, and other important information that you need to know to schedule the records.

Selecting Interviewees

As stated earlier, it is wise to include a cross-section of staff, managers, and frontline employees to get a rounded view of how records are created and used. Managers have a different perspective and may not know how workers utilize electronic records in their everyday operations.

A good lens to use is to focus on those who make decisions based on information contained in the electronic records and to follow those decision-based processes through to completion, observing and interviewing at each level.

For example, an application is received (mail room logs date and time), checked (clerk checks the application for completeness and enters into a computer system), verified (clerk verifies that the information on the application is correct), and approved (supervisor makes the decision to accept the application). These staff members may only be looking at specific pieces of the record and making decisions on those pieces.

Interview Scheduling and Tips

One rule to consider is this: Be considerate of other people's work time. Since they are probably not getting compensated for participating in the records inventory, the time you take to interview them is time taken away from compensated tasks they are evaluated on. So, once the interviewees are identified, provide as much advance notice as possible, follow up to confirm appointments, and stay within the scheduled time. Interviews should be kept to 20 to 60 minutes. Most of all—*never be late!*

Before starting any interviews, be sure to restate the goals and objectives of the inventorying process and how the resulting output will benefit people in their jobs.

In some cases, it may be advisable to conduct interviews in small groups, not only to save time but also to generate a discussion of how records are created, used, and stored. Some new insights may be gained.

Try to schedule interviews that are as convenient as possible for participants. That means providing participants with questions in advance and holding the interviews as close to their work area as possible. Do not schedule interviews back to back with no time for a break between. You will need time to consolidate your thoughts and notes, and, at times, interviews may exceed their planned time if a particularly enlightening line of questioning takes place.

If you have some analysis from the initial collection of surveys, share that with the interviewees so they can validate or help clarify the preliminary results. Provide it in advance, so they have some time to think about it and discuss it with their peers.

Sample Interview Questionnaire

You'll need a guide to structure the interview process. A good starting point is the sample questions presented in the questionnaire shown in Figure 9.3. It is a useful tool that has been used successfully in actual records inventory projects.

What is the mandate of the office?

What is the reporting structure of the department?

Who is the department liaison for the records inventory?

Are there any external agencies that impose guidelines, standards, or other requirements?

Is there a departmental records retention schedule?

Are there specific legislative requirements for creating or maintaining records? Please provide a copy.

What are the business considerations that drive recordkeeping? Regulatory requirements? Legal requirements?

Does the department have an existing records management policy? Guidelines? Procedures?

Please provide a copy.

Does the department provide guidance to employees on what records are to be created?

What is the current level of awareness of employees their responsibilities for records management?

How are nonrecords managed?

Does the department have a classification or file plans?

What are the business drivers for creating and maintaining records?

Where are records stored? On-site? Off-site? One location? Multiple locations?

Does the department have records in sizes other than letter (8½ × 11)?

What is the cutoff date for the records?

□ Fiscal Year □ Calendar □ Year Other

Are any tools used to track active records? Excel, Access, and so forth?

Does the department use imaging, document management, and so forth?

Is the department subject to audits? Internal? External? Who conducts the audits?

Are any records in the department confidential or sensitive?

Are there guidelines for destroying obsolete records?

What disposition methods are authorized or required?

How does disposition occur? Paper? Electronic? Other?

What extent does the department rely on each individual to destroy records?

□ Paper □ Electronic □ Other _____

What principles govern decisions for determining the scope of records that must be held or frozen for an audit or investigations?

How is the hold or freeze communicated to employees?

Figure 9.3 Sample Interview Questionnaire

Source: Charmaine Brooks, IMERGE Consulting, e-mail to author, March 20, 2012.

Analyze and Verify the Results

Once collected, some follow-up will be required to verify and clarify responses. Often this can be done over the telephone. For particularly complex and important areas, a follow-up in person visit can clarify the responses and gather insights.

Once the inventory draft is completed, a good practice is to go out into the business units and/or system areas and verify what the findings of the survey are. Once presented with findings in black and white, key stakeholders may have additional insights that are relevant to consider before finalizing the report. Do not miss out on the opportunity to allow power users and other key parties to provide valuable input.

Be sure to tie the findings in the final report of the records inventory to the business goals that launched the effort. This helps to underscore the purpose and importance of the effort, and will help in getting that final signoff from the executive sponsor that states the project is complete and there is no more work to do.

Depending on the magnitude of the project, it may (and *should*) turn into a formal IG program that methodically manages records in a consistent fashion in accordance with internal governance guidelines and external compliance and legal demands.

Be sure to tie the findings in the final report of the records inventory to the business goals that launched the effort.

Appraising the Value of Records

Part of the process of determining the retention and disposition schedule of records is to appraise their value. Records can have value in different ways, which affects retention decisions.

> **Records appraisal** is an analysis of all records within an agency [or business] to determine their administrative, fiscal, historical, legal, or other archival value. The purpose of this process is to determine for how long, in what format, and under what conditions a record series ought to be preserved. *Records appraisal is based upon the information contained in the records inventory*. Records series shall be either preserved permanently or disposed of when no longer required for the current operations of an agency or department, depending upon:
>
> - *Historical value* or the usefulness of the records for historical research, including records that show an agency [or business] origin, administrative development, and present organizational structure.
> - *Administrative value* or the usefulness of the records for carrying on [a business or] an agency's current and future work, and to document the development and operation of that agency over time.
> - *Regulatory and statutory* [value to meet] requirements.
> - *Legal value* or the usefulness of the records to document and define legally enforceable rights or obligations of [business owners, shareholders, or a] government and/or citizens.
> - *Fiscal value* or the usefulness of the records to the administration of [a business or] an agency's current financial obligations, and to document the development and operation of that agency over time.
> - Other archival value as determined by the State [or corporate] Archivist.[24] (Emphasis added.)

Records appraisal is based on the information contained in the records inventory.

Ensuring Adoption and Compliance of RM Policy

The inventorying process in not a one-shot deal: It is useful only if the records inventory is kept up to date, so it should be reviewed, at least annually. A process should be put in place so that business unit or agency heads notify the RM head/lead if a new file series or system has been put in place and new records collections are created. There are emerging approaches that utilize file/content analytics tools to automate this process.

Following are some tips to help ensure that a records management program achieves its goals:

1. *Records management is everyone's role*. The volume and diversity of business records, from e-mails to reports to tweets, means that the person who creates or receives a record is in the best [position] to classify it. Everyone in the organization needs to adopt the records management program.
2. *Don't micro-classify*. Having hundreds, or possibly thousands, of records classification categories may seem like a logical way to organize the multitude of different records in a company. However, the average information worker, whose available resources are already under pressure, does not want to spend any more time than necessary classifying records. Having a few broad classifications makes the decision process simpler and faster.
3. *Talk the talk from the top on down*. A culture of compliance starts at the top. Businesses should establish a senior-level steering committee comprised of executives from legal, compliance, and information technology (IT). A committee like this signals the company's commitment to compliant records management and ensures enterprise adoption.
4. *Walk the walk, consistently*. For compliance to become second nature, it needs to be clearly communicated to everyone in the organization, and policies and procedures must be accessible. Training should be rigorous and easily available, and organizations may consider rewarding compliance through financial incentives, promotions and corporate-wide recognition.
5. *Measure the measurable*. The ability to measure adherence to policy and adoption of procedures should be included in core business operations and audits. Conduct a compliance assessment, including a gap analysis, at least once a year, and prepare an action plan to close any identified holes.

The growth of information challenges a company's ability to use and store its records in a compliant and cost-effective manner. Contrary to current practices, the solution is not to hire more vendors or to adopt multiple technologies. The key to compliance is consistency, with a unified enterprise-wide approach for managing all records, regardless of their format or location.

Therefore a steady and consistent IG approach that includes controls, audits, and clear communication is key to maintaining an accurate and current records inventory.

Tracking Information Assets: Moving Toward an Information Asset Register

A relatively new concept in IG is the development of an **Information Asset Register** (IAR). An IAR is a sort of "general ledger of information assets" that lists all information assets, structured and unstructured, their lifecycle retention, privacy and security requirements, where they are housed, and even what hardware is used. Dennis Kessler, Data Governance Lead at the European Investment Bank states, "This asset-based approach to managing information helps to reveal:

- Which people or teams are responsible and accountable for maintaining the confidentiality, integrity, and availability of information
- Which people/teams and systems can access information, and for what purposes—whether to create, update, or consume information
- How information flows and is used throughout the organization—which business processes and which decision-making points depend on which information
- Regulatory, compliance, and other obligations"[25]

Briton Reynold Leming has been developing IAR software for several years. He has created a comprehensive list of benefits that an IAR provides:[26]

1. *Understanding Relationships:* A related series of records sharing the same purpose (an "asset collection," if you will) might have a variety of constituent entities ("assets") in different formats—for example, physical records, digital content, system data. Identifying these within an IAR, with a suitable narrative recorded, will enable an understanding of their relationships and purpose over time. This could include, for example, the "story" of document handling paper originals and resulting images within a document scanning process or the retirement and introduction of systems.[27]

   Allied to this is tagging assets to a business classification scheme of the functions and activities of your organization. This allows the assets to be categorized to a vocabulary of business activity that is neutral to and more stable than organizational structures (which can change more often than what an organization actually does), provides a collated corporate view of assets maintained based upon their purpose (e.g. many departments will hold invoice, staff, policy, and contract records), and supports cross-cutting processes involving different teams. It also allows the consistent inheritance and application of business rules, such as retention policies.

2. *Security Classification:* Assets can be classified within the IAR to an approved security classification/protective marking scheme, with current protective measures recorded, in order to identify if there are any risks relating to the handling of confidential personal or commercially sensitive information. You can assess that assets are handled, stored, transferred, and disposed of in an appropriate manner.

3. *Personal Data:* Specifically, you can identify confidential personal information to ensure that data protection and privacy obligations are met.

   The GDPR contains many obligations that require a thorough understanding of what personal data you process and how and why you do so. Many requirements for keeping records as a data controller for GDPR Article 30 can be supported by the information asset inventory. For example, the asset attributes can describe the purposes of the processing, the categories of data subjects and personal data, categories of recipients, envisaged time limits for erasure of the different categories of data, and a general description of the technical and organizational security measures.

   It will also help data processors keep a record of the categories of processing, transfers of personal data to a third country or an international organization, and a general description of the technical and organizational security measures.

   Much of the information about personal data required for Article 30 compliance is also useful to meet obligations under Article 13 and Article 14 on information to be provided, for example, via privacy notices or consent forms.

   Under Chapter 3 of the GDPR, data subjects have a number of rights. Understanding things such as the location, format, use of, and lawful basis of processing for different categories of personal data will enable will support responses to rights and requests.

   Under Article 25 of the GDPR there are requirements for Data Protection by design and by default. Additionally, under Article 35 there are requirements relating to Data Protection impact assessments. The inventory can provide insight into which processes and systems need to be assessed based upon, for example, the nature, scope, context, and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing.

   As aforementioned, it is important to identify who the personal data is shared with. The inventory can support this

as well as specifically enable monitoring of the existence or status or suitable agreements. For example, under Article 28 of the GDPR, processing by a processor shall be governed by a contract or other legal act under Union or Member State law.

Article 32 of the GDPR covers security of processing, with requirements to implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk. Then, using the inventory you can assess the security measures in place for assets against their level of confidentiality. It also can help with identifying the data sets where, if anything unfortunate were to happen, there are considerations regarding Article 33 Notification of a personal data breach to the supervisory authority and Article 34 Communication of a personal data breach to the data subject.

4. *Ownership:* The ability to know: Who owns what? This includes understanding ownership both in terms of corporate accountability and ownership of the actual information itself. You could also record who administers an asset on a day-to-day basis if this is different.
5. *Business Continuity:* An organization will have vital/business critical records that are necessary for it to continue to operate in the event of a disaster. They include those records that are required to recreate the organization's legal and financial status, to preserve its rights, and to ensure that it can continue to fulfill its obligations to its stakeholders. Assets can be classified within the IAR to an approved criticality classification scheme, with current protective measures recorded, in order to assess whether they are stored and protected in a suitable manner and identify if there are any risks relating to business critical ("vital record") information. You can also identify the Recovery Point Objective (RPO) and Recovery Time Objective (RTO) for assets to support a disaster recovery or data protection plan.
6. *Originality:* You can identify whether an asset is original or a copy, ascertaining its relative importance and supporting decisions on removing duplication and the optimization of business processes.
7. *Heritage:* You can identify records of historical importance that can be transferred at some stage to the custody of a corporate or third-party archive.
8. *Formats:* The ability to understand the formats used for information, supporting decisions on digital preservation or migration.
9. *Space Planning:* In order to support office moves and changes, data can be gathered for physical assets relating to their volume, footprint, rate of accumulation, use, filing methods, and so on.
10. *Subject Matter:* If assets are tagged to a business classification scheme of functions and activities, as well as potentially to a keyword list, the organization can understand the "spread" of record types (e.g. who holds personnel, financial, contractual records) and/or "discover" resources for knowledge management or e-discovery purposes.
11. *Archive Management:* The ability to understand what physical records (paper, backup tapes, etc.) are archived, where and when; this might, for example, identify risks in specific locations or issues with the regularity of archiving processes. The organization can also understand its utilization of third-party archive storage vendors—potentially supporting decisions on contract management/consolidation—and maintain their own future-proof inventory of archive holdings. Archive transactions can be recorded if there is no system to otherwise do so.
12. *Location:* The "location" of an asset can of course be virtual or physical. This (together with other questions relating to, for example, security measures) is important to ensure that information assets are suitably protected. It also helps in the planning of IT systems and physical filing/archiving services. The benefits for archive management are explored above and for maintaining a system catalogue below. Other examples might be to identify records to gather when doing an office sweep following vacation of a floor or building, or what assets are held in the cloud, or asset types within a given jurisdiction. It would also support the "discovery" of resources for knowledge management or e-discovery purposes.
13. *Retention:* An IAR can be used both to link assets with approved records retention policies and understand the policies and methods currently applied within the organization, therefore identifying queries, risks, and issues. The IAR can also be used to maintain the actual policies (across jurisdictions if applicable) and their citations; if a law changes or is enacted, relevant assets can be identified for any process changes to be made.
14. *Disposal:* An IAR can be used both to link assets with approved destruction or transfer policies and understand the processes and methods currently applied within the organization, therefore identifying queries, risks, and issues, particularly for confidential information. Disposal transactions can be recorded if there is no system to otherwise do so.
15. *Source:* The source of assets can be identified to understand where information is derived from and better manage the information supply chain. Under article 14 of the GDPR, part of the information the controller shall provide to the data subject to ensure fair and transparent processing includes from which source the personal data originate, and if applicable, whether it came from publicly accessible sources.
16. *Rights:* The rights held in and over assets can be identified, such as copyright and intellectual property, in order to

protect IPR and to avoid infringement of the rights of others.

17. *Applications Catalogue:* The application systems in use (e.g. content management, front and back office) can be identified and be linked in locations, people, activities, and, of course, assets. Licensing and upgrade criteria could also be managed. It would also be possible, for example, to identify system duplication or the use of homegrown databases.
18. *Condition:* Both physical and digital assets can degrade: this can be identified for assets with conservation/preservation actions taken accordingly.
19. *Age:* The age of assets can be established, with decisions made on their further retention/disposal, the need for archiving (historic or business), and potentially whether they need to be superseded with newer resources.
20. *Organization and Referencing:* An understanding can be gained of whether structured systems and approaches are in place to describe, reference, and organize physical and digital assets, identifying if there are likely to be any issues with the finding information.
21. *Utilization:* An understanding can be gained of whether assets are proposed, active, inactive, or discontinued/superseded, therefore enabling decisions on their format, storage, disposal, and so on.
22. *Sharing:* An IAR can be used to identify how information is shared within and without the organization, helping to ensure that it is available as required, and that suitable security measures and, where applicable, information sharing agreements are in place. This supports compliance with Article 30 of the GDPR as part of the records of processing activities.
23. *Provenance:* Fundamentally an IAR can provide an accountable audit trail of asset existence and activity, including any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.
24. *Publications:* Information produced for wider publication to an internal or external resource can be identified, including, for example, the audience for whom the resource is intended or useful, the channels used for distribution and the language(s) of the content, thus facilitating editorial, production, and dissemination planning and management.
25. *Quality:* Observations can be recorded on the quality of assets (e.g. accuracy, completeness, reliability, relevance, consistency across data sources, accessibility), with risks and issues identified and managed.

To help carry out Mr. Leming's approach to implementing and IAR, Mr. Kessler provided this succinct sample IAR survey, as shown in the following section.

> An IAR is a sort of general ledger of information assets that lists all information assets, structured and unstructured, their lifecycle retention, privacy and security requirements, where they are housed, and even what hardware is used.

Sample Information Asset Survey Questions[28]

1. **Information Asset Description**

   1.1 **Name**: A descriptive and meaningful label for the asset

   1.2 **Description**: What is the information asset and what it is used for

2. **Ownership**: This section covers the “stewardship” of the information asset, including the key roles and responsibilities of the owner and manager, together with any other stakeholders likely to be affected by the quality and availability of the asset.

   2.1 Owner has overall accountability for access, use, and management of the asset

   2.2 **Manager**: Hands-on manager responsible for day-to-day operations and administration. Expected to be familiar with the details of the asset content, structure and usage, and so likely to be quick to detect evidence of breach, tampering/corruption, and so on

   2.3 **Creator**: Source of the information; a person, application system, or external source

   2.4 **Stakeholders/customers**: Other stakeholders affected by the use, management, integrity, or availability of the asset

3. **Dates**

   3.1 **Creation date**: Date on which the asset was created (if involving a fixed lifecycle)

   3.2 **Last review date**: Date on which the asset was last reviewed for completeness and accuracy

3.3 **Date closed**: Date on which the asset was closed/completed/removed from production use

4. **Confidentiality**

   4.1 **Confidentiality**: Indicates the confidentiality classification of the information based on the Confidentiality of Information policy, which is distinct from the Risk/Impact Criticality rating in section 7 below.)

   4.2 **Data Privacy and Protection**: Indicates whether the asset contains or relates to Personally Identifiable Information (PII) and potential relevance to Data Protection or Data Privacy regulations and legal risk—especially the EU General Data Protection Regulation (GDPR).

5. **Retention**

   5.1 **Retention period**: Retention category (if useful—but avoid duplication and inconsistency)

   5.2 System of Record: Name of the record system used to store the asset (or a subset of related business records)

6. **Access and Use**

   6.1 **Applications and Interfaces**: List of applications and interfaces authorized to access the information asset, together with the corresponding access rights

   6.2 **User groups**: List of user groups authorized to access the asset, together with the corresponding access rights

   6.3 **Metadata**: List of any metadata needed to access or describe the context of the asset

7. **Risk/Impact of problems/issues**

   7.1 **Confidentiality**: Impact if the asset is accessed or disclosed without authorization

   7.2 **Integrity**: Impact if the information is corrupted, tampered with, or otherwise suffers a loss of integrity

   7.3 **Availability**: Impact if the information is lost or unavailable?

   7.4 **Criticality**: The overall criticality rating is the product of the combined scores of the above confidentiality, integrity, and availability ratings.

Of note: To truly have an IAR that stays up-to-date on a daily basis, file analysis software must be implemented on the back end to monitor and update the status of information assets.

General Principles of a Retention Scheduling

A series of basic principles common to all retention schedules, include:[29]

- The retention schedule must include all records.
- Records scheduling includes all records, regardless of media or location.[30]
- All legal and regulatory requirements for records must be reflected in the records scheduling process. For public entities, retention scheduling fosters and enables the agency to comply with information requests (e.g. FOIA in the United States, Freedom of Information Act 2000 in the United Kingdom, Freedom of Information and Protection of Privacy Act and the Health Information Act in Canada, and Freedom of Information Amendment [Reform] Act 2010 in Australia).
- Records scheduling is a “proactive” planning process, where schedules are set in place and standardized in advance.
- Periodic review of the retention schedule must take place when significant legislation, technology acquisitions, or other changes are being considered; but in any case, this should be at least annually or biannually.
- Records scheduling is a continuous process that needs updating and amending, based on legal, technology, or business changes over time.
- Classification and records scheduling are inextricably linked.
- File series with similar characteristics or value should be assigned consistent and appropriate retention periods.
- Records of historical value must be preserved.
- Records retention periods should reflect the business needs of users, the value of the records, and any legal or compliance requirements. The best way to make these determinations are with a team that includes cross-functional representatives from RM, legal, risk, compliance, IT, and business unit representatives, headed by an executive sponsor.
- RM resource use is optimized, and costs are minimized by keeping records a minimum amount of time under a planned and controlled set of processes.
- Records must be retained in a repository (file room or software system) where the record is protected (e.g. made

read-only and monitored with an audit trail), so that the integrity of the record is maintained in a manner that meets all evidence and legal admissibility standards if or when litigation is encountered.

- Senior management must approve of and sign off on the retention schedule and will be legally accountable for compliance with the schedule.
- Senior management must be able to readily review retention schedules, policy documentation, and audit information to ensure users are in compliance with the retention schedule.
- Complete documentation of scheduling requirements and activities must take place so that future users and archivists can view and track changes to the retention schedule.[31]

Developing a Records Retention Schedule

A **records retention schedule** *defines the length of time that records are to be kept and considers legal, regulatory, operational, and historical requirements.*[32] The retention schedule also includes direction as to how the length of time is calculated (i.e. the event or trigger that starts the clock [e.g. two years from completion of contract]). Legal research and opinions are required, along with consultation with owners and users of the records. Users typically overestimate the time they need to keep records, as they confuse the legal requirements with their own personal wishes. Some hard questioning has to take place, since having these records or copies of records lying around the organization on hard drives, thumb drives, or in file cabinets may create liabilities for the organization.

> Records retention defines the length of time that records are to be kept and considers legal, regulatory, operational, and historical requirements.[33]

Disposition typically means destruction of a file series once it has met its lifecycle retention requirements. However, it can mean not just destruction but also archiving, or transfer and a change in ownership and responsibility for the records (such as is often done in the US federal government when records are transferred to NARA). The processes of archiving and preserving are an example where records may be handed over to a historical recordkeeping unit. At this time, the records may be sampled and only selective parts of the group of records may be retained.

> Disposition means not just destruction but can also mean archiving and a change in ownership and responsibility for the records.

Why Are Retention Schedules Needed?

A retention schedule allows for uniformity in the retention and disposition process, regardless of the media or location of the records. Further, it tracks, enforces, and audits the retention and disposition of records while optimizing the amount of records kept to legal minimums, which saves on capital and labor costs and reduces liability (by discarding unneeded records that carry legal risk).[34] The **Generally Accepted Recordkeeping Principles** state the critical importance of having a retention schedule (see the section "Generally Accepted Recordkeeping Principles" in Chapter 3 for more details) and provides guidelines for open collaboration in developing one. In the public sector, holding records that have passed their legally required retention period also can have negative ramifications and liabilities in meeting information service requests made during litigation, compliance actions, or, for example, under the US FOIA, or similar acts in other countries.

> A retention schedule allows for uniformity in the retention and disposition process, regardless of the media or location of the records.

Information Included on Retention Schedules

A retention schedule consists of these components:

- *Title* of the record series
- *Descriptions* of the records series
- *Office responsible* for the retention of the record (default is usually the office of origin)
- *Disposal decision*—destroy, transfer to the archives, or, in exceptional circumstances, reconsider at a later (specified) date

- *Timing of disposal*—a minimum period for which the records should be retained in the office or in an off-site store before disposal action is undertaken
- *Event that triggers* the disposal action
- *Dates on which the schedule was agreed*, signed, or modified
- *Legal citations or a link to a citation* that reference the retention requirements of that group of records

A sample of a simple records retention schedule is shown in Figure 9.4.

Records Retention Schedule	**ENVIRONMENTAL HEALTH AND SAFETY**
December 10, 2015	

Record Type	Responsible Department	Event	Retention Period
Accident/Injury Reports *Includes:* *Accidents* *Diagnosis (Accident or Injury)* *First aid reports* *Injuries* *Medical reviews* *Occupational Health Incident* *Treatment and Progress (Accident or Injury)* *Work-related accidents* *Workers' health information* *Workers' Compensation Claims*	HR	Date of Incident	E+30
Employee Medical Files *Includes:* *Audiology* *Lung Function* *Return to Work Authorization* *Related to:* *Employee Files (Active)*	HR	Termination	E+30
Health and Safety Programs *Includes:* *Health and Safety Committee* *Health and Safety Reports*	Health and Safety		CY+10

Figure 9.4 Sample Records Retention Schedule

Source: IMERGE Consulting, Inc.

Steps in Developing a Records Retention Schedule

If you already have existing retention schedules but are revising and updating them, there may be useful information in those schedules that can serve as a good reference point—but be wary, as they may be out of date and may not consider current legal requirements and business needs.

According to the US National Archives, some key steps are involved in developing retention schedules:

1. Review the functions and recordkeeping requirements for the [business unit or] agency or the organizational component of the agency whose records will be included on the schedule.
2. Inventory the records.
3. Determine the period of time the records are needed for conducting [business or] agency operations and meeting legal obligations.
4. Draft disposition instructions including:
   - File cutoffs or file breaks (convenient points within a filing plan/system, such as end of a letter of the alphabet, end of year, or month, etc., at which files are separated for purposes of storage and/or disposition)
   - Retention periods for temporary records
   - Instructions for transferring permanent records to the National Archives of the United States [or corporate archive for businesses]
   - Instructions for sending inactive records to off-site storage
   - Organize the schedule and clear it internally
   - Obtain approval from [your corporate archivist or] NARA [for federal agencies], as well as from GAO if required by Title 8 of the GAO, "*Policy and Procedures Manual for the Guidance of Federal Agencies*."[35]

What Records Do You Have to Schedule? Inventory and Classification

Inventory and classification are prerequisites for compiling a retention schedule. Before starting work, develop an **information map** or **data map** that shows where information is created, where it resides, and the path it takes. What records are created, who uses them, and how is their disposition handled? Questions like these will provide key insights in the development of the retention schedule.[36] Confirm that the information map covers all the uses of the records by all parts of the organization, including use for accountability, audit, and reference purposes.

> An information map is a critical first step in developing a records retention schedule. It shows where information is created, where it resides, and who uses it.

In the absence of a formal information map, at a minimum *you must compile a list of all the different types of records in each business area*. This list should include information about who created them and what they are used for (or record **provenance**), which parts of the organization have used them subsequently and for what purpose (its **usage**), and the actual **content**.

In the absence of any existing documentation or records inventory, you will need to conduct a records inventory or survey to find out what records the business unit (or organization) holds. Tools are available to scan e-records folders to expedite the inventory process. A retention schedule developed in this way will have a shorter serviceable life than one based on an information map because it will be based on existing structures rather than functions and will remain usable only as long as the organizational structure remains unchanged.

> Tools are available to scan e-records folders to expedite the inventory process.

Once a records inventory or survey is complete, building a records retention schedule begins with **classification** of records.[37]

This basic classification can be grouped into three areas:

1. Business functions and activities
2. Records series
3. Document types

Business functions are basic business units such as accounting, legal, human resources, and purchasing. (See Appendix A, Information Organization and Classification: Taxonomies and Metadata for details on the process of developing classifications.) It basically answers this question: *What were you doing when you created the record?*

Business activities are the tasks *performed* to accomplish the business function. Several activities may be associated with each function.

A **records series** (or **file series**) *is a group or unit of identical or related records that are normally used and filed as a unit* and that can be evaluated as a unit or business function for scheduling purposes.

A **document type** is a term used by many software systems to refer to a grouping of related records. When the records are all created by similar processes, then the document type is equivalent to the business functions or activities mentioned previously. However, "document type" often refers to the format of the record (e.g. presentation, meeting minutes). In this case, there is not enough information to determine a retention period, because it is ambiguous regarding what type of work was being done when that document was created. Retention schedules require that record series be defined by business function and activity, not by record format or display type.

> After completing an inventory, developing a retention schedule begins with records classification.

Rationale for Records Groupings

Records are grouped together for fundamental reasons to improve information organization and access. These reasons include:

- Grouping by "similar theme" for improved completeness
- Improving information search speed and completeness
- Increasing organizational knowledge and memory by providing the "context" within which individual documents were grouped

- Clearly identifying who the record owner or creator is and assigning and tracking responsibility for a group of records
- Grouping records with the same retention requirements for consistent application of disposition processes to records

Records Series Identification and Classification

After completing a records inventory including characterizing descriptive information about the records such as their contents, use, file size, and projected growth volumes, you will need to interview staff in those target areas you are working with to determine more information about the specific organizational structure, its business functions, services, programs, and plans.[38]

In the course of business, there are several different types of records series. There are **case records**, for example, which are characterized as having a beginning and an end but are added to over time. Case records generally have titles that include names, dates, numbers, or places. These titles do not provide insight into the nature of the function of the record series. Examples of case records include personnel files, mortgage loan folders, contract and amendment/addendum records, accident reports, insurance claims, and other records that accumulate and expand over time. Although the contents of case files may be similar, you should break out each type of case record under a unique title.

Subject records (also referred to as **topic** or **function records**) "contain information relating to specific or general topics and that are arranged according to their informational content or by the function/activity/transaction they pertain to."[39] These types of records accumulate information on a particular topic or function to be added to the organization's memory and make it easier for knowledge workers to find information based on subject matter, topics, or business functions. Records such as those on the progression of relevant laws and statutes, policies, standard operating procedures, and education and training have long-term reference value and should be kept until they are no longer relevant or are displaced by more current and relevant records. In a record retention schedule, the trigger event often is defined as *superseded or obsolete*. Records of this type that relate to "routine operations of a [project], program or service" do not have as much enduring value and should be scheduled to be kept for a shorter period.

Retention of E-Mail Records

Are e-mail messages records? This question has been debated for years. *The short answer is no, not all e-mail messages constitute a record*. But how do you determine whether certain messages are a business record or not? The general answer is that a record documents a transaction or business-related event that may have legal ramifications or historic value. Most important are business activities that may relate to compliance requirements or those that could possibly come into dispute in litigation. Particular consideration should be given to financial transactions of any type.

Certainly evidence that required that governance oversight or compliance activities have been completed needs to be documented and become a business record. Also, business transactions, where there is an exchange of money or the equivalent in goods or services, are also business records. Today, these transactions are often documented by a quick e-mail. And, of course, any contracts (and any progressively developed or edited versions) that are exchanged through e-mail become business records.

> Not all e-mail messages are records; those that document a business transaction or progress toward it are clearly records and require retention.

The form or format of a potential record is irrelevant in determining whether it should be classified as a business record. For instance, if a meeting of the board of directors is recorded by a digital video recorder and saved to DVD, it constitutes a record. If photographs are taken of a groundbreaking ceremony for a new manufacturing plant, the photos are records too. If the company's founders tape-recorded a message to future generations of management on reel-to-reel tape, it is a record also, since it has historical value. But most records are going to be in the form of paper, microfilm, or an electronic document.

Here are three guidelines for determining whether an e-mail message should be considered a business record:

1. *The e-mail documents a transaction or the progress toward an ultimate transaction where anything of value is exchanged between two or more parties*. All parts or characteristics of the transaction, including who (the parties to it), what, when, how much, and the composition of its components are parts of the transaction. Often seemingly minor parts of a transaction are found buried within an e-mail message. One example would be a last-minute discount offered by a supplier based on an order being placed or delivery being made within a specified time frame.
2. *The e-mail documents or provides support of a business activity occurring that pertains to internal corporate governance policies or compliance* to externally mandated regulations.
3. *The e-mail message documents other business activities that may possibly be disputed in the future*, whether it

ultimately involves litigation or not. (Most business disputes actually are resolved without litigation, provided that proof of your organization's position can be shown.) For instance, your supplier may dispute the discount you take that was offered in an e-mail message and, once you forward the e-mail thread to the supplier, it acquiesces.

> E-mail messages that document business activities, especially those that may be disputed in the future, should be retained as records.

Managing e-mail business records is challenging, even for technology professionals. According to an AIIM and ARMA survey, *fully two-thirds of records managers doubt that their IT departments really understand the concept of electronic records life cycle management*. That is despite the fact that *70% of companies rely on IT professionals alone to manage their electronic records*.

Although the significance of e-mail in civil litigation cannot be overstated (it is the leading piece of evidence requested at civil trials today), *one-third of IT managers state that they would be incapable of locating and retrieving e-mails that are more than one year old*, according to Osterman Research.[40]

How Long Should You Keep Old E-Mails?

There are different schools of thought on e-mail retention periods and retention schedules. The retention and deletion of your electronic business records may be governed by laws or regulations. *Unless your organization's e-mail and ESI records are governed by law or regulations, your organization is free to determine the retention periods and deletion schedules that are most appropriate for your organization.*[41] If your organization's e-mail retention periods are not specified by law or regulation, consider keeping them for at least as long as you retain paper records. Many software providers provide automated software that allows e-mail messages to be moved to controlled repositories as they are declared to be records.

Destructive Retention of E-Mail

A destructive retention program is an approach to e-mail archiving where e-mail messages are retained for a limited time (say, 90 days), followed by the permanent manual or automatic deletion of the messages from the organization network, so long as there is no litigation hold or the e-mail has not been declared a record.

E-mail retention periods can vary from 90 days to as long as seven years:

- Osterman Research reports that almost one-quarter of companies delete e-mail after three months.
- Heavily regulated industries, including energy, technology, communications, and real estate, favor archiving for one year or more, according to Fulbright and Jaworski research.[42]

The most common e-mail retention period traditionally has been seven years; however, some organizations are taking a hard-line approach and stating that e-mails will be kept for only 90 days or six months, unless it is declared as a record, classified, and identified with a classification/retention category, and tagged or moved to a repository where the integrity of the record is protected (i.e. the record cannot be altered and an audit trail on the history of the record's usage is maintained).

> Destruction retention of e-mail is a method whereby e-mail messages are retained for a limited period and then destroyed.

Long-Term Archival Records

Inactive records that have historical value or are essential for maintaining corporate memory must be kept the longest. Although they are not needed for present operations, they still have some value to the organization and must be preserved. When it comes to preserving electronic records, this process can be complex and technical. (See Chapter 17 for details.) If you have a corporate or agency archivist, his or her input is critical.

Meeting Legal Limitation Periods

A key consideration in developing retention schedules is researching and determining the minimum time required to keep records that may be demanded in legal actions. “A **limitation period** is the length of time after which a legal action cannot be brought before the courts. Limitation periods are important because they determine the length of time records must be kept to support court action [including subsequent appeal periods]. It is important to be familiar with the purpose, principles, and special circumstances that affect limitation periods and therefore records retention.”[43]

Legal Requirements and Compliance Research

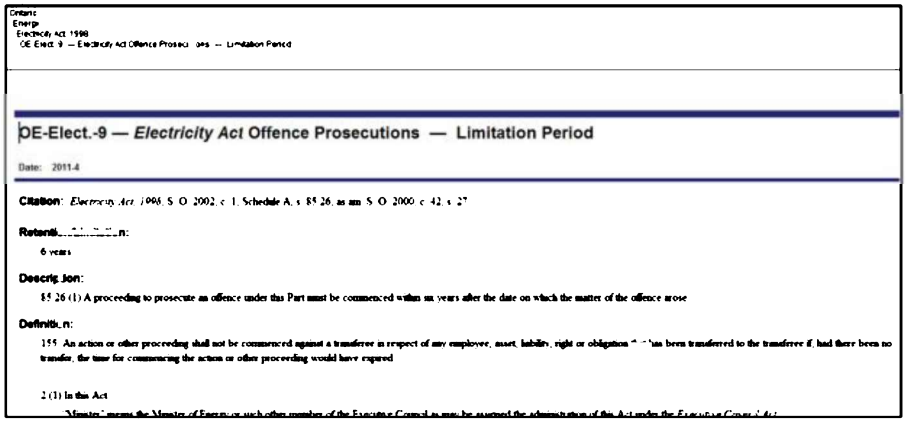
Legal requirements trump all others. The retention period for a particular records series must meet minimum retention requirements as mandated by law. Business needs and other considerations are secondary. So, legal research is required before determining retention periods. Legally required retention periods must be researched for each jurisdiction (state, country) in which the business operates, so that it complies with all applicable laws.

In order to locate the regulations and citations relating to retention of records, there are two basic approaches. The first approach is to use a records retention citation service, which publishes in electronic form all of the retention-related citations. These services usually are bought on a subscription basis, as citations are updated on an annual or more frequent basis as legislation and regulations change.

Figure 9.5 is an excerpt from a Canadian records retention database product called FILELAW®. In this case, the act, citation, and retention periods are clearly identified.

Another approach is to search the laws and regulations directly using online or print resources. Records retention requirements for corporations operating in the United States may be found in the **Code of Federal Regulations** (CFR), the annual edition of which

> is the codification of the general and permanent rules published in the Federal Register by the departments and agencies of the federal government. It is divided into 50 titles that represent broad areas subject to federal regulation. The 50 subject matter titles contain one or more individual volumes, which are updated once each calendar year, on a staggered basis. The annual update cycle is as follows: titles 1 to 16 are revised as of January 1; titles 17 to 27 are revised as of April 1; titles 28 to 41 are revised as of July 1, and titles 42 to 50 are revised as of October 1. Each title is divided into chapters, which usually bear the name of the issuing agency. Each chapter is further subdivided into parts that cover specific regulatory areas. Large parts may be subdivided into subparts. All parts are organized in sections, and most citations to the CFR refer to material at the section level.[44]


Ontario
Energy
Electricity Act 1998
OE-Elect.-9 — Electricity Act Offence Prosecutions — Limitation Period

OE-Elect.-9 — *Electricity Act* Offence Prosecutions — Limitation Period

Date: 2011-4

Citation: *Electricity Act 1998*, S.O. 2002, c. 1, Schedule A, s. 85.26, as am. S.O. 2000, c. 42, s. 27

Retenti... n:

6 years

Descrig..tion:

85.26 (1) A proceeding to prosecute an offence under this Part must be commenced within six years after the date on which the matter of the offence arose

Definiti..n:

155 An action or other proceeding shall not be commenced against a transferee in respect of any employee, asset, liability, right or obligation ... has been transferred to the transferee if, had there been no transfer, the time for commencing the action or other proceeding would have expired

2 (1) In this Act

Figure 9.5 Excerpt from Canadian Records Retention Database

Source: Ontario, Electricity Act, FILELAW database, Thomson Publishers, May 2012.

There is an up-to-date version that is not yet a part of the official CFR but is updated daily, the **Electronic Code of Federal Regulations** (e-CFR). "It is not an official legal edition of the CFR. The e-CFR is an editorial compilation of CFR material and Federal Register amendments produced by the National Archives and Records Administration's Office of the Federal Register (OFR) and the Government Printing Office."[45]

Event-Based Retention Scheduling for Disposition of E-Records

Event-based disposition is kicked off with the passage of an event, such as hiring or firing an employee, the end of a project, or the initiation of a lawsuit.

Event-based disposition can have an associated retention schedule, and the clock starts running once the event occurs. The required retention period begins only after the triggering event occurs. The length of the retention period may be regulated by law, or it may be determined by IG guidelines set internally by the organization. So, when an employee is terminated, and personnel files are destroyed after (say) five years, the retention schedule entry would be "Termination + 5 years."

One other definition of event-based disposition comes from the US e-records standard, Department of Defense 5015.2, which states that a disposition instruction in which a record is eligible for the specified disposition (transfer or destroy) upon or immediately after the specified event occurs. No retention period is applied and there is no fixed waiting period, as with "timed" or combination "timed-event" dispositions. Example: "Destroy when no longer needed for current

operations."[46]

> Event-based disposition begins with the passage of a triggering event.

Some hardware vendors provide solutions that assist in executing event-based disposition with assistance from firmware (fixed instructions on a microchip). The firmware-assisted solution should be considered if your RM or IG team aims to perform a complete and thorough retention solution analysis. These hardware-based solutions can potentially streamline the event-based disposition process.[47]

Triggering events may be record-related, "such as supersession or obsolescence." This is common to a policy statement. For example, if a group of policies are to be destroyed five years after superseded or obsolete, the old policy would be held for five years after the new policy has been created.

Sounds simple. But in an attempt to meet retention requirements, organizations handle event-based triggers in different ways, ways that often are problematic. For instance, the trigger events often are not captured electronically and fed directly into the retention scheduling software or records repository to start the clock running, or the event itself is not well documented in the retention schedule so it is not consistently being applied and tracked. In other cases, the organization simply does not have the ERM functionality it needs to manage event-based triggers.

This causes many organizations to simply overretain and keep the records indefinitely, or until disk storage is full, which means that those records are retained for an incorrect—and indefensible—time. The period is either too long or possibly too short, but it always is *always* inconsistent. *And inconsistent means legally indefensible.*

The only prudent and defensible approach is to implement the proper IG policies to manage and control the implementation of event-based disposition.

Prerequisites for Event-Based Disposition

Three key prerequisite tasks must be completed before event-based disposition can be implemented:

1. *Clarify trigger events*. Not all of the events that can trigger the beginning of a retention period are as clear as the date an employee is terminated. For instance, "contract completion date" could be the day a vendor finishes work, when a final invoice is rendered, when the invoice is paid, or some other period, such as 30 days following the payment of the final invoice. These definitions, depending on the record series in question, may be regulated by law or governed by IG policies.

   What is needed is an agreement as to what the definition is, so that the retention period will be uniform among the record series in question, providing a defensible policy.

   To gain this agreement on these blurry areas, the RM lead/manager or team will need to work with the relevant business unit representatives, IT, compliance, risk management, and any other stakeholders.

   The event triggers must be clear and agreed on so that they may kick off a retention period and disposition process.

   In a number of cases, the answer to these questions will rely on trigger points, such as one year after completion or four months after the board of directors' meeting. *It is important to choose a* trigger point *that you can implement*. For example, there is no point in saying that records should be kept until an individual dies, if you have no reliable way of knowing the person is alive. Instead, choose a trigger point based on the information you have about the individual; in this case, the 100th birthday might be a suitable trigger point.

2. *Automated capture of agreed-on trigger events must be performed and sent to the ERM*. It is easy to know an employee's termination date—most human resources management systems or payroll systems can supply it—but other types of events are not so easily captured and may require some customization in order that this information is fed into an ERM. The metadata about the event must be seamlessly entered into the ERM so that it may launch the beginning of the retention period. If systems external to the ERM need to be interfaced, a common locater (e.g. contract number) can link the two.

3. *The ERM systems must have complete retention and disposition capabilities*. In order for the retention to start properly and run to final disposition, this tracking capability must be an inherent feature of the software. (In some cases, organizations may use specialized retention and disposition software that can perform this task minimally without complete ERM functionality, but it falls short of the type of richness that a robust ERM system provides. What is needed is the ability to include the details or retention rules beyond simple date calculations (i.e. to store descriptive data or scope notes and records series code in addition to retention requirements, which are automatically associated with the retention rule, and to have a records hold-and-release capability). If destruction is the final disposition, then the system must be able to perform a deletion of the record (so long as there is no preservation or legal hold) with no traces that can allow reconstruction of it, and this process must be verifiable.

   To accomplish clarity and agreement on event-based triggers requires close consultation and collaboration among

RM staff, business units, IT, legal, compliance, risk management, and other stakeholders, as relevant.

Final Disposition and Closure Criteria

After completing the records values analysis and legislative and legal research, you must determine the closure criteria and final disposition (e.g., destroy, transfer, archive) for each records series. To minimize costs and litigation risk, retention periods should be kept as short as possible while meeting all applicable regulatory, legal, and business requirements.[48]

Retention Periods: Online versus Offline

For e-records, retention periods may be segmented into active and inactive, or online and offline. Offline may be segmented further into on-site and off-site or archival storage.

Going back and combing through records retrieval requests and usage logs may provide helpful insights as to the needs of records users—but bear in mind that these logs may be misleading as users may have (in the past, before a formal IG program was implemented) kept shadow copies of files on their local hard drives or backed up to flash drives or other storage devices.

Closure Dates

A clear closure start date is required to kick off a retention period for any record, whether the retention is scheduled for on- or off-site. Calendar or fiscal year-ends are typical and practical closure dates for subject or topical records. The date used to indicate the start year is usually the date the file closed or the date of last use or update. In a university setting, school year-end may be more logical. Still, a reasoned analysis is required to determine the best closure start date for subject records in your organization.

Case records are different; logically, their closure date is set when a case record is completed (e.g. the date when an employee resigns, retires, or is terminated).

Future dates may be used, such as an employee promotion date, student graduation, or project completion. After consulting those who create and handle the records series you are analyzing, apply good business judgment and common sense when determining closure dates.

Retaining Records Indefinitely

There may be some vital, historical, or other critical records that, in the best interests of the organization, need to be retained permanently. This is rare, and storing records long term must be scrutinized heavily. If certain electronic records are to be retained indefinitely or permanently, then LTDP policies and techniques must be used. (See Chapter 17 for more details.)

Retaining Transitory Records

Transitory documents usually do not rise to the level of becoming a record; they are temporary and are useful only in the short term, such as direct mail or e-mail advertising (brochures, price lists, etc.), draft documents (although not all are transitory, and some may need longer retention periods, such as draft contracts) and work in progress, duplicates, external publications (e.g. magazines, journals, newspapers, etc.), and temporary notices (e.g. company picnic, holiday party, or football pool). You must consider transitory records in your master records retention schedule.

Implementation of the Retention Schedule and Disposal of Records

Automated programs that interpret these retention periods are the best way to ensure that records are disposed of at the correct time and that an audit trail of the disposition is maintained.

Getting Acceptance and Formal Sign-off of the Retention Schedule

Upon completion of the records retention schedule, project management best practices dictate that it be signed off by an executive or project sponsor, to indicate it has been completed and there is no more work to be done on that phase of the project. In addition, you may want to gain the sign-off and acceptance by other key stakeholders, such as senior representatives from legal, IT, the board of directors or executive committee, and perhaps audit and information governance. The schedule should be updated when new record types are introduced and, in any case, at least annually.

Disposition Timing: Records Disposal

It is much easier to time or schedule the disposal of e-records than of paper or physical records, but true and complete destruction of all traces of a record cannot be done by hitting a simple “delete” key. There must be a process in place to verify the total destruction of all copies of the record. (See Chapter 17 for more details.) Records destruction can occur daily, routinely, or be scheduled at intervals (i.e. monthly or quarterly).

Automating Retention/Disposal Actions

ERM systems typically are capable of automatically executing a record deletion when a record has reached the end of its life cycle. Often these systems have a safety feature that allows an operator who has the authority to review deletions before they are performed.

Disposal Date Changes

To make a retention schedule change, such as extending the life of a record series, IG controls must be in place. So, usually, ERM systems require that a person of higher authority than the system operator make these approvals. Every subsequent delay in destroying the records often requires an escalation in approval period to extend the time that records are kept past the destruction date.

Proving Record Destruction

In some environments, especially in the public sector, a certificate of destruction or other documentation is required to prove that a record and all its copies have been completely deleted (including its metadata—although at times it is beneficial to retain metadata longer than the record itself; see Appendix A "Information Organization and Classification" for more details). ERM systems can be configured to keep an audit trail and prove that destruction has occurred.

Ongoing Maintenance of the Retention Schedule

Records series are not static; they change, are added to, and are amended. New record functions emerge, based on changes in business, acquisitions, and divestitures. So it is necessary for organizations to review and update—at least annually—their records retention schedule.

In addition, retention requirements change as legislation changes, lawsuits are filed, and the organization refines and improves its IG policies. Development of a records retention schedule is not a one-time project; it requires attention, maintenance, and updating on a regular schedule, and using a controlled change process.

Audit to Manage Compliance with the Retention Schedule

Once your organization establishes records retention schedules for business units, or a master retention schedule, there must be IG policies in place to audit and ensure that policies are being followed. *This is a key requirement of maintaining a legally defensible retention schedule that will hold up to legal challenges*.

CHAPTER SUMMARY: KEY POINTS

- According to ISO, a record is "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business."
- RM is "[the] field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use, and disposition of records, including the processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records."
- ERM includes the management of electronic and nonelectronic records, such as paper and other physical records.
- ERM has become much more critical to enterprises with increased compliance legislation and massively increasing volumes of electronic information.
- ERM follows the same basic principles as paper-based records management.
- A number of factors provide the business rationale for ERM, including facilitating compliance, supporting IG, and providing backup capabilities in the event of a disaster.
- Implementing ERM is challenging since it requires user support and compliance, adherence to changing laws, and support for new information delivery platforms like mobile and cloud computing.
- ERM benefits are both tangible and intangible or difficult to calculate.
- Improved professionalism, preserving corporate memory, support for better decision making, and safeguarding vital records are key intangible benefits of ERM.
- NARA recommends that e-records are inventoried by information system rather than file series, which is the traditional approach for physical records.
- Generally Accepted Recordkeeping Principles are "information management and governance of record creation, organization, security, maintenance and other activities used to effectively support recordkeeping of an organization."
- It may be helpful to use a recordkeeping methodology such as the Principles or DIRKS to guide inventorying

efforts.

- Perhaps the organization has a handle on their paper and microfilmed records, but e-records have been growing exponentially and spiraling out of control.
- Whatever the business goals for the inventorying effort are, they must be conveyed to all stakeholders, and that message must be reinforced periodically and consistently, and through multiple means.
- An appropriate scope might enumerate the records of a single program or division, several functional series across divisions, or records that fall within a certain time frame versus an entire enterprise.
- The completed records inventory contributes toward the pursuit of an organization's IG objectives in a number of ways.
- There are four basic ways to conduct the inventory: surveys, interviews, observation, and software tools. Combining these methods yields the best results.
- Additional information not included in inventories of physical records must be collected in any inventory of e-records.
- Be sure to tie the findings in the final report of the records inventory to the business goals that launched the effort.
- Records appraisal is based on the information contained in the records inventory.
- An Information Asset Register is a sort of general ledger of information assets that lists all information assets, structured and unstructured, their lifecycle retention, privacy and security requirements, where they are housed, and even what hardware is used.
- Records can have different types of value to organizations: historical, administrative, regulatory and statutory, legal, fiscal, or other archival value as determined by an archivist.
- Consistency in managing records across an enterprise, regardless of media, format, or location, is the key to compliance.
- A complete, current, and documented records retention program reduces storage and handling costs and improves searchability for records by making records easier and faster to find.
- Retention schedules are developed by records series—not for individual records.
- Retention schedules are basic tools that allow an organization to prove that it has a legally defensible basis on which to dispose records.
- The master retention schedule contains all records series in the entire enterprise.
- Records retention defines the length of time that records are to be kept and considers legal, regulatory, operational, and historical requirements.
- “Disposition” means not just destruction but can also mean archiving and a change in ownership and responsibility for the records.
- An information map is a critical first step in developing a records retention schedule. It shows where information is created, where it resides, and who uses it.
- After inventorying, developing a retention schedule begins with records classification.
- All e-mail messages are not records; those that document a business transaction, or progress toward it, are clearly records and require retention.
- E-mail messages that document business activities, especially those that may be disputed in the future, should be retained as records.
- Destruction retention of e-mail is a method whereby e-mail messages are retained for a limited period and then destroyed.
- Tools are available to scan e-records folders to expedite the inventorying process.
- Assessing the relative value of records is key to determining their retention periods and disposition path.
- Records have different types of value, such as financial, legal, technical, and administrative/operational.
- Event-based disposition begins with a triggering event.
- Retention schedules, once established, must be maintained and updated to add new records series, as appropriate, and to comply with new or changed legislation and regulatory requirements.
- Auditing to ensure compliance with established retention policies is key to maintaining a legally defensible records retention program.

Notes

1. International Organization for Standardization, *ISO 15489-1: 2001 Information and Documentation—Records Management. Part 1: General* (Geneva: ISO, 2001), section 3.15.
2. Ibid., section 3.16.
3. Ibid.
4. Ibid.
5. U.S. Environmental Protection Agency, "Why Records Management? Ten Business Reasons," updated March 8, 2012, https://www.epa.gov/records.
6. U.S. National Archives and Records Administration, *Disposition of Federal Records: A Records Management Handbook*, 2000, Web edition, www.archives.gov/records-mgmt/publications/disposition-of-federal-records/chapter-3.html.
7. Ibid.
8. State and Consumer Services Agency Department of General Services, *Electronic Records Management Handbook*, State of California Records Management Program, February 2002, www.documents.dgs.ca.gov/osp/recs/ermhbkall.pdf.
9. US Environmental Protection Agency, "Six Steps to Better Files," updated March 8, 2012, www.epa.gov/records/tools/toolkits/6step/6step-02.htm.
10. Moore's Law was coined by Gordon Moore (co-founder of Intel). In 1963, Moore predicted that computer chip–design technology would exponentially double every 24 months, while the cost of computers would decline by 50% during the same period. Moore's law has been accurate, although the doubling is occurring at 18-month intervals.
11. In the United States, "private information" or "personally identifiable information" is defined by state law, unless specifically defined by a federal statute. Types of information usually specified include social security numbers, driver's license numbers, e-mail accounts with passwords, health insurance, medical records, and biometrics.
12. Target's chief executive, Greg Steinhafel, was forced to resign when 40 million card accounts were breached along with 70 million instances of PII breach. The CFO of the Fortelus hedge fund, Thomas Meston, lost his job when he was duped into allowing the change of security codes by a phone caller late on a Friday afternoon. Most recently, Donna Seymore, CIO for the OPM, was named as a defendant in a class action suit brought by federal employees' union.
13. *Remijas et al. v. Neiman Marcus Group*, case 14-3122 (7th Cir. 2015). In December 2014, 350,000 credit card accounts were breached. Plaintiffs allege that Neiman Marcus deliberately delayed its notification of customers to take advantage of the holiday sales. On a motion to dismiss, the court ruled that the breach itself was enough to infer damages to plaintiffs.
14. In its 2012 complaint against Wyndham, the Federal Trade Commission alleged that Wyndham's privacy policy misrepresented the security measures that the company and its subsidiaries took to protect consumers' personal information, and that its failure to safeguard personal information caused substantial consumer injury. The agency charged that the security practices were unfair and deceptive and violated the FTC Act.
15. Immediate pressures often undermine long-term strategies to build solid foundations for future information growth, particularly regarding defensible deletion.
16. Margaret Rouse, "Generally Accepted Recordkeeping Principles," updated March 2011, https://searchcompliance.techtarget.com/definition/Generally-Accepted-Recordkeeping-Principles (accessed December 12, 2018).
17. Ibid.
18. Ibid.
19. Public Record Office, "Guidance for an Inventory of Electronic Record Collections: A Toolkit," September 2000, www.humanrightsinitiative.org/programs/ai/rti/implementation/general/guidance:for_inventory_elect_rec_collection.pdf, pp. 5–6. (accesssed December 12, 2018).
20. Ibid.
21. National Archives, "Frequently Asked Questions about Records Inventories," updated October 27, 2000, https://www.archives.gov/records-mgmt/scheduling/inventory-intro (accessed December 12, 2018).

22. William Saffady, "Managing Electronic Records," 4th ed., *Journal of the Medical Library Association*, 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2947138/ (accessed December 12, 2018).

23. Ibid.

24. Maryland State Archives, "Retention Schedule Preparation," https://msa.maryland.gov/msa/intromsa/html/record_mgmt/toc.html (accessed December 12, 2018)

25. Dennis Kessler, "Where Do You Keep the Crown Jewels? Identifying, Classifying and Managing Your Information Assets," in *Information Governance for Executives,* ed. Robert Smallwood (San Diego, CA: Bacchus Business, 2016), 153–154.

26. Reynold Leming, "25 Exciting Things to Do With an Information Asset Register," in *Information Governance for Healthcare Professionals,* ed. Robert Smallwood (Boca Raton, FL: CRC Press, 2018), 106–109

27. Ibid.

28. Kessler, "Where Do You Keep the Crown Jewels?," 156–158.

29. Government of Alberta, "Developing Retention and Disposition Schedules."

30. National Archives, "Disposition of Federal Records."

31. Government of Alberta, "Developing Retention and Disposition Schedules."

32. National Archives, "Frequently Asked Questions about Records Scheduling and Disposition," http://www.archives.gov/records-mgmt/faqs/scheduling.html#steps (accessed December 12, 2018).

33. Ibid.

34. University of Edinburgh, Records Management Section, https://www.ed.ac.uk/records-management (accessed June 7, 2019).

35. National Archives, "Frequently Asked Questions about Records Scheduling and Disposition."

36. University of Edinburgh, Records Management Section.

37. National Archives, "Frequently Asked Questions about Records Scheduling and Disposition."

38. Government of Alberta, "Developing Retention and Disposition Schedules."

39. Ibid.

40. Marty Foltyn, "Getting Up to Speed on FRCP," June 29, 2007, www.enterprisestorageforum.com/continuity/features/article.php/3686491/Getting-Up-To-Speed-On-FRCP.htm.

41. Nancy Flynn, *The E-Policy Handbook* (New York: AMACOM, 2009), pp. 24–25.

42. Mary Flood, Survey: They see a more litigious future, posted on October 18, 2010, http://blog.chron.com/houstonlegal/2010/10/survey-they-see-a-more-litigious-future/.

43. Government of Alberta, "Developing Retention and Disposition Schedules," p. 122.

44. U.S. Government Printing Office, *Code of Federal Regulations*, www.gpo.gov/help/index.html#about_code_of_federal_regulations.htm (accessed December 12, 2018).

45. U.S. National Archives and Records Administration, "Electronic Code of Federal Regulations," https://www.archives.gov/about/regulations/regulations.html (accessed December 12, 2018).

46. Department of Defense, "Design Criteria Standard for Electronic Records Management Software Applications," https://www.archives.gov/records-mgmt/policy/cots-eval-guidance.html (accessed December 12, 2018).

47. Craig Rhinehart, IBM, e-mail to author, July 30, 2012.

48. Government of Alberta, "Records and Information Management."

* Portions of this chapter are adapted from Chapters 1, 5, and 7 of Robert F. Smallwood, *Managing Electronic Records: Methods, Best Practices, and Technologies*, © John Wiley & Sons, Inc., 2013. Reproduced with permission of John Wiley & Sons, Inc.

CHAPTER 10
Information Governance and Information Technology Functions

Information technology (IT) is a core function that contributes to, and is impacted by information governance (IG) program efforts. IT departments typically have been charged with keeping the "plumbing" of IT intact—the network, servers, applications, and data. However, while the output of IT is in their custody, they have not been held to account for it; that is, the information, reports, and databases they generate have long been held to be owned by users in business units. This has left a gap of responsibility for governing the information that is being generated and managing it in accordance with legal and regulatory requirements, standards, and best practices.

Certainly, on the IT side, shared responsibility for IG means the IT department itself must take a closer look at IT processes and activities with an eye to IG. A focus on improving IT efficiency, software development processes, and data governance and quality will help contribute to the overall IG program effort. IT is an integral piece of the program.

Debra Logan, vice president and distinguished analyst at Gartner, states:

> Information governance is the only way to comply with regulations, both current and future, and responsibility for it lies with the CIO and the chief legal officer. When organizations suffer high-profile data losses, especially involving violations of the privacy of citizens or consumers, they suffer serious reputational damage and often incur fines or other sanctions. IT leaders will have to take at least part of the blame for these incidents.[1]

Gartner predicted that the need to implement IG is so critical that significant numbers of chief information officers (CIOs) will be terminated for their inability to implement IG successfully. Data breaches, ransomware attacks, and significant system downtime all end up on the CIO's doorstep. And if serious enough, the CEO can even be held to account for IT department deficiencies and mistakes.

Aaron Zornes, chief research officer at the MDM (Master Data Management) Institute, stated: "While most organizations' information governance efforts have focused on IT metrics and mechanics such as duplicate merge/purge rates, they tend to ignore the industry- and business-metrics orientation that is required to ensure the economic success of their programs."[2]

Four IG best practices in this area can help CIOs and IT leaders to be successful in delivering business value as a result of IG efforts:

1. *Make the business case for IG by tying its objectives to business objectives.*

   To garner the resources and time needed to implement an IG program, you must develop a business case in real, measurable terms and tie it to corporate objectives. When executives see this alignment of objectives, they are more likely to support an IG program. The business case must be presented in order to gain executive sponsorship, which is an essential component of any IG effort. Without executive sponsorship, the IG effort will fail. Making the business case and having metrics to measure progress and success toward meeting business objectives are absolute musts.

2. *Don't focus on technology, focus on business impact.*

   Technology often fascinates those in IT—to the point of obfuscating the reason that technologies are leveraged in the first place: to deliver business benefit. Therefore IT needs to reorient its language, its vernacular, its very focus when implementing IG programs. IT needs to become more business savvy, more businesslike, more focused on delivering business benefits that can help the organization to meet its business goals and achieve its business objectives. "Business leaders want to know why they should invest in an information governance program based on the potential resulting business outcomes, which manifest as increased revenues, lower costs and reduced risk."[3]

3. *Customize your IG approach for your specific business, folding in any industry-specific best practices possible.*

   You cannot simply take a boilerplate IG plan, implement it in your organization, and expect it to be successful. Sure, there are components that are common to all industries, but tailoring your approach to your organization is the only way to deliver real business value and results. That means embarking on

an earnest effort to develop and sharpen your business goals, establishing business objectives that consider your current state and capabilities and external business environment and legal factors unique to your organization. It also means developing a communications and training plan that fits with your corporate culture. And it means developing meaningful metrics to measure your progress and the impact of the IG program, to allow for continued refinement and improvement.

4. *Standardize the use of business terms.*

   IG requires a cross-functional effort, so you must be speaking the same language, which means the business terms you use in your organization must be standardized. This is the very minimum to get the conversation started. But IG efforts will delve much more deeply into the organization of information and seek to standardize the taxonomy for organizing documents and records and even the metadata fields that describe in detail those documents and records across the enterprise.

Overall, being able to articulate the business benefits of your planned IG program will help you recruit an executive sponsor, help the program gain traction and support, and help you implement the program successfully.[4]

Several key foundational programs should support your IG effort in IT, including data governance, master data management (MDM), IT governance, and implementing accepted IT standards and best practices.

> Focusing on business impact and customizing your IG approach to meet business objectives are key best practices for IG in the IT department.

Data Governance

We touched on **data governance** in Chapter 2. In today's business environment, data is mountainous, data is growing, data is valuable, and the insights that can be gained by analyzing clean, reliable data with the latest analytics tools are a sort of new currency. This is where the principles of infonomics enter into play. There are nuggets of gold in those mountains of data. Some insights can be monetized or leveraged for economic advantage. And leveraging those discoveries can provide a sustainable competitive advantage for the organization in areas such as customer acquisition, customer retention, and customer service.

The challenge is largely in garnering control over data and in cleaning, securing, and protecting it; doing so requires effective data governance strategies. But data governance is not only about cleaning and securing data; it is also about delivering it to the right people at the right time (sometimes this means in real time) to provide strategic insights and opportunities. If a data governance program is successful, it can add profits directly to the bottom line, while improving productivity for knowledge workers.[5]

> Effective data governance can yield bottom-line benefits derived from new insights.

Data governance involves processes and controls to ensure that information at the *data* level—raw data that the organization is gathering and inputting—is true and accurate, and unique (not redundant). It involves **data cleansing** (or **data scrubbing**) to strip out corrupted, inaccurate, or extraneous data and **de-duplication** to eliminate redundant occurrences of data.

Data governance focuses on **information quality** from the ground up (at the lowest or root level), so that subsequent reports, analyses, and conclusions are based on clean, reliable, trusted data (or records) in database tables. Data governance is the most fundamental level at which to implement IG. Data governance efforts seek to ensure that formal management controls—systems, processes, and accountable employees who are stewards and custodians of the data—are implemented to govern critical data assets to improve data quality and to avoid negative downstream effects of poor data.

Data governance is a newer, hybrid *quality control discipline* that includes elements of data quality, data management, IG policy development, business process improvement, and compliance and risk management.

Good data governance programs should extend beyond the enterprise to include external stakeholders

(suppliers, customers) so an organization has its finger on the pulse of its extended operations. In other words, enforcing data governance at the earliest possible point of entry—even external to the organization—can yield significant efficiencies and business benefits downstream. And combining data governance with real-time **business intelligence** (BI) and **data analytics** software not only can yield insights into significant and emerging trends but also can provide solid information for decision makers to use in times of crisis—or opportunity.

Steps to Governing Data Effectively

Nine key steps you can take to govern data effectively are listed next:

1. *Recruit a strong executive sponsor:* As in broader IG efforts, data governance requires cross-functional collaboration with a variety of stakeholders. To drive and facilitate this sometimes contentious conversation, a strong executive sponsor is required. This is not an easy task since executives generally do not want to deal with minutia at the data level. You must focus on the realizable business benefits of improved data governance (i.e. specific applications that can assist in customer retention, revenue generation, and cost cutting).
2. *Assess your current state:* Survey the organization to see where the data repositories or silos of data are, what problems related to data exist, and where some opportunities to improve lie. Document where your data governance program stands today and then map out your road to improvement in fundamental steps.
3. *Set the ideal state vision and strategy:* Create a realistic vision of where your organization wants to go in its data governance efforts, and clearly articulate the business benefits of getting there. Articulate a measureable impact. Track your progress with metrics and milestones.
4. *Compute the value of your data.* Try to put some hard numbers to it. Calculate some internal numbers on how much value data—good data—can add to specific business units. Apply some of the formulas for calculating the value of information presented in Doug Laney's groundbreaking *Infonomics* book. Data is unlike other assets that you can see or touch (cash, buildings, equipment, etc.), and it changes daily, but it has real value.
5. *Assess risks.* What is the likelihood and potential cost of a data breach? A major breach? Are ransomware attacks on the rise in your industry? What factors come into play, and how might you combat these potential threats? Perform a risk assessment to rank and prioritize threats and assign probabilities to those threats so you may develop appropriate countermeasures.
6. *Implement a going-forward strategy.* It is a significantly greater task to try to improve data governance across the enterprise for existing data, versus focusing on smaller business units, one at a time.[6] Remember, you may be trying to fix years if not decades of bad behavior, mismanagement, and lack of governance. Taking an "incremental approach with an eye to the future" provides for a clean starting point and can substantially reduce the pain required to implement. A strategy where new data governance policies for handling data are implemented beginning on a certain future date is a proven best practice.
7. *Assign accountability for data quality to business units, not IT.* Typically, IT has had responsibility for data quality, yet the data generation is mostly not under that department's control, since most data is created in the business units. A pointed effort must be made to push responsibility and ownership for data to data stewards in the business units that create and use the data.
8. *Manage the change.* Educate, educate, educate. People must be trained to understand why the data governance program is being implemented and how it will benefit the business. The new policies represent a cultural change, and supportive program messages and training are required to make the shift.
9. *Monitor your data governance program.* See where shortfalls might be, and continue to fine-tune the program.[7]

From a risk management perspective, data governance is a critical activity that supports decision makers and can mean the difference between retaining a customer and losing one. Protecting your data is protecting the lifeblood of your business, and improving the quality of the data will improve decision making, foster compliance efforts, and yield competitive advantages.

Good data governance ensures that downstream negative effects of poor data are avoided and that subsequent reports, analyses, and conclusions are based on reliable, trusted data.

Data Governance Framework

The Data Governance Institute has created a **data governance framework**, a visual model to help guide planning efforts and a "logical structure for classifying, organizing, and communicating complex activities involved in making decisions about and taking action on enterprise data."[8] (See Figure 10.1.) The framework applies more to larger organizations, which have greater complexity, greater internal requirements, and greater, more complex regulatory demands. It allows for a conceptual look at data governance processes, rules, and people requirements.

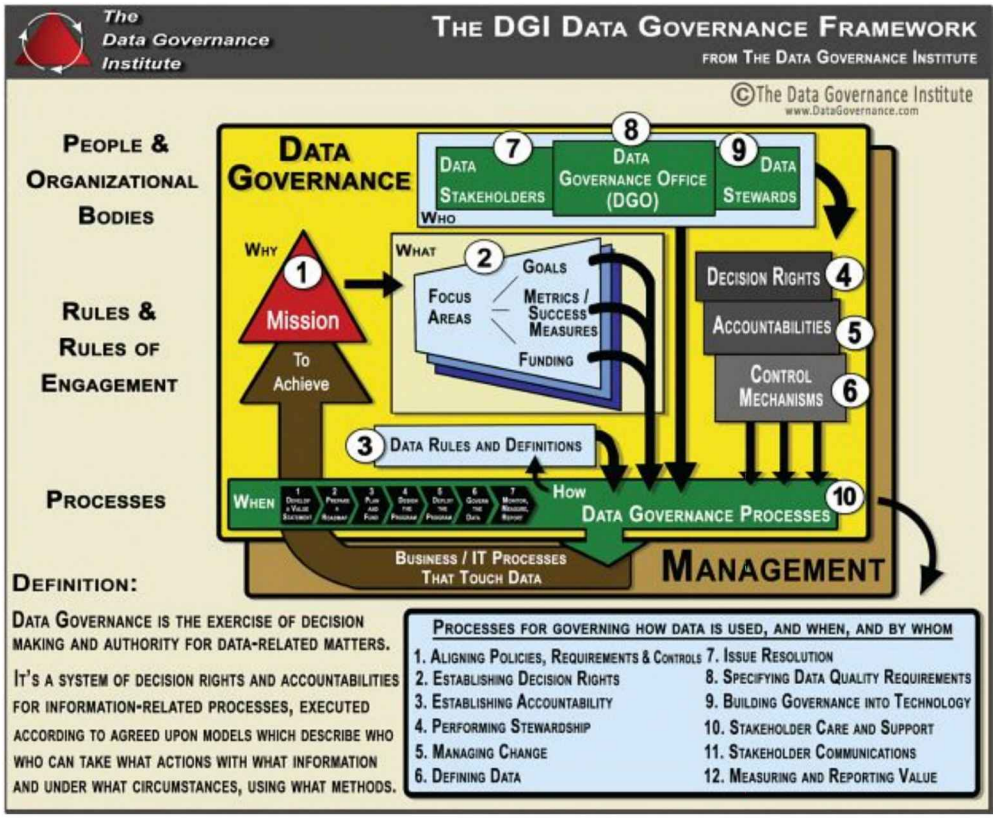


Figure 10.1 DGI Data Governance Framework™

Source: The Data Governance Institute (datagovernance.com).

Information Management

Information management is a principal function of IT. It is complex and spans a number of subdisciplines but can be defined as the "application of management techniques to collect information, communicate it within and outside the organization, and process it to enable managers to make quicker and better decisions."[9] It is about managing information, which is more than just collecting and processing data from varying sources and distributing it to various user audiences. It includes a number of subcomponent tasks, including these four key functions:

1. *Master data management* (MDM) is a key process for IG success in the IT department, which extends to involved business units. An emerging discipline, MDM came into prominence around 2010–2012, coinciding with the Big Data trend. The goal of MDM is to ensure that reliable, accurate data from a *single source* is leveraged across business units. That is, a key aim is to establish a "single version of the

truth"[10] and eliminate multiple, inconsistent versions of data sets, which are more common than most might think, especially in larger organizations with physically distributed operations and large numbers of servers and databases. MDM gets to the core of **data integrity** issues, essentially asking "Is this data true and accurate? Is this the best and only, final version?" MDM grew from the need to create a standardized, "discrete discipline" to ensure there was a single version to base analytics calculations on and to base decisions on.[11] According to Gartner, MDM:

is a technology-enabled discipline in which business and IT work together to ensure the uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise's official shared master data assets. Master data is the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise including customers, prospects, citizens, suppliers, sites, hierarchies and chart of accounts.[12]

What is the business impact? How are operations enhanced and how does that contribute to business goals? One set of reliable, clean data is critical to delivering quality customer service, reducing redundant efforts and therefore operational costs, improving decision making, monetizing data, and even potentially to lower product and marketing costs. A unified view of customers, products, or other data elements is critical to turning these business goals into reality.

Again, the larger the organization, the greater the need for MDM.

Master data management is a key IG process in IT.

2. *Information life cycle management* (ILM) is managing information appropriately and optimally at different stages of its useful life, from creation, through distribution and use, including meeting legal and regulatory requirements, and through its final disposition, which can be destruction, archiving, or transfer to another entity. Organizations historically over-retain information; however, studies show that information quickly loses its value and that once data has aged 10 to 15 days, the likelihood it will be used again is around 1%.[13] Based on its use characteristics, differing storage management strategies are appropriate. It defies business logic to manage information that has little value with as much IT resource as information that is high value. *Doing so is a misuse of resources*. To execute ILM properly, the value of certain data sets and records must be appraised and policies must be formed to manage it, recognizing that information value changes over the life cycle, which requires varying strategies and resource levels.[14] ILM conceptually includes and can begin with MDM and is linked to compliance requirements and capabilities.
3. *Data architecture* refers to the "design of structured and unstructured information systems"[15] in an effort to optimize data flow between applications and systems so that they are able to process data efficiently. Further, data architecture uses data modeling, standards, IG policies, and rules for governing data and how it populates databases and how those databases and applications are structured.[16] Some key issues to uncover when researching data architecture and design include data structure, or **schema**, which databases are used (e.g. Oracle Database 18.1.0, IBM Db2, MS SQL Server 2017), methods of query and access (e.g. SQL), the operating systems the databases operate on, and even their hardware (which can affect data architecture features and capabilities).
4. *Data modeling* can be complex, yet it is an important step in overall IG for the IT department. It "illustrates the relationships between data." Data modeling is an application software design process whereby data processes and flows between applications are diagrammed graphically in a type of flowchart that formally depicts where data is stored, which applications share it, where it moves, and the interactions regarding data movement between applications. "Data modeling techniques and tools capture and translate complex system designs into easily understood representations of the data flows and processes, creating a blueprint for construction and/or re-engineering."[17] Good data models allow for troubleshooting *before* applications are written and implemented.

   The importance of data modeling as a foundation for the application development process is depicted in Figure 10.2.

   Once the data model is developed, business rules and logic can be applied through application development. A user interface is constructed for the application, followed by movement of data or e-

documents through work steps using work flow capabilities, and then integration with existing applications (e.g. enterprise resource planning or *customer relationship management* systems). Typically this is accomplished through an **application programming interface,** a sort of connector that allows interaction with other applications and databases.

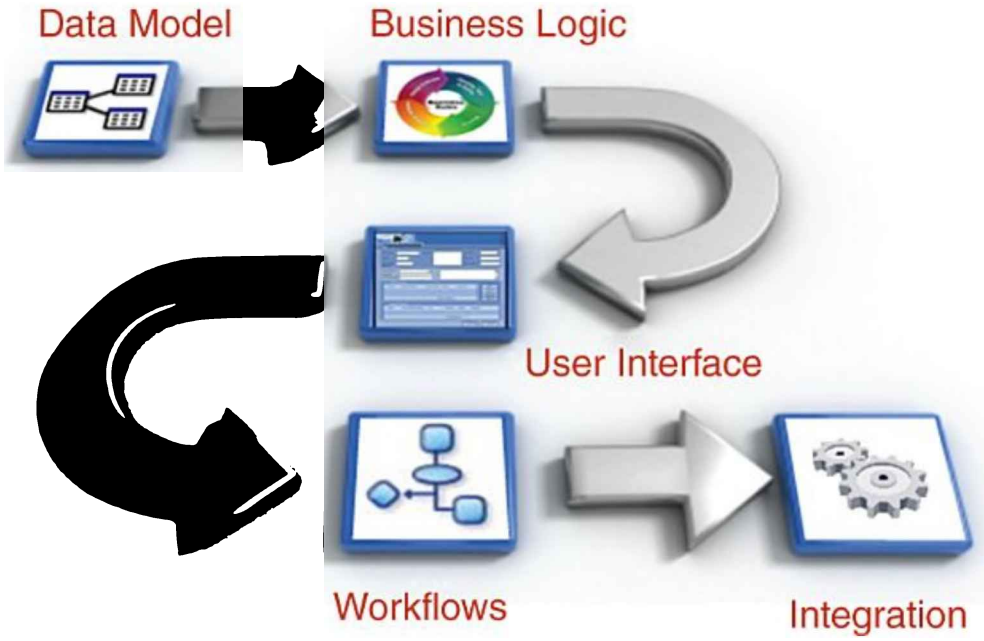


Figure 10.2 Key Steps from Data Modeling to Integration

Source: Adapted from Orangescape.com (http://www.orangescape.com/wp-content/uploads/2010/10/Application-Development-Lifecycle-OrangeScape.png).

There are six approaches to data modeling:

1. *Conceptual.* The conceptual approach merely diagrams data relationships at the "highest level"[18] showing the storage, warehousing, and movement of data between applications.
2. *Enterprise.* The enterprise approach is a more business-oriented version of conceptual data modeling that includes specific requirements for an enterprise or business unit.
3. *Logical.* Pertinent to the design and architecture of physical storage, logical data modeling "illustrates the specific entities, attributes, and relationships involved in a business function."
4. *Physical.* The physical approach depicts the "implementation of a logical data model" relative to a specific application and database system.
5. *Data integration.* This approach is just what it says; it involves merging data from two or more sources, processing the data, and moving it into a database. "This category includes Extract, Transform, and Load (ETL) capabilities."[19]
6. *Reference data management.* This approach often is confused with MDM, although they do have interdependencies. Reference data is a way to refer to data in categories (e.g. having lookup tables—standard industry classification or SIC codes) to insert values,[20] and is used only to "categorize other data found in a database, or solely for relating data in a database to information beyond the boundaries of the enterprise."[21] So reference data is not your actual data itself but a reference to categorize data.

Figure 10.3 shows different categories of data.

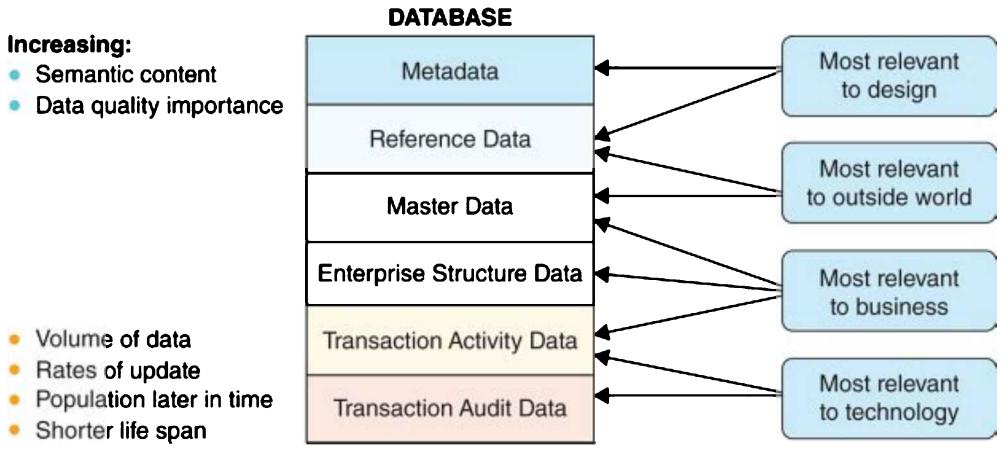


Figure 10.3 Categories of Data

Source: IBM.com.

IT Governance

As introduced in Chapter 2, IT governance is about efficiency and value creation. *IT governance is the primary way that stakeholders can ensure that investments in IT create business value and contribute toward meeting business objectives.*[22] This strategic alignment of IT with the business is challenging yet essential. IT governance programs go further and aim to improve IT performance, deliver optimum business value, and ensure regulatory compliance.

Although the CIO typically has line responsibility for implementing IT governance, the chief executive officer and board of directors must receive reports and updates to discharge their responsibilities for IT governance and to see that the program is functioning well and providing business benefits.

> IT governance seeks to align business objectives with IT strategy to deliver business value.

The focus of governance in IT is on the actual software development and maintenance activities of the IT department or function, and IT governance efforts focus on making IT efficient and effective. That means minimizing costs by following proven software development methodologies and best practices, principles of data governance and information quality, and project management best practices while aligning IT efforts with the business objectives of the organization.

IT Governance Frameworks

Several IT governance frameworks can be used as a guide to implementing an IT governance program.

Although frameworks and guidance like COBIT® and ITIL have been widely adopted, there is no absolute standard IT governance framework; the combination that works best for your organization depends on business factors, corporate culture, IT maturity, and staffing capability. The level of implementation of these frameworks will also vary by organization.

COBIT 2019®

COBIT (Control Objectives for Information and Related Technology) is a process-based IT governance framework that represents a consensus of experts worldwide. It was codeveloped by the IT Governance Institute and ISACA. COBIT addresses business risks, control requirements, compliance, and technical issues.[23] The latest version is COBIT 2019.[24] Some changes and updates include:

- New concepts are introduced and terminology is explained—the COBIT Core Model and its 40 governance and management objectives provide the platform for establishing your governance program
- The performance management system is updated and allows the flexibility to use maturity measurements as well as capability measurements

- Introductions to design factors and focus areas offer additional practical guidance on flexible adoption of COBIT 2019, whether for specific projects or full implementation.[25]

COBIT offers IT controls that:

- Cut IT risks while gaining business value from IT under an umbrella of a globally accepted framework.
- Assist in meeting regulatory compliance requirements.
- Utilize a structured approach for improved reporting and management decision making.
- Provide solutions to control assessments and project implementations to improve IT and information asset control.

COBIT consists of detailed descriptions of processes required in IT and tools to measure progress toward maturity of the IT governance program. It is industry agnostic and can be applied across all vertical industry sectors, and it continues to be revised and refined.[26]

COBIT is broken into three basic organizational levels and their responsibilities: (1) board of directors and executive management; (2) IT and business management; and (3) line-level governance, security, and control knowledge workers.

The COBIT model draws on the traditional "plan, build, run, monitor" paradigm of traditional IT management, only with variations in semantics. There are four IT domains in the COBIT framework, which contain 40 governance and management objectives for IT processes and also control objectives that map to the four specific IT processes of:

1. Plan and organize
2. Acquire and implement
3. Deliver and support
4. Monitor and evaluate

Specific goals and metrics are assigned, and responsibilities and accountabilities are delineated.

The COBIT framework maps to ISO 17799 of the International Organization for Standardization and is compatible with Information Technology Infrastructure Library (ITIL) and other accepted practices in IT development and operations.

> COBIT 2019 is the latest version of the business framework for the governance of IT. It has been developed and refined over a 25-year span.

ValIT®

ValIT is a newer value-oriented framework that is compatible with and complementary to COBIT. Its principles and best practices focus is on leveraging IT investments to gain maximum value. Forty key ValIT essential management practices (analogous to COBIT's control objectives) support three main processes: value governance, portfolio management, and investment management. ValIT and COBIT "provide a full framework and supporting tool set" to help managers develop policies to manage business risks and deliver business value while addressing technical issues and meeting control objectives in a structured, methodic way."

> COBIT is process-oriented and has been widely adopted as an IT governance framework. ValIT is value-oriented and compatible and complementary with COBIT yet focuses on value delivery.

> COBIT is process oriented and has been widely adopted as an IT governance framework. ValIT is value oriented and compatible and complementary with COBIT yet focuses on value delivery.

ITIL

ITIL is a set of process-oriented best practices and guidance originally developed in the United Kingdom to standardize delivery of IT service management. ITIL is applicable to both the private and public sectors and is the "most widely accepted approach to IT service management in the world."[27] As with other IT governance frameworks, ITIL provides essential guidance for delivering business value through IT, and it "provides guidance to organizations on how to use IT as a tool to facilitate business change, transformation and growth."[28]

ITIL best practices form the foundation for ISO/IEC 20000 (previously BS 15000), the International Service Management Standard for organizational certification and compliance.[29] ITIL 2011 is the latest revision (as of this writing). It consists of five core published volumes that map the IT service cycle in a systematic way:

1. ITIL Service Strategy
2. ITIL Service Design
3. ITIL Service Transition
4. ITIL Service Operation
5. ITIL Continual Service Improvement

> ITIL is the "most widely accepted approach to IT service management in the world."

ISO 38500

ISO/IEC 38500:2015 is an international standard that provides high-level principles and guidance for senior executives and directors, and those advising them, for the effective and efficient use of IT.[30] Based primarily on AS 8015, the Australian IT governance standard, it "applies to the governance of management processes" performed at the IT service level, but the guidance assists executives in monitoring IT and ethically discharging their duties with respect to legal and regulatory compliance of IT activities.

The ISO 38500 standard comprises three main sections:

1. Scope, Application and Objectives
2. Framework for Good Corporate Governance of IT
3. Guidance for Corporate Governance of IT

It is largely derived from AS 8015, the guiding principles of which were:

- Establish responsibilities
- Plan to best support the organization
- Acquire validly
- Ensure performance when required
- Ensure conformance with rules
- Ensure respect for human factors

The standard also has relationships with other major ISO standards, and embraces the same methods and approaches.[31]

> ISO 38500 is an international standard that provides high-level principles and guidance for senior executives and directors responsible for IT governance.

IG Best Practices for Database Security and Compliance

Although security is a topic primarily for Chapter 11, it is a technical topic that we address here as well. Best

practices have been developed over the past few years and can prevent leakage of structured data from databases and Web services due to SQL injections (where hackers attack SQL databases) and other types of attacks.

An organization and its data need to be connected to its stakeholders—employees, customers, suppliers, and strategic partners. In this interconnected world that keeps expanding (e.g. cloud, mobile devices), proprietary data is exposed to a variety of threats. It is critical to protect the sensitive information assets that reside in your databases.[32]

Perimeter security often is easily penetrated. Web apps are vulnerable to attacks such as SQL injection (a favorite among malicious approaches). Hackers also can gain access by spear phishing (very specific phishing attacks that include personal information) to glean employee login credentials in order to get access to databases.

Streamlining your approach to database security by implementing a uniform set of policies and processes helps in compliance efforts and reduces costs. Here are some proven database security best practices:[33]

- *Inventory and document.* You must first identify where your sensitive data and databases reside in order to secure them. So a discovery and mapping process must take place. You can begin with staff interviews but also use tools such as **data loss prevention** (DLP) to map out data flows. Include all locations, including legacy applications, and intellectual property such as price lists, marketing and strategic plans, product designs, and the like. This inventorying/discovery process must be done on a regular basis with the assistance of automated tools, since the location of data can migrate and change.
- *Assess exposure/weaknesses.* Look for security holes, missing updates and patches, and any irregularities on a regular basis, using standard checklists such as the CIS Database Server Benchmarks and the DISA Security Technical Implementation Guides (STIGs). Do not forget to check OS-level parameters such as file privileges for database configuration files and database configuration options such as roles and permissions, or how many failed logins result in a locked account. (These types of database-specific checks are typically not performed by network vulnerability assessment scanners.)
- *Shore up the database.* Based on your evaluation of potential vulnerabilities, take proper steps and also be sure to that used database functions are disabled.
- *Monitor.* On a regular basis, monitor and document any configuration changes, and make sure the "gold" configuration is stable and unchanged. "Use change auditing tools that compare configuration snapshots and immediately alert whenever a change is made that affects your security posture."
- *Deploy monitoring/auditing tools.* Deploy these tools to immediately detect intrusions or suspicious activity, use your database's database activity monitoring (DAM) and database auditing tools continuously and in real time. Note any anomalies, such as unusually large numbers of records being downloaded even by authorized users—this could indicate, for instance, a rogue employee gathering information. But also higher-level "privileged users—such as database administrators (DBAs), developers, and outsourced personnel" must be monitored to comply with certain regulations. Watch for attackers who have gained access through authorized credentials. DAM creates an audit trail generated in real time that can be the forensic smoking gun in investigations after attacks have occurred. Also, monitor the application layer, as well-designed DAM solutions associate specific database transactions performed by the application with specific end-user IDs, in order to deterministically identify individuals violating corporate policies. In addition, combining database auditing information with OS [operating system] and network logs via a security information and event management ... system to see everything that a user has done can also provide critical information for forensic investigations.
- *Verify privileged access.* In your audit process, periodically review the list of privileged users and entitlement reports to ensure that superusers and those with access to sensitive information are still authorized.
- *Protect sensitive data.* Known sensitive data should be encrypted, so that even if attackers gain access, it is unreadable. "File-level encryption at the OS layer, combined with granular real-time monitoring and access control at the database layer, is typically accepted as a practical alternative to column-level encryption and a compensating control for Requirement 3.3 of PCI-DSS."
- *Deploy masking.* Hide your live production data by masking test data. "Masking is a key database security technology that de-identifies live production data, replacing it with realistic but fictional data

that can then be used for testing, training and development purposes, because it is contextually appropriate to the production data it has replaced."

- *Integrate and automate standardized security processes*. To pass compliance audits, you need to show that processes and system are in place to reduce risks and detect potential intrusions, attacks, and unauthorized use. Standardizing and automating these tasks as much as possible help minimize compliance costs while protecting the organization's data.

Implementing these best practices will help keep sensitive data in your databases secure.

> Identifying sensitive information in your databases and implementing database security best practices help reduce organizational risk and the cost of compliance.

Tying It All Together

Multiple frameworks and standards can be applied to the IT process to more effectively govern it and focus the processes on business impact. Beginning with a robust data governance program, organizations can ensure, at the more fundamental level, that the information they are using to base decisions on is clean, reliable, and accurate. Implementing an MDM program will help larger organizations with complex IT operations ensure that they are working with consistent data from a single source. Implementing the COBIT 5 business framework for delivering IT results will help support a more efficient IT operation and include other major frameworks, standards, and best practices. Leveraging the use of the ISO 38500 standard will help senior executives to better manage and govern IT operations, and employing database security best practices will help guard against outside threats.

CHAPTER SUMMARY: KEY POINTS

- Focusing on business impact and customizing your IG approach to meet business objectives are key best practices for IG in the IT department.
- Effective data governance can yield bottom-line benefits derived from new insights.
- Good data governance ensures that downstream negative effects of poor data are avoided and that subsequent reports, analyses, and conclusions are based on reliable, trusted data.
- Master data management is a key IG process in IT.
- IT governance seeks to align business objectives with IT strategy to deliver business value.
- COBIT 2019 is the latest version of the business framework for the governance of IT. It has been developed and refined over a 25-year span.
- CobiT is process oriented and has been widely adopted as an IT governance framework. ValIT is value oriented and compatible and complementary with CobiT yet focuses on value delivery.
- ValIT is a framework that focuses on delivering IT value.
- ITIL is the "most widely accepted approach to IT service management in the world."
- ISO 38500 is an international standard that provides high-level principles and guidance for senior executives and directors responsible for IT governance
- Identifying sensitive information in your databases and implementing database security best practices help reduce organizational risk and the cost of compliance.

Notes

1. "Gartner Says Master Data Management Is Critical to Achieving Effective Information Governance," January 19, 2012,www.gartner.com/newsroom/id/1898914.
2. IBM, "Selling Information Governance to Business Leaders," https://www.information-

management.com/news/selling-information-governance-to-business-leaders (accessed December 13, 2018).

3. Ibid.

4. Ibid.

5. Steven Adler, "Six Steps to Data Governance Success," May 31, 2007, https://www.cio.com/article/2438861/enterprise-architecture/six-steps-to-data-governance-success.html.

6. "New Trends and Best Practices for Data Governance Success," SeachDataManagement.com e-book, http://viewer.media.bitpipe.com/1216309501_94/1288990195_946/Talend_sDM_SO_32247_EBook_1104.pdf (accessed December 12, 2018).

7. Ibid.

8. "The DGI Data Governance Framework," DataGovernance.com, www.datagovernance.com/the-dgi-framework/ (accessed December 13, 2018).

9. "Information Management," BusinessDictionary.com, www.businessdictionary.com/definition/information-management.html (accessed December 13, 2018).

10. Sunil Soares, *Selling Information Governance to the Business* (Ketcham, ID: MC Press, 2011), 4.

11. Andrew White, "We Are Only Half Pregnant with MDM," April 17, 2013, https://blogs.gartner.com/andrew_white/2013/04/17/we-are-only-half-pregnant-with-master-data-management/.

12. Gartner IT Glossary, "Master Data Management," https://www.gartner.com/it-glossary/master-data-management-mdm (accessed December 13, 2018).

13. Bill Tolson, "Information Governance 101," May 21, 2013, https://informationgovernance101.com/2013/05/21/the-lifecycle-of-information/.

14. Gartner IT Glossary, "Information Lifecycle Management," www.gartner.com/it-glossary/information-life-cycle-management-ilm (accessed December 13, 2018).

15. Soares, *Selling Information Governance to the Business*.

16. "Data Architecture," BusinessDictionary.com, www.businessdictionary.com/definition/data-architecture.html (accessed December 13, 2018).

17. "Data Modeling," http://searchdatamanagement.techtarget.com/definition/data-modeling (accessed December 13, 2018).

18. Ibid.

19. Soares, *Selling Information Governance to the Business*.

20. Ibid.

21. Malcolm Chisholm, "Master Data Versus Reference Data," *Information Management*, April 1, 2006, https://www.information-management.com/news/master-data-versus-reference-data.

22. M. N. Kooper, R. Maes, and E.E.O. Roos Lindgreen, "On the Governance of Information: Introducing a New Concept of Governance to Support the Management of Information," *International Journal of Information Management* 31 (2011): 195–200, https://www.sciencedirect.com/science/article/pii/S0268401210000708.

23. Bryn Phillips, "IT Governance for CEOs and Members of the Board," 2012, p. 26.

24. www.isaca.org/cobit/pages/default.aspx (accessed December 12, 2018).

25. "COBIT 2019 Framework: Introduction and Methodology," www.isaca.org/COBIT/Pages/COBIT-2019-Framework-Introduction-and-Methodology.aspx (accessed December 12, 2018).

26. Phillips, “IT Governance for CEOs and Members of the Board.”

27. ITIL, “Welcome to the Official ITIL® Website,” https://www.axelos.com/best-practice-solutions/itil (accessed December 13, 2018).

28. ITIL, “What Is ITIL?” https://www.axelos.com/best-practice-solutions/itil/what-is-itil (accessed December 13, 2018).

29. Ibid.

30. ISO, “ISO/IEC 38500:2015: Governance of IT for the Organization,” https://www.iso.org/standard/62816.html (accessed December 12, 2018).

31. “ISO 38500 IT Governance Standard” (2008), www.38500.org/ (accessed December 13, 2018).

32. The following discussion and quotes are from Phil Neray, “Beating the Breach: 10 Best Practices for Database Security and Compliance,” November 3, 2011, https://datasafestorage.wordpress.com/2011/11/15/beating-the-breach-10-best-practices-for-database-security-and-compliance/.

33. Ibid.

CHAPTER 11
Information Governance and Privacy and Security Functions*

Privacy and security go hand in hand. Privacy cannot be protected without implementing proper security controls and technologies. Organizations must not only make reasonable efforts to protect privacy of data, but they must go much further as privacy breaches are damaging to customers and reputation. Potentially, they could put companies out of business.

Privacy and data protection awareness skyrocketed in 2018 with the implementation of the EU General Data Protection Regulation (GDPR), which gave new privacy rights to individuals in the EU and EU citizens everywhere, while creating significant new regulatory burdens on companies that handle personal data (PD), personally identifiable information (PII), and protected health information (PHI). Major corporations, after decades of automation, suddenly were being held to account for all instances and uses of personal consumer data. To do so, data maps and information flow diagrams had to be created to inventory all instances of stored personal data and learn how it flows through the organization.

This inventorying step is often one of the first in launching information governance (IG) programs, so the trend provided a significant increase in support for formal IG programs.

Information Privacy

By Andrew Ysasi

In a 2018 survey, Americans stated that they were more concerned with privacy than with healthcare or economic growth.[1] Privacy came of age in 2018, when the EU GDPR went into effect. Its impact was felt across the globe, as citizens became more aware of privacy concerns.

Information privacy refers to individuals or corporations controlling what others know about them. Unlike information security, privacy is not objective, but subjective. What one believes needs to be private can vary. (Some think that privacy is a moral or legal right, while others have argued that privacy isn't about controlling information about oneself.)[2] Privacy has been a widespread debate for the past century, and as the Internet has become engrained in societies and cultures, it will continue to be a concern for individuals and organizations.[3] In the United States, personally identifiable information (PII) is used to determine privacy attributes, for example, last name, home address, place of birth, and so forth. Often, PII is referred to as information that is not publicly available or in public works.[4]

Privacy came of age in 2018, when the EU GDPR went into effect.

In the digital age, individuals who are concerned about their privacy are often at the mercy of the corporations and developers of software to control how an individual's information is used. Individuals often use apps to work, manage finances, socialize, and play games on smartphones or tablets. The apps on portable devices often require permission from the user to use the information to operate at their full potential. As a result, individuals permit the organizations that manage the apps to store passwords, credit card information, bank accounts, digital cookies, fingerprints, and a myriad of personal information. If apps or software have privacy controls, an individual may choose to restrict what information is stored, how it is used, and who their information can be shared with. In extreme cases, individuals may opt to share personal information on social media sites with little regard for their privacy. Facebook, Twitter, Tinder, Foursquare, and LinkedIn are some examples of social media where individuals may reveal a great deal of personal information.[5]

Organizations should have an interest in privacy. Whether consumers or employees demand privacy or privacy is regulated in the countries they operate is something an organization should understand and have plans to address. Many organizations have a responsibility to their shareholders to be profitable or, if not an entity to gain profit, ensure that they are meeting their mission within the guidelines of their corporate structure. Privacy concerns can have a direct impact on an organization. Organizations that are more driven by profits may have fewer privacy concerns or controls than organizations that provide significant privacy protection.[6]

Further, organizations may operate in jurisdictions or industries where there are laws or rules around privacy.

For example, organizations should determine if they operate in an “opt-in” or “opt-out” jurisdiction, or if they are required to protect information because they operate in the healthcare or financial industries. Opt-in climates typically favor the privacy of the individual whereas opt-out favor an organization. Hospitals and insurance companies usually have laws and rules they need to follow to protect PII (personally identifiable information) or PHI (protected health information). Laws and regulations define what constitutes as PII and provide further guidance on how the information should be handled.

Criminals often target organizations to gain access to private information to sell or disrupt the organization. A database of privacy data breaches can be found at privacyrights.org. Privacyrights.org reports over 11.5 billion records have been breached from over 9000 breaches since 2005.[7] As data breaches become more prevalent, governments and privacy professionals have advocated for strict laws to protect individuals. Scholars believe that net neutrality, the Internet of Things (IoT), the human genome (medical), and cryptocurrencies will impact privacy for individual and organizations throughout the next decade.[8]

Generally Accepted Privacy Principles

The Generally Accepted Privacy Principles (GAPP) can be used to guide privacy programs. Please see Chapter 3 for more detail.

Privacy Policies

Privacy policies are ways for organizations to explain what they do with PII. Privacy policies may be found on websites or may be used internally only at an organization. Researchers predict that organizations will have tools available for individuals to choose how their information is used via privacy policies.[9] The International Association of Privacy Professionals (IAPP) has provided a template for organizations,[10] and it includes the following:

1. Why the policy exists—to comply with a law or protection from a data breach.
2. Data protection laws—specific examples of laws the organization is subject to follow.
3. Policy scope—who the policy applies to, employees, contractors, vendors, and individual.
4. Data protection risks—identifying to users what could happen if private information is provided.
5. Responsibilities—an explanation of what the organization is responsible to protect, who or whom is ultimately is responsible (e.g. board of directors, data protection officer, privacy officer, IT manager, marketing manager) and what will be done to protect information.
6. Guidelines for staff—not sharing information, using a strong password, not sharing credentials, not disclosing information unnecessarily.
7. Data storage—how information is stored and where it may be stored.
8. Data use—why information is needed for businesses and what is done to the information.
9. Data accuracy—that data collected is accurate, updated when wrong, and a way for individuals to report inaccurate information.
10. Subject access requests—how individuals can determine what is collected about them and how they can retrieve or edit the information.
11. Disclosure—the possibility of disclosing information to authorities or for legal reasons.
12. Providing information—clarification on how an individual's information is being processed and what their rights are.[11]

The IAPP has a privacy policy version that includes the European Union's GDPR provisions for organizations that must comply with the GDPR privacy requirements that can be found at the same site as the template outlined above.

> Privacy policies are ways for organizations to explain what they do with PII.

Privacy Notices

Privacy notices are typically exclusive to external facing stakeholders or to the public, where privacy policy

could be both.[12] The Better Business Bureau in the United States advises that a privacy notice should include five elements:

1. *Notice* (what personal information is being collected on the site or by the organization)
2. *Choice* (what options the customer has about how/whether personal data is collected and used)
3. *Access* (how a customer can see what data has been collected and change/correct it if necessary)
4. *Security* (state how any data that is collected is stored/protected)
5. *Redress* (what customer can do if the privacy policy is not met)

Privacy notices are typically exclusive to external facing stakeholders or to the public, where privacy policy could be both.

Fair Information Practices (FIPS)

HEW Report

"The first steps toward formally codifying Fair Information Practices began in July 1973, when an advisory committee of the US Department of Health, Education and Welfare proposed a set of information practices to address a lack of protection under the law at that time."[13]

As a result of this group, the HEW report was created formerly known as the Records, Computers and the Rights of Citizens. The HEW report summarized fair information practices as:

- There must be no personal data recordkeeping systems whose very existence is secret.
- There must be a way for an individual to find out what information about him is in a record and how it is used.
- There must be a way for an individual to prevent information about him obtained for one purpose from being used or made available for other purposes without his consent.
- There must be a way for an individual to correct or amend a record of identifiable information about him.
- Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data.[14]

OCED Privacy Principles

In 1980, the Organisation for Economic Co-operation and Development (OECD) published guidelines on the protection of privacy and personal data and were recently updated in 2013.[15] The eight fair information principles outlined by the OECD are as follows.

Collection Limitation Principle

There should be limits to the collection of personal data, and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

Data Quality Principle

Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete, and kept up to date.

Purpose Specification Principle

The purposes for which personal data are collected should be specified not later than at the time of data collection, and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.

Use Limitation Principle

Personal data should not be disclosed, made available, or otherwise used for purposes other than those

specified in accordance with Paragraph 9 except:

a. With the consent of the data subject; or

b. By the authority of law.

Security Safeguards Principle

Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification, or disclosure of data.

Openness Principle

There should be a general policy of openness about developments, practices, and policies concerning personal data. Means should be readily available for establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

Individual Participation Principle

An individual should have the right:

a. To obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;

b. To have communicated to him, data relating to him within a reasonable time;

   at a charge, if any, that is not excessive;

   in a reasonable manner; and

   in a form that is readily intelligible to him;

c. To be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and

d. To challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed, or amended.

Accountability Principle

A data controller should be accountable for complying with measures which give effect to the principles stated above.

Madrid Resolution 2009

In 2009, the Madrid Resolution brought 50 countries together to provide further guidance on information privacy. The resolution was designed and signed by executives from 10 international organizations, and one of the most important recommendations is governments "promoting better compliance with the applicable laws regarding data protection matters."[16]

In 2009, the Madrid Resolution brought 50 countries together to provide further guidance on information privacy.

While the Madrid Resolution was a big step in achieving global privacy awareness and guidance, it was merely viewed as a starting point for addressing the dynamic landscape of personal privacy across the globe. In 2010, the United States Department of Health Services Privacy Office stated an opinion on the Madrid Resolution, specifically criticizing the fact that governments and regulatory authorities were not included, and recommending that all stakeholders be involved in the development of a global privacy framework.[17] It is also worth noting the Director of the Spain Data Protection Agency (AEPD), Artemi Rallo: "These standards are a proposal of international minimums, which include a set of principles and rights that will allow the achievement of a greater degree of international consensus and that will serve as reference for those countries that do not have a legal and institutional structure for data protection."[18] It appeared in 2010 that a global privacy framework was on the horizon.

EU General Data Protection Regulation

The European Union's Data Protection Regulation (GDPR) provided the most significant privacy framework, policy, and regulatory overhaul history.

“In 2016, the EU adopted the General Data Protection Regulation (GDPR), one of its greatest achievements in recent years. It replaces the 1995 Data Protection Directive which was adopted at a time when the Internet was in its infancy.”[19]

Compliance for GDPR and enforceability requirements went into effect on May 28, 2018, and required organizations to seek consent from individuals before collecting their personal information.[20] Further, GDPR requires organizations to assign a Data Controller and provides specific principles related to the processing of personal data found in article 5:

Personal data shall be:

1. Processed lawfully, fairly and in a transparent manner in relation to the data subject (“lawfulness, fairness and transparency”);
2. Collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific, or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes (“purpose limitation”);
3. Adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed (“data minimization”);
4. Accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (“accuracy”);
5. Kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject (“storage limitation”);
6. Processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures (“integrity and confidentiality”).[21]

The May 2018 EU GDPR provided the most significant privacy framework, policy, and regulatory overhaul in history.

Failure to comply with GDPR requirements comes with hefty penalties. Fines are not assessed without an investigation and guidance. GDPR outlines the 10 criteria to determine the amount an organization can be fined:

1. *Nature of infringement:* The number of people affected, damage they suffered, duration of infringement, and purpose of processing.
2. *Intention:* Whether the infringement is intentional or negligent.
3. *Mitigation:* Actions taken to mitigate damage to data subjects.
4. *Preventative measures:* How much technical and organizational preparation the firm had previously implemented to prevent noncompliance.
5. *History:* Past relevant infringements, which may be interpreted to include infringements under the Data Protection Directive and not just the GDPR, and past administrative corrective actions under the GDPR, from warnings to bans on processing and fines.
6. *Cooperation:* How cooperative the firm has been with the supervisory authority to remedy the infringement.
7. *Data type:*: What types of data the infringement impacts.

8. *Notification:* Whether the infringement was proactively reported to the supervisory authority by the firm itself or a third party.
9. *Certification:* Whether the firm had qualified under approved certifications or adhered to approved codes of conduct.
10. *Other:* Other aggravating or mitigating factors may include the financial impact on the firm from the infringement.[22]

Lower-level fines can be up to €10 million, or 2% of the worldwide annual revenue of the prior fiscal year, whichever is higher. Higher-level fines can be up to €20 million, or 4% of the worldwide annual revenue of the prior fiscal year, whichever is higher.[23] The EU is quite evident through GDPR that consumer privacy is of the utmost importance, and organizations that choose to mishandle the data of EU citizens will suffer substantial financial losses. On the day that organizations were required to be compliant, Facebook and Google were hit with over $8 billion in lawsuits.[24] Facebook and other large companies seem to be making strides to comply, but other organizations are trying to block EU citizens from their sites to mitigate risk.[25] GDPR provides comprehensive oversight and guidance on how to protect consumer information and what happens if an organization does not comply. However, it is yet to be determined if GDPR will provide adequate privacy protection of consumer data without hindering EU citizens from experience foreign sites or companies wishing to expand their market in the EU.

GDPR: A Look at Its First Year

By Mark Driskill

The EU implemented sweeping new data privacy and protection laws meant to protect the personal data (PD) of those in the EU—importantly—be they citizens, temporary residents or visitors, from unauthorized use, *and*, extraterritorially, *wherever in the world their personal data is stored or used*.

The issues stem from the EU's broad definition of PD and the long history in Europe of privacy being viewed as a fundamental human right, against too much history of dictatorships and fascist control. The EU's General Data Protection Regulation (GDPR) took effect, provoking a new era of tech-company corporate accountability.

The GDPR didn't just standardize data privacy and protection across all (current) 28 member states of Europe, but refined both how to seek permission to use personal data and refresh the personal rights of each person in the EU to view and take control of their own personal data.

As 2018 came to a close, it was revealed that some major tech companies use personal data in ways that violate personal privacy in many ways.

Large data handlers like Facebook, Google, and Amazon have come under close examination by EU regulators, forcing CEOs in the "personal surveillance data business" to defend, and even rethink, their business models (e.g. Google then cited privacy regulation as a major threat to their business model in corporate documents). These have included both Privacy Regulators around GDPR (e.g. UK ICO, Ireland DPC, etc.) and EU competition regulators. Under the new GDPR these companies, without exception, must follow EU privacy law. The issues rest primarily with the advertising data insights these companies have created using proprietary algorithms. The invasiveness is secretive and at times unsettling, as these companies seem to know when someone will buy a pair of socks!

At first glance, it might seem as if the first year of GDPR compliance was largely uneventful, at least in terms of other leading global news stories. It's really a journey, as the EU regulators and analysts have shared. With almost 95,000 privacy complaints filed, they have only just started to process those investigations, findings, and enforcements. So many of the "privacy fines" we've seen since GDPR went live were really cases that occurred pre-GDPR and were thus much smaller in scope and penalties under the prior EU privacy regulation. What has been happening quietly, almost behind the scenes, is a tacit acceptance that data privacy from the person-centered perspective must begin with forcing larger companies such as Facebook, Google, and Amazon to comply. This hangs over companies in the consumer tech sector like thick fog. American businesses and culture do not like anyone telling them how to run things. Apparently, this is also true for GDPR compliance, adding to a persistent lack of full compliance.

A December 2018 Forrester survey commissioned by Microsoft found that more than half of businesses failed to meet GDPR compliance checkpoints.[26] Other highlights included:

- 57% instituted "privacy by design."

- 59% “collected evidence of having addressed GDPR compliance risks.”
- 57% “trained business personnel on GDPR requirements.”
- 62% “vetted third-party vendors.”

This last item is perhaps the most troubling: *38% have yet to vet their third-party software vendors*. This means that a significant portion of the global economy is not meeting GDPR compliance. The Forrester survey's primary findings were that only 11% of global companies are prepared to undergo the type of digital transformation needed to fully comply with GDPR-based privacy needs of citizens. In its entirety, GDPR has yet to make a significant impact, at least one beyond large tech company compliance.

A key implied issue that ultimately influences GDPR compliance checkpoints is the balance between intrusion into a company's business practices and its ability for profitmaking. Industry leaders note that in order to truly protect personal data, you must know exactly where and whose it is. This necessarily requires intrusion.

Enforcement and Precedent Setting

With the new GDPR mandate in place, EU member countries have a valuable tool for ensuring compliance even as these companies undertake actions to protect their business models. Ireland, for example, has “opened 10 statutory inquiries into Facebook and other Facebook-owned platforms in the first seven months” since GDPR was adopted in May 2018.[27]

The Irish Data Protection Commission (DPC) commissioner Helen Dixon notes that the inquiries match the public's interest in “understanding and controlling” their own personal data. The Irish DPC fully intends that these be precedent-setting. Given the widespread global use of Facebook and its plethora of connected apps, such inquiries from other EU member countries cannot be far behind.

In perhaps the most egregious case yet, a whistleblower forced Facebook to reveal that “as many as 600 million users’ passwords were stored in plain text and accessible to 20,000 employees, of which 2000 made more than 9 million searches that accessed the passwords going back to 2012.”[28] Added to this blatant breach of basic cybersecurity practices is the fact that Facebook knew about the issue back in January and spent several months trying to keep it from the public.[29] They would surely have been embarrassing questions to answer during the recent US Congressional hearings.

As *Forbes* points out, cybersecurity at Facebook just might be obsolete. In the wake of the sensational stories regarding recent Russian interference into American elections, “Facebook did not conduct a top-down security audit of its authentication systems.” This is a profound, if not provocative, revelation, particularly given Mark Zuckerberg's promise to reform Facebook's business practices.

That promise, made to Congress just prior to GDPR's May 2018 rollout, seems now to be empty. While Zuckerberg testified, his company continued its intrusive practices, even as he tried to simplify for legislators Facebook's business practices.

In the business world, laws and regulations are street signs to setting precedent. During this initial phase of GDPR compliance, it is crucial that leading EU countries, such as Germany, take positions of authority. Germany's Federal Cartel Office, the federal agency that regulates Germany's competition laws, set a new precedent in a February 2019 court ruling. In an anti-competition class-action case, the German court severely limited Facebook's ability to collect user data inside Germany. This essentially walls off Germany's Facebook users from the rest of Facebook's user base. The precedent set by German regulators was substantial. Facebook (at least in Germany) can longer use tactics such as using user data to make fictitious profiles. Moreover, it can no longer use Facebook Pixel, a single character imbedded in a page that transmits data back to the company's servers. With the German precedent, Facebook can no longer claim that what it does with user data on its platform is proprietary.

In some ways, the first year of “GDPR-live” was marked by both confusion and denial that such regulation was really needed. Today, the establishment of a nation-specific precedent is the exception, not the rule. However, enough cannot be said about the fact that Germany is one of the main economic powers of the globe. Without German leadership, GDPR might die an unceremonious death. The same must happen in other countries involved in setting global economic policy.

In short, GDPR-style privacy must come to the United States. Thankfully, California is leading the way with its California Consumer Privacy Act (CCPA), which went live in January 2020.

Privacy Programs

Privacy programs are often required to manage the national and international requirements to protect consumer information. The IAPP has a rigorous certification called the Certified Information Privacy Manager (CIPM), and the outline[30] for the exam provides a framework to build a privacy program.

- Vision—The purpose of the privacy program and what the program will achieve.
- Team—Key stakeholders who have direct responsibility for privacy matters that may include a Chief Privacy Officer (CPO), Data Protection Officer (DPO), Data Controller (DC), executive sponsorship, inside or outside legal counsel, technology leadership, compliance leadership, and records and information leadership.
- Policies—Privacy policies for websites, social media, e-mail, mobile apps, internal practices, and privacy policies addressing applicable international, national, and local laws should be reviewed and examined.
- Activities—Examples of activities are training, awareness, updating agreements with key stakeholders, addressing cross-border concerns, reviewing insurance and assurance options, evaluating privacy compliance software tools, and conducting privacy impact assessments along with specific risk assessments that relate to privacy concerns or regulations.
- Metrics—Collection, responses time to inquiries, retention, PIA metrics, maturity levels, and resource utilization are some examples of how organizations can measure privacy.

Privacy in the United States

FCRA

Privacy in the United States has a fairly recent history. In 1970, the Fair Credit Reporting Act (FCRA) was placed into law to protect consumers. With the FCRA, consumers were able to correct errors on their credit report.[31] FCRA was amended in 1996 and again in 2003 by the Fair and Accurate Credit Transaction Act that addresses issues related to identity theft. Under FCRA there are obligations users of credit reports must follow, which include:

1. Users must have a permissible purpose—as in the uses of credit reports are limited.
2. Users must provide certifications—they have a right to request the report.
3. Users must notify consumers when adverse actions are taken—users must be notified if their credit was a result of something not occurring.

The Federal Trade Commission (FTC) provided updates to the FCRA on their website located at www.ftc.gov.[32] Further, the FTC is the enforcement arm of FCRA and FACTA.

HIPAA

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 was created to improve the delivery of healthcare services and to provide standards on how patient records are handled in data exchanges. HIPAA oversight is provided by the Department of Health and Human Services (DHHS) under the Office for Civil Rights (OCR).[33] Entities covered under HIPAA are:

1. Health care providers
2. Health plans
3. Health care clearinghouses

The entities were broken up into two Covered Entities (CE) and Business Associates (BA) and personal information they managed is referred to as Protected Health Information (PHI).

Protected health information is any individually identifiable health information transmitted or maintained in any form or medium, which is held by a covered entity or its business associate, identifies an individual or offers a reasonable basis for identification, is created or received by a covered entity or an employer, and relates to a past, present or future physical or mental condition, provision of health care or payment for health care to that individual.[34]

HIPAA was enacted in 1996 to improve the delivery of healthcare services and to provide standards on how patient records are handled in data exchanges.

In 2000, HIPAA was amended to include the "Privacy Rule" that required:

1. Privacy Notices
2. Authorization of Uses and Disclosures
3. Access and Accountings of Disclosures
4. "Minimum Necessary" Use or Disclosure
5. De-identification
6. Safeguards
7. Business Associates accountability
8. Exceptions[35]

In 2003, HIPAA was amended further with the *Security Rule*, which focused on the protection of electronic medical records. It required covered entities to:

1. Ensure confidentiality of electronic medical records
2. Protect against any threats or hazards to the security or integrity of records
3. Protect against any reasonably anticipated users or disclosures that are not permitted
4. Ensure compliance with the Security Rule by the staff
5. Identify an individual responsible for implementation and oversight of the Security Rule and compliance program
6. Conduct an initial and ongoing risk assessments
7. Implement a security awareness program
8. Incorporate Security Rule requirements into Business Associate Contracts required by the Privacy Rule

In 2000, HIPAA was amended to include the "Privacy Rule"; in 2003, the "Security Rule" was added that focused on the protection of EMRs.

In 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act was enacted as part of the American Recovery and Reinvestment Act of 2009 to promote the adoption and meaningful use of health information technology.[36] Further, HITECH separated out four categories of how fines would be distributed, and another noteworthy change was the correction of a violation within 30 days.[37] Below is the penalty structure for HIPAA violations:

- *Category 1:* A violation that the covered entity was unaware of and could not have realistically avoided, had a reasonable amount of care had been taken to abide by HIPAA Rules.
- *Category 2:* A violation that the covered entity should have been aware of but could not have avoided even with a reasonable amount of care (but falling short of willful neglect of HIPAA Rules).
- *Category 3:* A violation suffered as a direct result of "willful neglect" of HIPAA Rules, in cases where an attempt has been made to correct the violation.
- *Category 4:* A violation of HIPAA Rules constituting willful neglect, where no attempt has been made to correct the violation.[38]

The HITECH Act was enacted as part of the American Recovery and Reinvestment Act of 2009 to promote the adoption of health information technology.

Other US Regulations

The United States has scores of other regulations that have shaped the US privacy landscape:

The Graham-Leach-Bliley Act (GLBA) issued in 1999 prohibited the sale of detailed customer information to, for example, telemarketing firms. GLBA is enforced by the FTC and allows consumers to "opt out" of having their information shared with affiliates. Like HIPAA, GLBA has safeguards and requires administrative, technical, and physical security of consumer information as well as privacy notice requirements.

The Children's Online Privacy Protection Act of 2000 (COPPA) was the result of the FTC's *1998 Privacy Online: A Report to Congress*.[39] Privacy related to children in the United States was the primary focus of COPPA, and required website companies and other operaters to adhere to a set of requirements, including privacy notices, or be subject to fines.

Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003 (CAN-SPAM) applies to anyone who advertises products and services by e-mail generated in the United States.[40] CAN-SPAM's aim is to reduce deceptive advertising by e-mail, include warnings for sexually explicit content, and the ability to opt-out from future e-mails. CAN-SPAM is governed by the FTC.

Certain states in the United States have taken action to protect the privacy of citizens.

Massachusetts rolled out 940 CMR 27.00: Safeguard of personal information in 2010[41] and the California Consumer Privacy Act in 2018, which will be enforceable in 2020,[42] focus on privacy and security of citizen information as a result of ongoing data breaches. What information is collected, specifically by social media entities Facebook and Google, why it is collected, and obtaining personal information in usable formats are some highlights of the California law. The law also requires children under 16 to opt in to allowing companies to collect their information.[43] Further, it is thought that future privacy laws from other states may modify their privacy laws to mirror California.

California Consumer Privacy Act

California's new privacy law went into effect on January 1, 2020. This act is designed to give California residents *a better way to control and protect their personal information*. California consumers will have the right to order companies to delete their personal data—similar to what Europe's all-encompassing GDPR regulation calls for (but not as strict). Many US states have begun debating new privacy laws using the CCPA and GDPR as models to protect the personal rights of individuals and consumers.[44]

Privacy regulations are rapidly spreading worldwide in countries such as India, Brazil, and Australia. Even the US Congress has been working on a bill that could soon become federal law.

California consumers will have the legal right to force companies to not only delete their personal information but also disclose what personally identifiable information (PII) has been collected about them, demand the reasons for collecting it, and order the companies to refrain from selling any of it. The personal information protected in these regulations contains a lot more than just financial or banking data; PII includes all "information that identifies, relates to, describes, is associated with, or could be reasonably linked, directly or indirectly, to a consumer or household." This consists of many different types of information, including IP addresses, biometric data, personal characteristics, browsing history, geolocation data, and much more.

On June 28, 2018, the California Congress passed Assembly Bill 375, the CCPA. The act will apply to any "for-profit" organization that grosses at least $25 million annually and interacts with 50,000 or more Californians, or derives at least half of its annual revenue from selling personal information. Most importantly, the CCPA applies to businesses "regardless of location" that meet the above criteria. You must comply if you process personal information of Californians whether your corporation is located in California or not.

What was interesting is how the CCPA was rushed into law and signed by Governor Jerry Brown in June of 2018, just days before a deadline to withdraw a state ballot measure on a privacy proposition coming up in the November election. Tech companies like Google and Facebook were ready to fight against this voter initiative because it would have been more strict—holding them more accountable with more far-reaching rules and heavier fines. These same tech giants are currently lobbying congress in Washington, DC, to create new federal privacy laws. Not surprisingly, big tech companies are only looking out for themselves as they try to preserve their "surveillance" business model by watering down impending privacy legislation.

It is important to note that the CCPA has already been amended and politicians promise to make more changes before the CCPA goes into full effect in January 2020.

Privacy in Asia

The Asia-Pacific region is no stranger to privacy concerns. The Asia-Pacific Economic Cooperation (APEC) endorsed the 1998 Blueprint for Action on Electronic Commerce that discussed the creation of a fair and open digital economy that promoted confidence and trust with consumers. In 2005, APEC published their privacy principles that included:

- Preventing harm
- Notice
- Collection limitations
- Uses of personal information choice
- Integrity of personal information
- Security safeguards
- Access and correction
- Accountability[45]

In 2015, APEC added the principle of *Choice*, which emphasized that individuals should have a choice in how their information is used.[46] The APEC privacy framework also provides implementation guidelines to guide organizations of all types on how to implement the privacy framework.

Infonomics and Privacy

Doug Laney explored a concept known as infonomics in his seminal 2018 book that discussed how organizations could monetize, manage, and measure information as an asset.[47] While the use of internal or consumer data raises privacy concerns, Laney suggests viewing data as an asset and working within the privacy and legal parameters to maximize the value of information. Sometimes this may involve spinning off a new legal entity to provide some distance between the parent company and a data monetization venture, for legal and branding purposes. Laney's book discusses how information is used to generate new revenue for marketing purposes, providing access to third parties and streamlining operations. While these concepts may go against general privacy principles, it is possible to navigate privacy concerns with Laney's ideas. He provides “Seven Steps to Monetizing Your Information Assets,” and goes on to explain how to report your information assets on your balance sheets, and provides formulas. While doing so, it is essential to understand the impact of privacy risks associated with monetizing information.

Privacy Laws

The protection of personally identifiable information (PII) is a core focus of IG efforts. PII is any information that can identify an individual, such as name, Social Security number, medical record number, credit card number, and so on. Various privacy laws have been enacted in an effort to protect privacy. You must consult your legal counsel to determine which laws and regulations apply to your organization and its data and documents.

The protection of personally identifiable information (PII) is a core focus of IG efforts.

The CCPA is waking up other US states to the need for more robust privacy legislation, while at the same time there is a move underfoot for the first national privacy legislation. If this goes forward, then there will be legal battles regarding preemption of the federal law in states having stricter privacy laws, using a “states’ rights” argument, and the federal government will argue that they have preemptive legal authority.

The Federal Wiretap Act “prohibits the unauthorized interception and disclosure of wire, oral, or electronic communications.” The Electronic Communications Privacy Act (ECPA) of 1986 amended the Federal Wiretap Act significantly and included specific on e-mail privacy.[48] The Stored Communications and Transactional Records Act (SCTRA) was created as a part of ECPA and is “sometimes useful for protecting the privacy of e-mail and other Internet communications when discovery is sought.” The Computer Fraud and Abuse Act makes it a crime to intentionally breach a “protected computer” (one used by a financial institution or for interstate commerce).

In the United States, the California Consumer Privacy Act (CCPA) is to go into effect January 1, 2020. The CCPA is waking up other US states to the need for more robust privacy legislation.

Also relevant for public entities is the Freedom of Information Act, which allows US citizens to request government documents that have not previously been released, although sometimes sensitive information is redacted (blacked out) and specifies the steps for disclosure as well as the exemptions. In the United Kingdom, the Freedom of Information Act 2000 provides for similar disclosure requirements and mandatory steps.

In the United Kingdom, privacy laws and regulations include the following:

- Data Protection Act 1998
- Freedom of Information Act 2000
- Public Records Act 1958
- Common law duty of confidentiality
- Confidentiality National Health Service (NHS) Code of Practice
- NHS Care Record Guarantee for England
- Social Care Record Guarantee for England
- Information Security NHS Code of Practice
- Records Management NHS Code of Practice

Also, the international information security standard ISO/IEC 27002: 2005 comes into play when implementing security.

Cybersecurity

Breaches are increasingly being carried out by malicious attacks, but a significant source of breaches is internal mistakes caused by poor information governance (IG) practices, software bugs, and carelessness. The average cost of a data breach in 2018 was nearly $4 million,[49] and some spectacular breaches have occurred, such as the 87 million–plus Facebook accounts and 37 million Panera accounts that were hacked in 2018,[50] but perhaps the most colossal was the breach of over 500 million customer records, including credit card and passport numbers, suffered by the Marriott Hotel chain that same year.[51] Millions of breaches occur each year: There were an estimated 179 million privacy breaches in the United States in 2017 alone.[52]

The average cost of a data breach in 2013 was over $5 million.

Cyberattacks Proliferate

Online attacks and snooping continue at an increasing rate. Organizations must be vigilant about securing their internal, confidential documents and e-mail messages. In one assessment, security experts at Intel/McAfee "discovered an unprecedented series of cyberattacks on the networks of 72 organizations globally, including the United Nations, governments and corporations, over a five-year period."[53] Dmitri Alperovitch of McAfee described the incident as "the biggest transfer of wealth in terms of intellectual property in history."[54] The level of intrusion is ominous.

The targeted victims included governments, including the United States, Canada, India, and others; corporations, including high-tech companies and defense contractors; the International Olympic Committee; and the United Nations. "In the case of the United Nations, the hackers broke into the computer system of its secretariat in Geneva, hid there for nearly two years, and quietly combed through reams of secret data, according to McAfee."[55] *Attacks can be occurring in organizations for years before they are uncovered—if they are discovered at all.* This means that an organization may be covertly monitored by criminals or competitors for extended periods of time.

And they are not the only ones spying—look no further than the US National Security Agency (NSA) scandal of 2013. With Edward Snowden's revelations, it is clear that governments are accessing, monitoring, and storing

massive amounts of private data.

> Attacks can continue in organizations for years before they are uncovered—if they are discovered at all.

Where this stolen information is going and how it will be used is yet to be determined. But it is clear that possessing this competitive intelligence could give a government or company a huge advantage economically, competitively, diplomatically, and militarily.

The information assets of companies and government agencies are at risk globally. Some are invaded and eroded daily, without detection. The victims are losing economic advantage and national secrets to unscrupulous rivals, so it is imperative that IG policies are formed, followed, enforced, tested, and audited. It is also imperative to use the best available technology to counter or avoid such attacks.[56]

> Information assets are invaded and eroded daily, often without detection. This compromises competitive position and has real financial impact.

Insider Threat: Malicious or Not

Ibas, a global supplier of data recovery and computer forensics, conducted a survey of 400 business professionals about their attitudes toward intellectual property (IP) theft:

- Nearly 70% of employees have engaged in IP theft, taking corporate property upon (voluntary or involuntary) termination.
- Almost one-third have taken valuable customer contact information, databases, or other client data.
- Most employees send e-documents to their personal e-mail accounts when pilfering the information.
- Almost 60% of surveyed employees believe such actions are acceptable.
- Those who steal IP often feel that they are entitled to partial ownership rights, especially if they had a hand in creating the files.[57]

These survey statistics are alarming, and by all accounts the trend continuing to worsen today. Clearly, organizations have serious cultural challenges to combat prevailing attitudes toward IP theft. A strong and continuous program of IG aimed at securing confidential and sensitive information assets can educate employees, raise their IP security awareness, and train them on techniques to help secure valuable IP. And the change needs to be driven from the top: from the CEO and boardroom. However, the magnitude of the problem in any organization cannot be accurately known or measured. Without the necessary IG monitoring and enforcement tools, executives cannot know the extent of the erosion of information assets and the real cost in cash and intangible terms over the long term.

Countering the Insider Threat

Frequently ignored, the insider has increasingly become the main threat—more than the external threats outside of the perimeter. *Insider threat breaches can be more costly than outsider breaches*. Most of the insider incidents go unnoticed or unreported.[58]

> Security professionals state that insider threat breaches are often more costly than outsider ones.

Companies have been spending a lot of time and effort protecting their perimeters from outside attacks. In recent years, most companies have realized that the insider threat is something that needs to be taken more seriously.

Malicious Insider

Malicious insiders and saboteurs comprise a very small minority of employees. A disgruntled employee or sometimes an outright spy can cause a lot of damage. Malicious insiders have many methods at their disposal to harm the organization by destroying equipment, gaining unsanctioned access to IP, or removing sensitive

information by USB drive, e-mail, or other methods.

Nonmalicious Insider

Fifty-eight percent of Wall Street workers say they would take data from their company if they were terminated, and believe they could get away with it, according to a recent survey by security firm CyberArk.[59] Frequently, they do this without malice. The majority of users indicated having sent out documents *accidentally* via e-mail. So, clearly it is easy to leak documents without meaning to do any harm, and that is the cause of most leaks.

Solutions

Trust and regulation are not enough. In the case of a nonmalicious user, companies should invest in security, risk education, and **Security Awareness Training** (SAT). A solid IG program can reduce IP leaks through education, training, monitoring, and enforcement. SAT raises user awareness and can be gamified to increase engagement and effectiveness. Newer SAT programs utilize animated cartoon-like videos to keep users interested and engaged.

In the case of the malicious user, companies need to take a hard look and see whether they have any effective IG enforcement and **document life cycle security** (DLS) technology such as information rights management (IRM) in place. Most often, the answer is no.[60]

Information Security Assessments and Awareness Training

By Baird Brueseke

Employees' human errors are the weakest link in securing an organization's confidential information. However, there are some small, inexpensive steps (through employee training) that can reduce information risk.

Security Awareness Training (SAT) programs educate an organization's workforce about the risks to information and potential schemes employed by hackers. SAT provides them with the skills to act consistently in a way that protects the organization's information assets. Bad actors target an employee's natural human tendencies with phishing e-mails and spear-phishing campaigns. SAT training programs often include phishing simulation and other social-engineering tactics such as text message "**smishing**" and unattended USB drives. SAT products provide a comprehensive approach to employee training, which empowers them to recognize and avoid a broad range of threat vectors.

SAT is an easy effective and easy way to reduce risk. Corporate risk is reduced by changing the (human) behavior of employees. Leading products in this market use innovative methods such as short, animated videos and pop quizzes to teach employees about information security threats.

SAT is not a one-and-done activity. In order to be effective, SAT must be implemented as an ongoing process. Workplace safety programs implemented to meet OSHA requirements serve as a good metaphor. SAT is a continuous improvement process; new threats emerge every day. The leading products incorporate new content on a regular basis and provide employee engagement opportunities that go well beyond the traditional computer-based training activities.

SAT Is a Quick Win for IG Programs

One of the quick—and low cost—wins that an Information Governance (IG) program can bring to an organization is the implementation of an SAT program. IG programs are implemented to reduce risk and maximize information value. Security Awareness Training programs are an excellent way to reduce risk and they are easy to implement. Employees have many bad habits that can leave a company vulnerable to data breach scenarios.

In response to the ever-increasing cyber security threat faced by business, a new subsegment of the Information Security market has emerged and matured in the last five years. The Security Awareness Training market grew 54% from 2015 to 2017. Projected revenues for 2018 were $400 million and growth was strong.

Cybersecurity threats are constantly evolving. One of the important things to understand when evaluating Security Awareness Training programs is the vendor's cycle for new content development and deployment in the training platform. Some of the features to look for and evaluate when selecting a SAT product are:

- Interactive content in varied formats designed to keep learners engaged
- Training designed to teach resistance to multiple forms of social engineering

- Optimization for smartphone and tablet usage
- Gamification and other methods to engage employees and increase participation
- Prestructured campaigns for different types/levels of employees
- Role-based training with optional customization based on corporate environment
- Robust library of existing content and flexible micro-learning topics
- Internal marketing and communication tools for use by the HR department
- Short lessons, approximately 5–10 minutes in length
- Integrated quizzes and metrics to track employee participation and retention
- Integration with corporate LMS
- Integration with end-point security systems

It is important to understand that SAT products typically include not only training, but also simulated attacks. Therefore, the way in which the SAT product interacts with existing cybersecurity defenses is a serious consideration. For example, if the training program administrator sends out a simulated phishing attack e-mail, that e-mail needs to make it through the SPAM filter and into the employee's e-mail inbox before the employee can be tempted into potentially clicking on the bad link.

In smaller companies, it may be sufficient to whitelist to the domain from which the phishing e-mail is being sent. In larger organizations that have Security Information Event Monitoring (SIEM) and other automated cyber defense systems, the company's IT/Security Team would likely request integration of a notification process for the simulated attack campaign in order to avoid a rash of false alarms from the security monitoring systems.

Security Awareness Training can provide a quick win for IG programs. The training immediately reduces risk. At the same time, management can point to the employee participation metrics as proof that proactive efforts are being made to enhance the organizations' security posture.

Cybersecurity Assessments

In today's cyber threat landscape, companies have a fiduciary duty to assess their cyber security posture. This is the root function of a Cybersecurity Assessment. Typically, third-party vendors are contracted to perform the Assessment. These firms have expertise in a variety of cybersecurity skills that they use to tailor the engagement to a scope appropriate for the organization being assessed.

One of the first steps when starting a Cybersecurity Assessment project is to select a framework. This choice will become part of the project requirements and in large part define the scope of work to be performed by the third-party vendor. There are several frameworks to choose from including: ISO 27001, COBIT 2019, NIST Cybersecurity Framework, NIST 800-53, DOD 8570, DCID 6/3, HITRUST CSF, and the Cloud Security Alliance's Cloud Controls Matrix. Even the Motion Picture Association of America has defined a cybersecurity framework to protect their member's intellectual property.

The NIST Cybersecurity Framework consists of five "functions," which are: Identify, Protect, Detect, Respond, and Recover, as shown in the following.

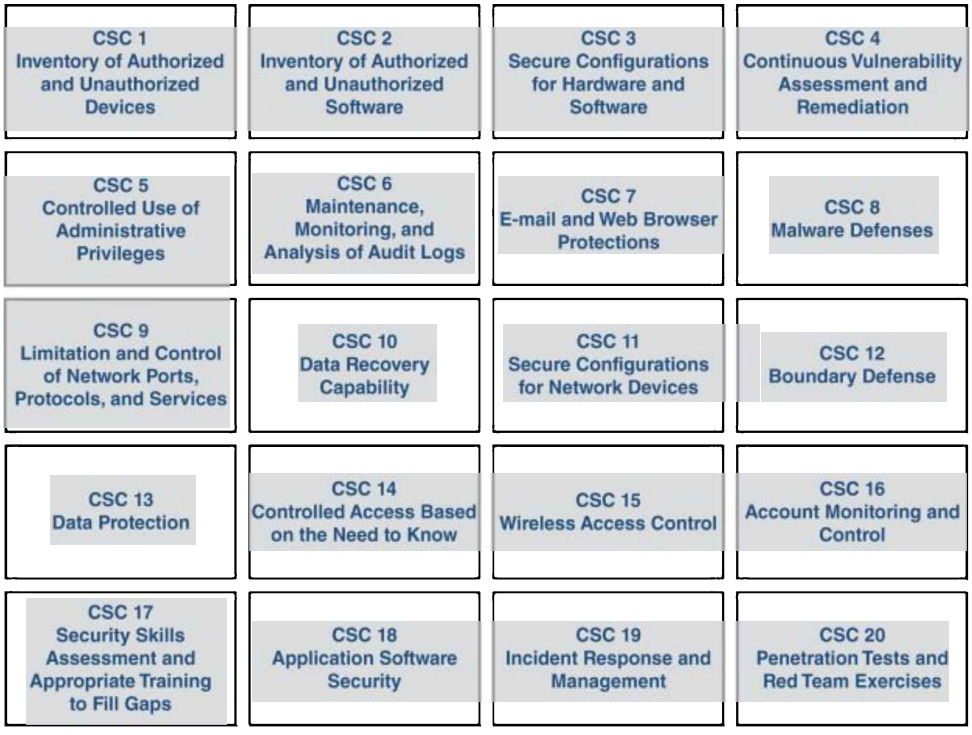
One of the first steps when starting a Cybersecurity Assessment project is to select a framework.



These five functions are subdivided into 22 categories—and then each category has multiple controls. One issue with the NIST framework is that a comprehensive Security Assessment using this framework can quickly become a big project, often too big for the organization's size.

For small and medium-sized business, a good step forward is to specify the Center for Internet Security (CIS) Top 20 controls as the framework the independent Cyber Security team will assess. The CIS Top 20 controls provide an easy to understand assessment tool that senior executives will understand.

For small and medium-sized business, a good step forward is to specify the Center for Internet Security (CIS) Top 20 controls.



Once the CIS controls are evaluated, the organization's security posture can be easily visualized using color coded infographics and risk score heat charts. Many Security Assessments include an evaluation of the business's people, process, and technology. There is no point in spending technology dollars if the existing corporate processes do not support their use. These decisions can be explored using Radar charts to visualize the cyber readiness of three metrics: people, process, and technology. Radar charts depict cybersecurity assessment scores in a circular chart with gradient ranking that shows executives the information they need to act on to enhance their security posture.

The term **vulnerability assessment** applies to a broad range of systems. For example, in the context of a disaster recovery plan, the vulnerability assessment would include the likelihood of flooding, earthquakes, and other potential disasters. In the digital sphere, a vulnerability assessment is an evaluation of an organization's cybersecurity weaknesses. This process includes identifying and prioritizing specific computer configuration issues that represent vulnerable aspects of an organization's computing platforms.

The Institute for Security and Open Methodologies (ISECOM; www.isecom.org/research/) publishes the Open-Source Security Testing Methodology Manual that documents the components of a vendor neutral approach to a wide range of assessment methods and techniques. A vulnerability assessment project typically includes the following:

1. Inventory of computing assets and networked devices
2. Ranking those resources in order of importance
3. Identification of vulnerabilities and potential threats
4. Risk assessment
5. Prioritized remediation plan

A vulnerability assessment starts with an inventory of computer systems and other devices connected to the network. Once the items on the network have been enumerated, the network is scanned using an automated tool to look for vulnerabilities. There are two types of scans: credentialed and noncredentialed. A credentialed scan uses domain admin credentials to obtain detailed inventories of software applications on each of the computers. This method provides the security team with the information necessary to identify operating system versions and required patches.

Often, a company's website is an overlooked corporate asset for a vulnerability assessments. The Open Web Application Security Project (OWASP) maintains a list of the top-10 vulnerabilities most commonly found on websites. Surprisingly, many websites fail to properly implement user authentication and data input checking.

These types of vulnerabilities have the potential to expose corporate data to anyone with Internet access. Performing a vulnerability assessment exposes these issues so they may be resolved.

The final output of a vulnerability assessment project is the prioritized remediation plan. This plan uses the results of the risk assessment to determine which vulnerabilities represent the greatest risk to the organization. The total list of vulnerabilities is often numbered in the hundreds, if not thousands. However, not all of the vulnerabilities are big problems requiring immediate attention. The prioritized remediation plan allows IT administrators to reduce corporate risk quickly by focusing on the most important weaknesses first.

InfoSec Penetration Testing

Penetration testing ("pen test") is a technique used by information security (InfoSec) professionals to find weaknesses in an organization's InfoSec defenses. In a penetration test, authorized cybersecurity professionals play the hacker's role.

Penetration testing attempts to circumvent digital safeguards and involves the simulation of an attack by hackers or an internal bad actor. The same techniques used by hackers to attack companies every day are used. The results of a penetration test reveal (in advance) the vulnerabilities and weaknesses that could allow a malicious attacker to gain access to a company's systems and data.

Some techniques used include brute-force attacks, exploitation of unpatched systems, and password-cracking tools. Organizations hire InfoSec experts with specialized training credentials––such as Certified Ethical Hacker (CEH) and Offensive Security Certified Profession (OSCP)––to conduct authorized attempts to breach the organization's security safeguards. These experts begin the pen test by conducting reconnaissance, often creating an attack surface and Internet footprint analysis to passively identify exposures, risks, and gaps in security. Once potential vulnerabilities are identified, the penetration testing team initiates the exploit attempts using automated tools to probe websites, firewalls, and e-mail systems.

Penetration testing attempts to circumvent digital safeguards and involves the simulation of an attack by hackers or an internal bad actor.

Successful exploits often involve multiple vulnerabilities, which are attacked over several days. Individually, none of the weaknesses are a wide-open door. However, when combined together by an expert penetration tester, the result is a snowball effect that provides the pen test expert with an initial foothold inside the network from which they can pivot and gain access to additional systems.

Penetration testing is a useful technique for evaluating the potential damage from a determined attacker, as well as assess the organizational risks posed. Most hackers and criminals go after low-hanging fruit––easy targets. Regular penetration tests ensure that the efforts required to gain access to internal networks are substantial. The result? Most hackers will give up after a few hours and move on to other targets that are not so well defended.

Cybersecurity Considerations and Approaches

By Robert Smallwood

Limitations of Perimeter Security

Traditionally, central computer system security has been primarily perimeter security—securing the firewalls and perimeters within which e-documents are stored and attempting to keep intruders out—rather than securing e-documents directly upon their creation. *The basic access security mechanisms implemented, such as passwords, two-factor authentication, and identity verification, are rendered totally ineffective once the confidential e-documents or records are legitimately accessed by an authorized employee.* The documents are usually bare and unsecured. This poses tremendous challenges if the employee is suddenly terminated, if the person is a rogue intent on doing harm, or if outside hackers are able to penetrate the secured perimeter. And, of course, it is common knowledge that they do it all the time. *The focus should be on securing the documents themselves, directly.*

Restricting access is the goal of conventional perimeter security, but it does not directly protect the information inside. Perimeter security protects information the same way a safe protects valuables; if safecrackers get in, the contents are theirs. There are no protections once the safe is opened. Similarly, if hackers penetrate the

perimeter security, they have complete access to the information inside, which they can steal, alter, or misuse.[61] The perimeter security approach has four fundamental limitations:

1. *Limited effectiveness*. Perimeter protection stops dead at the firewall, even though sensitive information is sent past it and circulates around the Web, unsecured. Today's extended computing model and the trend toward global business means that business enterprises and government agencies frequently share sensitive information externally with other stakeholders, including business partners, customers, suppliers, and constituents.
2. *Haphazard protections*. In the normal course of business, knowledge workers send, work on, and store copies of the same information outside the organization's established perimeter. Even if the information's new digital environment is secured by other perimeters, each one utilizes different access controls or sometimes no access control at all (e.g. copying a price list from a sales folder to a marketing folder; an attorney copying a case brief or litigation strategy document from a paralegal's case folder).
3. *Too complex*. With this multiperimeter scenario, there are simply too many perimeters to manage, and often they are out of the organization's direct control.
4. *No direct protections*. Attempts to create boundaries or portals protected by perimeter security, within which stakeholders (partners, suppliers, shareholders, or customers) can share information, cause more complexity and administrative overhead while they fail to protect the e-documents and data directly.[62]

Despite the current investment in e-document security, it is astounding that once information is shared today, it is largely unknown who will be accessing it tomorrow.

Defense in Depth

Defense in depth is an approach that uses multiple layers of security mechanisms to protect information assets and reduce the likelihood that rogue attacks can succeed.[63] The idea is based on military principles that an enemy is stymied by complex layers and approaches compared to a single line. That is, hackers may be able to penetrate one or two of the defense layers, but multiple security layers increase the chances of catching the attack before it gets too far. Defense in depth includes a firewall as a first line of defense and also antivirus and anti-spyware software, **identity and access management** (IAM), hierarchical passwords, intrusion detection, and biometric verification. Also, as a part of an overall IG program, physical security measures are deployed, such as smartcard or even biometric access to facilities and intensive IG training and auditing.

Controlling Access Using Identity Access Management

IAM software can provide an important piece of the security solution. It aims to prevent unauthorized people from accessing a system and to ensure that only authorized individuals engage with information, including confidential e-documents.

Today's business environment operates in a more extended and mobile model, often including stakeholders outside of the organization. With this more complex and fluctuating group of users accessing information management applications, the idea of identity management has gained increased importance.

The response to the growing number of software applications using inconsistent or incompatible security models is strong identity management enforcement software. These scattered applications offer opportunities not only for identity theft but also for *identity drag*, where the maintenance of identities does not keep up with changing identities, especially in organizations with a large workforce. This can result in theft of confidential information assets by unauthorized or out-of-date access and even failure to meet regulatory compliance, which can result in fines and imprisonment.[64]

> "IAM addresses 'access creep' where employees move to a different department of business unit and their rights to access information fail to get updated."

IAM—along with sharp IG policies—"manages and governs user access to information through an automated, continuous process."[65] Implemented properly, good IAM does keep access limited to authorized users while increasing security, reducing IT complexity, and increasing operating efficiencies.

Critically, *"IAM addresses 'access creep' where employees move to a different department of business unit and*

their rights to access information fail to get updated" (emphasis added).[66]

In France in 2007, a rogue stock trader at Société Générale had in-depth knowledge of the bank's access control procedures from his job at the home office.[67] He used that information to defraud the bank and its clients out of over €7 billion (over $10 billion). If the bank had implemented an IAM solution, the crime may not have been possible.

A robust and effective IAM solution provides for:

- *Auditing*. Detailed audit trails of *who* attempted to access *which information*, and *when*. Stolen identities can be uncovered if, for instance, an authorized user attempts to log in from more than one computer at a time.
- *Constant updating*. Regular reviews of access rights assigned to individuals, including review and certification for user access, an automated recertification process (*attestation*), and enforcement of IG access policies that govern the way users access information in respect to segregation of duties.
- *Evolving roles*. Role life cycle management should be maintained on a continuous basis, to mine and manage roles and their associated access rights and policies.
- *Risk reduction*. Remediation regarding access to critical documents and information.

Enforcing IG: Protect Files with Rules and Permissions

One of the first tasks often needed when developing an IG program that secures confidential information assets is to define roles and responsibilities for those charged with implementing, maintaining, and enforcing IG policies. Corollaries that spring from that effort get down to the nitty-gritty of controlling information access by rules and permissions.

Rules and permissions specify *who* (by roles) is allowed access to *which* documents and information, and even contextually, *from where* (office, home, travel), and *at what times* (work hours, or extended hours). Using the old policy of the *need-to-know* basis is a good rule of thumb to apply when setting up these access policies (i.e. only those who are at a certain level of the organization or are directly involved in certain projects are allowed access to confidential and sensitive information). The roles are relatively easy to define in a traditional hierarchical structure, but today's flatter and more collaborative enterprises present challenges.

To effectively wall off and secure information by management level, many companies and governments have put in place an information security framework—a model that delineates which levels of the organization have access to specific documents and databases as a part of implemented IG policy. This framework shows a hierarchy of the company's management distributed across a range of defined levels of information access. The US Government Protection Profile for Authorization Server for Basic Robustness Environments is an example of such a framework.

Challenge of Securing Confidential E-Documents

Today's various document and content management systems were not initially designed to allow for secure document sharing and collaboration while also preventing document leakage. These software applications were mostly designed before the invention and adoption of newer business technologies that have extended the computing environment. The introduction of cloud computing, mobile PC devices, smartphones, social media, and online collaboration tools all came after most of today's document and content management systems were developed and brought to market.

Thus, vulnerabilities have arisen that need to be addressed with other, complementary technologies. We need to look no further than the WikiLeaks incident and the myriad of other major security breaches resulting in document and data leakage to see that there are serious information security issues in both the public and private sectors.

Technology is the tool, but without proper IG policies and a culture of compliance that supports the knowledge workers following IG policies, any effort to secure confidential information assets will fail. An old IT adage is that even *perfect technology will fail without user commitment*.

Protecting Confidential E-Documents: Limitations of Repository-Based Approaches

Organizations invest billions of dollars in IT solutions that manage e-documents and records in terms of security, auditing, search, records retention and disposition, version control, and so on. These information

management solutions are predominantly repository-based, including enterprise content management (ECM) systems and collaborative workspaces (for unstructured information, such as e-documents). With content or document repositories, the focus has always been on perimeter security—keeping intruders out of the network. But that provides only partial protection. Once intruders are in, they are *in* and have full access to confidential e-documents. For those who are authorized to access the content, there are no protections, so they may freely copy, forward, print, or even edit and alter the information.[68]

The glaring vulnerability in the security architecture of ECM systems is that few protections exist once the information is legitimately accessed.

These confidential information assets, which may include military plans, price lists, patented designs, blueprints, drawings, and financial reports, often can be printed, e-mailed, or faxed to unauthorized parties without any security attached.[69]

The glaring vulnerability in the security architecture of ECM systems is that few protections exist once the information is legitimately accessed.

Also, in the course of their normal work processes, knowledge workers tend to keep an extra copy of the electronic documents they are working on stored at their desktop, or they download and copy them to a tablet or laptop to work at home or while traveling. *This creates a situation where multiple copies of these e-documents are scattered about on various devices and media, which creates a security problem, since they are outside of the repository and no longer secured, managed, controlled, or audited.*

It also creates records management issues in terms of the various versions that might be out there and determining which one is the official business record.

Technologies like firewalls, access controls, and gateway filters can grant or deny access but cannot provide granular enforcement of acceptable use policies that define what users can and cannot do with confidential data and documents.

Apply Better Technology for Better Enforcement in the Extended Enterprise

Protecting E-Documents in the Extended Enterprise

Sharing e-documents and collaborating are essential in today's increasingly mobile and global world. Businesses are operating in a more distributed model than ever before, and they are increasingly sharing and collaborating not only with coworkers but also with suppliers, customers, and even at times competitors (e.g. in pharmaceutical research). This reality presents a challenge to organizations dealing in sensitive and confidential information.[70]

Basic Security for the Microsoft Windows Office Desktop

The first level of protection for e-documents begins with basic protections at the desktop level. Microsoft Office provides ways to password-protect Microsoft Office files, such as those created in Word and Excel, quickly and easily. Many corporations and government agencies around the world use these basic protections. A key flaw or caveat is that *passwords used in protecting documents cannot be retrieved if they are forgotten or lost*.

Where Do Deleted Files Go?

When you delete a file it is gone, right? Actually, it is not (with the possible exception of solid-state hard drives). For example, after a file is deleted in Windows, a simple undelete DOS command can bring back the file, if it has not been overwritten. That is because when files are deleted, they are not really deleted; rather, the space where they reside is marked for reuse and can be overwritten. If it is not yet overwritten, the file is still there. The same process occurs as drafts of documents are created and temp (for *temporary*) files are stored. The portions of a hard drive where deleted or temp files are stored can be overwritten. This is called unallocated space. *Most users are unaware that deleted files and fragments of documents and drafts are stored temporarily on their computer's unallocated space.* So it must be wiped clean and completely erased to ensure that any confidential documents or drafts are completely removed from the hard drive.

IG programs include the highest security measures, which means that an organization must have a policy that includes deleting sensitive materials from a computer's unallocated space and tests that verify such deletion actions are successful periodically.

Lock Down: Stop All External Access to Confidential E-Documents

Organizations are taking other approaches to stop document and data leakage: physically restricting access to a computer by disconnecting it from any network connections and forbidding or even blocking use of any ports. Although cumbersome, these methods are effective in highly classified or restricted areas where confidential e-documents are held. Access is controlled by utilizing multiple advanced identity verification methods, such as biometric means.

Secure Printing

Organizations normally expend a good amount of effort making sure that computers, documents, and private information are protected and secure. However, if your computer is hooked up to a network printer (shared by multiple knowledge workers), all of that effort might have been wasted.[71]

Some basic measures can be taken to protect confidential documents from being compromised as they are printed. You simply invoke some standard Microsoft Office protections, which allow you to print the documents once you arrive in the copy room or at the networked printer. This process varies slightly, depending on the printer's manufacturer. (Refer to the documentation for the printer for details.)

In Microsoft Office, there is an option in the Print Dialog Box for delayed printing of documents (when you physically arrive at the printer).

Serious Security Issues with Large Print Files of Confidential Data

According to Canadian output and print technology expert William Broddy, in a company's data center, a print file of, for instance, investment account statements or bank statements contains all the rich information that a hacker or malicious insider needs. *It is distilled information down to the most important core data about customers, which has been referred to as data syrup since it has been boiled down and contains no mountains of extraneous data, only the culled, cleaned, essential data that gives criminals exactly what they need.*[72]

What most managers are not aware of is that entire print files and sometimes remnants of them stay on the hard drives of high-speed printers and are vulnerable to security breaches. Data center security personnel closely monitor calls to their database. To extract as much data as is contained in print files, a hacker requires hundreds or even thousands of calls to the database, which sets off alerts by system monitoring tools. But retrieving a print file takes only one intrusion, and it may go entirely unnoticed. The files are sitting there; a rogue service technician or field engineer can retrieve them on a routine service call.

> A print file contains all the distilled customer information a hacker might want. Retrieving a print file takes only one intrusion and may go entirely unnoticed.

To help secure print files, specialized hardware devices designed to sit between the print server and the network and cloak server print files are visible only to those who have a cloaking device on the other end.

Organizations must practice good IG and have specific procedures to erase sensitive print files once they have been utilized. For instance, in the example of preparing statements to mail to clients, files are exposed to possible intrusions in at least six points in the process (starting with print file preparation and ending with the actual mailing). These points must be tightly monitored and controlled. Typically an organization retains a print file for about 14 days, though some keep files long enough for customers to receive statements in the mail and review them. *Organizations must make sure that print files or their remnants are secured and then completely erased when the printing job is finished.*

> Files are exposed to possible intrusions in at least six points between print file preparation and final hard-copy mailing.

E-Mail Encryption

Encrypting (scrambling using advanced algorithms) sensitive e-mail messages is an effective step to securing confidential information assets while in transit. Encryption can also be applied to desktop folders and files and even entire disk drives (full disk encryption, or FDE). All confidential or sensitive data and e-documents that are exposed to third parties or transferred over public networks should be secured with file-level encryption, at a minimum.[73]

Secure Communications Using Record-Free E-Mail

What types of tools can you use to encourage the free flow of ideas in collaborative efforts without compromising your confidential information assets or risking litigation or compliance sanctions?

Stream messaging is an innovation that became commercially viable around 2006. It is similar in impact to IRM software, which limits the recipients' ability to forward, print, or alter data in an e-mail message (or reports, spreadsheets, etc.) *but goes further by leaving no record on any computer or server*.

> With stream messaging, no record or trace of communication is left.

Stream messaging is a simple, safe, secure electronic communications system ideal for ensuring that sensitive internal information is kept confidential and not publicly released. Stream messaging is not intended to be a replacement for enterprise e-mail but is a complement to it. If you need an electronic record, e-mail it; if not, use stream messaging.[74]

What makes stream messaging unique is its recordlessness. Streamed messages cannot be forwarded, edited, or saved. A copy cannot be printed as is possible with e-mail. That is because *stream messaging separates the sender's and receiver's names and the date from the body of the message, never allowing them to be seen together*. Even if the sender or receiver were to attempt to make a copy using the print-screen function, these elements are never captured together.[75]

The instant a stream message is sent, it is placed in a temporary storage buffer space. When the recipient logs in to read the message, it is removed from the buffer space. By the time the recipient opens it, the complete stream message no longer exists on the server or any other computer.

This communications approach is Web based, meaning that no hardware or software purchases are required. It also works with existing e-mail systems and e-mail addresses and is completely immune to spam and viruses. Other solutions (both past and present) have been offered, but these have taken the approach of encrypting e-mail or generating e-mail that disappears after a preset time. Neither of these approaches is truly recordless.

Stream messaging is unique because its technology effectively eliminates the ability to print, cut, paste, forward, or save a message. It may be the only electronic communications system that separates the header information —date, name of sender, name of recipient—from the body of the message. This eliminates a traceable record of the communication. Soon many other renditions of secure messaging will be developed.

In addition, stream messaging offers the added protection of being an indiscriminate Web-based service, meaning that the messages and headers are never hosted on the subscribing companies' networks. This eliminates the risk that employers, competitors, or hackers could intercept stream messages, which is a great security benefit for end users.[76]

Digital Signatures

Digital signatures are more than just digitized autographs—they carry detailed audit information used to "detect unauthorized modifications" to e-documents and to "authenticate the identity of the signatory."[77]

Online transactions can be conducted with full trust that they are legal, proper, and binding. They prove that the person whose signature is on the e-document did, in fact, authorize it. A digital signature provides evidence in demonstrating to a third party that the signature was genuine, true, and authentic, which is known as *nonrepudiation*. To repudiate is to dispute, and with digital signatures, a signatory is unable to claim that the signature is forged.

Digital signatures can be implemented a variety of ways—not just through software but also through firmware (programmed microchips), computer hardware, or a combination of the three. Generally, hardware- and firmware-based implementations are more difficult to hack, since their instructions are hardwired.

There is a big difference between digital and electronic signatures. Digital signatures contain additional authenticating information.

Here is a key point: for those who are unfamiliar with the technology, *there is a big difference between electronic signatures and digital signatures*.[78]

An "electronic signature is likely to be a bit-map image, either from a scanned image, a fax copy or a picture of someone's signature, or may even be a typed acknowledgment or acceptance." *A digital signature contains "extra data appended to a message which identifies and authenticates the sender and message data using public-key encryption."*[79]

So digital signatures are the only ones that offer any real security advantages.

Digital signatures are verified by the combination of applying a signatory's private signing key and the public key that comes from the signatory's personal ID certificate. After that, only the public key ID certificate is required for future verifications. "In addition, a checksum mechanism confirms that there have been no modifications to the content."[80]

A formal, trusted **certificate authority** (CA) issues the certificate associated with the public-private key. It is possible to generate self-certified public keys, but these do not verify and authenticate the recipient's identity and are therefore flawed from a security standpoint. The interchange of verified signatures is possible on a global scale, as "digital signature standards are mature and converging internationally."[81]

Requiring a physical signature can disrupt and slow business processes. Digital signatures speed that up and add a layer of security.

After more than 30 years of predictions, the paperless office is almost here. Business process cycles have been reduced, and great efficiencies have been gained since the majority of documents today are created digitally and spend most of their life cycle in digital form, and they can be routed through work steps using business process management (BPM) and work flow software. *However, the requirement for a physical signature frequently disrupts and holds up these business processes*. Documents have to be printed out, physically routed, and physically signed—and often they are scanned back into a document or records management (or contract management) system, which defeats the efficiencies sought.

Often *multiple* signatures are required in an approval process, and some organizations require each page to be initialed, which makes the process slow and cumbersome when it is executed without the benefit of digital signatures. Also, multiple copies are generated—as many as 20—so digital signature capability injected into a business process can account for significant time and cost savings.[82]

Document Encryption

There is some overlap and sometimes confusion between digital signatures and document encryption. Suffice it to say, they work differently, in that document encryption secures a document for those who share a secret key, and digital signatures prove that the document has not been altered and the signature is authentic.

There are e-records management implications of employing document encryption:

> Unless it is absolutely essential, full document encryption is often advised against for use within electronic records management systems as it prevents full-text indexing, and requires that the decryption keys (and application) are available for any future access. Furthermore, if the decryption key is lost or an employee leaves without passing it on, encrypted documents and records will in effect be electronically shredded as no one will be able to read them.
>
> Correctly certified digital signatures do not prevent unauthorized persons reading a document nor are they intended to. They do confirm that the person who signed it is who they say they are, and that the document has not been altered since they signed it. Within a records management system a digital signature is often considered to be an important part of the metadata of a document, confirming both its heritage and its integrity.[83]

Data Loss Prevention (DLP) Technology

The aforementioned document security challenges have given rise to an emerging but critical set of capabilities by a new breed of IT companies that provide **data loss prevention** (DLP) (also called data *leak* prevention). DLP providers create software and hardware appliances that thoroughly inspect all e-documents and e-mail messages before they leave the organization's perimeter and attempt to stop sensitive data from exiting the firewall.

This filtering is based on several factors, but mostly using specified critical content keywords that are flagged by the implementing organization. DLP can also stop the exit of information assets by document types, origin, time of day, and other factors.

DLP systems are designed to detect and prevent unauthorized use and transmission of confidential information.[84] In more detail, DLP is a computer security term referring to systems that identify, monitor, and protect data/documents in all three states: (1) *in use* (endpoint actions), (2) *in motion* (network actions), and (3) *at rest* (data/document storage). DLP accomplishes this by deep content inspection and contextual security analysis of transaction data (e.g. attributes of the originator, the data object, medium, timing, recipient/destination, etc.) with a centralized management framework.

Promise of DLP

The global enterprise data loss prevention market is anticipated to grow at a compound annual growth rate of more than 16% from 2018 to 2023, from about $1.2 billion to $2.5 billion by 2023.[85] Gartner states, that "with adoption of DLP technologies moving quickly down to the small to medium enterprise, DLP is no longer an unknown quantity."[86] Although the DLP market has matured, it suffers from confusion about how DLP best fits into the new mix of security approaches, how it is best utilized (endpoint or gateway), and even the definition of DLP itself.[87]

Data loss is very much on managers' and executives' minds today. The series of WikiLeaks incidents exposed hundreds of thousands of sensitive government and military documents. According to the Ponemon Institute (as reported by DLP experts), data leaks continue to increase annually. Billions of dollars are lost every year as a result of data leaks, with the cost of each breach ranging from an average of $700,000 to $31 million. Some interesting statistics from the study include:

- Almost half of breaches happen while an enterprise's data was in the hands of a third party.
- Over one-third of breaches involved lost or stolen mobile devices.
- The cost per stolen record is approximately $200 to $225.
- One-quarter of breaches were conducted by criminals or with malicious intent.
- More than 80% of breaches compromised more than 1000 records.[88]

What DLP Does Well (and Not So Well)

DLP has been deployed successfully as a tool used to map the flow of data inside and exiting the organization to determine the paths that content takes, so that more sophisticated information mapping, monitoring, and content security can take place.

This use as a traffic monitor for analysis purposes has been much more successful than relying on DLP as the sole enforcement tool for compliance and to secure information assets. Today's technology is simply not fast enough to catch everything. It catches many e-mail messages and documents that users are authorized to send, which slows the network and the business down. This also adds unnecessary overhead, as someone has to go back and release each and every one of the e-mails or documents that were wrongly stopped.

Another downside: *since DLP relies on content inspection, it cannot detect and monitor encrypted e-mail or documents*.

Basic DLP Methods

DLP solutions typically apply one of three methods:

1. Scanning traffic for keywords or regular expressions, such as customer credit card or social security numbers.
2. Classifying documents and content based on a predefined set to determine what is likely to be confidential

and what is not.

3. Tainting (in the case of agent-based solutions), whereby documents are tagged and then monitored to determine how to classify derivative documents. For example, if someone copies a portion of a sensitive document into a different document, this document receives the same security clearance as the original document.[89]

All these methods involve the network administrator setting up a policy clearly defining what is allowed to be sent out and what should be kept in confidence. This policy creating effort is extremely difficult: defining a policy that is *too broad* means accidentally letting sensitive information get out, and defining a policy that is *too narrow* means getting a significant number of false positives and stopping the flow of normal business communications.

Although network security management is well established, defining these types of IG policies is extremely difficult for a network administrator. Leaving this job to network administrators means there will be no collaboration with business units, no standardization, and no real forethought. As a result, many installations are plagued with false positives that are flagged and stopped, which can stifle and frustrate knowledge workers. *The majority of DLP deployments simply use DLP for monitoring and auditing purposes*.

Examining the issue of the dissolving perimeter more closely, a deeper problem is revealed: DLP is binary; it is black or white. Either a certain e-document or e-mail can leave the organization's boundaries or it cannot. This process has been referred to as outbound content compliance.

But this is not how the real world works today. Now there is an increasing need for collaboration and for information to be shared or reside outside the organization on mobile devices or in the cloud.

Most of today's DLP technology cannot address these complex issues on its own. Often additional technology layers are needed.

Data Loss Prevention: Limitations

DLP has been hyped in the past few years, and major security players have made several large acquisitions—especially those in the IRM market. Much like firewalls, DLP started in the form of network gateways that searched e-mail, Web traffic, and other forms of information traveling out of the organization for data that was defined as internal. When it found such data, the DLP blocked transmission or monitored its use.

Soon agent-based solutions were introduced, performing the same actions locally on users' computers. The next step brought a consolidation of many agent- and network-based solutions to offer a comprehensive solution.

IG policy issues are key. What is the policy? All these methods depend on management setting up a policy that clearly defines what is acceptable to send out and what should be kept in confidence.

With DLP, a certain document can either leave the organization's boundaries or it can't. But this is not how the real world works. In today's world there is an increasing need for information to be shared or reside outside the organization on mobile devices or in the cloud. Simply put, *DLP is not capable of addressing this issue on its own, but it is a helpful piece of the overall technology solution*.

Missing Piece: Information Rights Management (IRM)

Another technology tool for securing information assets is information rights management (IRM) software (also referred to as enterprise rights management [ERM] and previously as enterprise digital rights management [e-DRM].) *For purposes of this book, we use the term "IRM" when referring to this technology set, so as not to be confused with electronic records management. Major software companies also use the term "IRM."*

IRM technology provides a sort of security wrapper around documents and protects sensitive information assets from unauthorized access.[90] We know that DLP can search for key terms and stop the exit of sensitive data from the organization by inspecting its content. But it can also prevent confidential data from being copied to external media or sent by e-mail if the person is not authorized to do so. If IRM is deployed, files and documents are protected wherever they may be, with persistent security. *The ability to apply security to an e-document in any state* (in use, in motion, and at rest), across media types, inside or outside of the organization, *is called persistent security*. This is a key characteristic of IRM technology, and it is all done transparently without user intervention.[91]

The ability to secure data at any time, in any state, is called persistent protection.

IRM has the ability to protect e-documents and data wherever they may reside, however they may be used, and in all three data states (at rest, in use, and in transit).[92]

IRM allows for e-documents to be remote controlled, meaning that security protections can be enforced even if the document leaves the perimeter of the organization. This means that e-documents (and their control mechanisms) can be separately created, viewed, edited, and distributed.

IRM provides persistent, ever-present security and manages access to sensitive e-documents and data. IRM provides embedded file-level protections that travel with the document or data, regardless of media type.[93] These protections prevent unauthorized viewing, editing, printing, copying, forwarding, or faxing. So, even if files are somehow copied to a thumb drive and taken out of the organization, e-document protections and usage are still controlled.

The major applications for IRM services include cross-protection of e-mails and attachments, dynamic content protection on Web portals, secure Web-based training, secure Web publishing, and secure content storage and e-mail repositories all while meeting compliance requirements of Sarbanes-Oxley, the *Health Insurance Portability and Accountability Act*, and others. Organizations can comply with regulations for securing and maintaining the integrity of digital records, and IRM will restrict and track access to spreadsheets and other financial data too.

In investment banking, research communications must be monitored, according to National Association of Securities Dealers (NASD) rule 2711, and IRM can help support compliance efforts. In consumer finance, personal financial information collected on paper forms and transmitted by fax (e.g. auto dealers faxing credit applications), or other low-security media can be secured using IRM, directly from a scanner or copier. Importers and exporters can use IRM to ensure data security and prevent the loss of cargo from theft or even terrorist activities, and they also can comply with U.S. Customs and trade regulations by deploying IRM software. Public sector data security needs are numerous, including intelligence gathering and distribution, espionage, and Homeland Security initiatives. Firms that generate intellectual property (IP), such as research and consulting groups, can control and protect access to IP with it. In the highly collaborative pharmaceutical industry, IRM can secure research and testing data.

IRM protections can be added to nearly all e-document types including e-mail, word processing files, spreadsheets, graphic presentations, computer-aided design (CAD) plans, and blueprints. This security can be enforced globally on all documents or granularly down to the smallest level, protecting sensitive fields of information from prying eyes. This is true even if there are multiple copies of the e-documents scattered about on servers in varying geographic locations. Also, the protections can be applied permanently or within controlled time frames. For instance, a person may be granted access to a secure e-document for a day, a week, or a year.

Key IRM Characteristics

Three requirements are recommended to ensure effective IRM:

1. *Security* is foremost; documents, communications, and licenses should be encrypted, and documents should require authorization before being altered.
2. *The system can't be any harder to use* than working with unprotected documents.
3. *It must be easy to deploy and manage*, scale to enterprise proportions, and work with a variety of common desktop applications.[94]

IRM software enforces and manages document access policies and use rights (view, edit, print, copy, e-mail forward) of electronic documents and data. Controlled information can be text documents, spreadsheets, financial statements, e-mail messages, policy and procedure manuals, research, customer and project data, personnel files, medical records, intranet pages, and other sensitive information. IRM provides persistent enforcement of IG and access policies to allow an organization to control access to information that needs to be secured for privacy, competitive, or compliance reasons. *Persistent content security is a necessary part of an end-to-end enterprise security architecture.*

Well, it sounds like fabulous technology, but is IRM really so new? No, it has been around for a decade or more, and continues to mature and improve. It has essentially entered the mainstream around 2004/2005 (when this

author began tracking its development and publishing researched articles on the topic).

IRM software currently is used for persistent file protection by thousands of organizations throughout the world. Its success depends on the quality and consistency of the deployment, which includes detailed policy-making efforts. *Difficulties in policy maintenance and lack of real support for external sharing and mobile devices have kept first-wave IRM deployments from becoming widespread, but this aspect is being addressed by a second wave of new IRM technology companies*.

Other Key Characteristics of IRM

Policy Creation and Management

IRM allows for the creation and enforcement of policies governing access and use of sensitive or confidential e-documents. The organization's IG team sets the policies for access based on role and organizational level, determining what employees can and cannot do with the secured e-documents.[95] The IG policy defined for a document type includes these following controls:

1. Viewing
2. Editing
3. Copy/Paste (including screen capture)
4. Printing
5. Forwarding e-mail containing secured e-documents

Access to sensitive e-documents may be revoked at any time, no matter where they are located or what media they are on, since each time a user tries to access a document, access rights are verified with a server or cloud IRM application. This can be done remotely—that is, when an attempt is made to open the document, an authorization must take place. In cloud-based implementations, it is a matter of simply denying access.

Decentralized Administration

One of the key challenges of e-document security traditionally is that a system administrator had access to documents and reports that were meant only for executives and senior managers. With IRM, the e-document owner administers the security of the data, which considerably reduces the risk of a document theft, alteration, or misuse.

Auditing

Auditing provides the smoking-gun evidence in the event of a true security breach. Good IRM software provides an audit trail of how all documents secured by it are used. Some go further, providing more detailed document analytics of usage.

Integration

To be viable, IRM must integrate with other enterprisewide systems, such as ECM, customer relationship management, product life cycle management, enterprise resource planning, e-mail management, message archiving, e-discovery, and a myriad of cloud-based systems. This is a characteristic of today's newer wave of IRM software.

This ability to integrate with enterprise-based systems does not mean that IRM has to be deployed at an enterprise level. *The best approach is to target one critical department or area with a strong business need and to keep the scope of the project narrow to gain an early success before expanding the implementation into other departments*.

IRM embeds protection into the data (using encryption technology), allowing files to protect themselves. IRM may be the best available security technology for the new mobile computing world of the permeable perimeter.[96]

> IRM technology protects e-documents and data directly rather than relying on perimeter security.

With IRM technology, a document owner can selectively prevent others from viewing, editing, copying, or printing it. Despite its promise, most enterprises do not use IRM, and if they do, they do not use it on an enterprise-wide basis. This is due to the high complexity, rigidity, and cost of legacy IRM solutions.

It is clearly more difficult to use documents protected with IRM—especially when policymaking and maintenance is not designed by role but rather by an individual. Some early implementations of IRM by first-to-market software development firms had as many as 200,000 different policies to maintain (for 200,000 employees). These have since been replaced by newer, second-wave IRM vendors, who have reduced that number to a mere 200 policies, which is much more manageable. Older IRM installations require intrusive plug-in installation; they are limited in the platforms they support, and they largely prevent the use of newer platforms, such as smartphones, iPads, and other tablets. This is a real problem in a world where almost all executives carry a smartphone and use of tablets (especially the iPad) is growing.

Moreover, due to their basic design, first-wave or legacy IRM is not a good fit for organizations aiming to protect documents shared outside company boundaries. These outdated IRM solutions were designed and developed in a world where organizations were more concerned with keeping information inside the perimeter than protecting information beyond the perimeter.

Most initial providers of IRM focused on internal sharing and are heavily dependent on Microsoft Active Directory (AD) and lightweight directory access protocol (LDAP) for authentication. Also, the delivery model of older IRM solutions involves the deployment and management of multiple servers, SQL databases, AD/LDAP integration, and a great deal of configuration. This makes them expensive and cumbersome to implement and maintain. Furthermore, these older IRM solutions do not take advantage of or operate well in a cloud computing environment.

Although encryption and legacy IRM solutions have certain benefits, they are extremely unwieldy and complex and offer limited benefits in today's technical and business environment. Newer IRM solutions are needed to provide more complete DLS.

Embedded Protection

IRM embeds protection into the data (using encryption technology), allowing files to protect themselves. IRM may be the best available security technology for the new mobile computing world of the permeable perimeter.[97]

> IRM technology protects e-documents and data directly rather than relying on perimeter security.

Is Encryption Enough?

Many of the early solutions for locking down data involved encryption in one form or another:

- E-mail encryption
- File encryption
- Full disk encryption (FDE)
- Enterprise-wide encryption

These encryption solutions can be divided into two categories: encryption *in transit* (e.g. e-mail encryption) and encryption *at rest* (e.g. FDE).

The various encryption solutions mitigate some risks. In the case of data in transit, these risks could include an eavesdropper attempting to discern e-mail or network traffic. In the case of at-rest data, risks include loss of a laptop or unauthorized access to an employee's machine. The most advanced solutions are capable of applying a policy across the organization and encrypting files, e-mails, and even databases. However, encryption has its caveats.

Most simple encryption techniques necessarily involve the decryption of documents so they can be viewed or edited. At these points, the files are essentially exposed. Malware (e.g. Trojan horses, keystroke loggers) installed on a computer may use the opportunity to send out the plain-text file to unauthorized parties. Alternatively, an employee may copy the contents of these files and remove them from the enterprise.

Device Control Methods

Another method that is related to DLP is **device control**. Many vendors offer software or hardware that prevents users from copying data via the USB port to portable drives and removing them from the organization in this manner. These solutions are typically as simple as blocking the ports; however, some DLP solutions,

when installed on the client side, can selectively prevent the copying of certain documents.[98]

Thin Clients

One last method worth mentioning is the use of thin clients to prevent data leaks. These provide a so-called walled garden containing only the applications users require to do their work, via a diskless terminal. This prevents users from copying any data onto portable media; however, if they have e-mail or Web access applications, they still can send information out via e-mail, blogs, or social networks.

Note about Database Security

Database security and monitoring is addressed in Chapter 10, "Information Governance and Information Technology Functions."

Compliance Aspect

Compliance has been key in driving companies to invest in improving their security measures, such as firewalls, antivirus software, and DLP systems. More than 400 regulations exist worldwide mandating a plethora of information and data security requirements. One example is the *Payment Card Industry Data Security Standard* (PCI-DSS), which is one of the strictest regulations for credit card processors. Companies that fail to comply with these regulations are subject to penalties of up to $500,000 per month for lost financial data or credit card information. Forrester Research estimated the per-record cost of a breach is $90 to $305. But do compliance activities always result in adequate protection of your sensitive data? In many cases the answer is no. It is important to keep in mind that *being formally compliant does not mean the organization is actually secure*. In fact, compliance is sometimes used as a fig leaf, covering a lack of real document security. One needs to look no further than to the recent series of major document leakage incidents to understand this. Those all came from highly secure and regulated entities, such as banks, hospitals, and the military.

Hybrid Approach: Combining DLP and IRM Technologies

An idea being promoted recently is to make IRM an enforcement mechanism for platforms like DLP. Together, DLP and IRM accomplish what they independently cannot. Enterprises may be able to use their DLP tools to discover data flows, map them out, and detect transmissions of sensitive information. They can then apply their IRM or encryption protection to enforce their confidentiality and information integrity goals.[99]

Several vendors in the fields of DLP, encryption, and IRM have already announced integrated products. However, at this point in time, most IRM solutions are by no means ready for prime time when it comes to this use. Only a select few select second-wave IRM software providers can offer comprehensive, streamlined, persistent security across many platforms.

As the enterprise perimeter dissolves, document and data security should become the focus of the Internet security field. However, most legacy solutions, such as encryption and legacy IRM, are complex and expensive and provide only a partial solution to the key problems. Combining several methods offers effective countermeasures, but an ultimate solution has not yet arrived.

Securing Trade Secrets After Layoffs and Terminations

In today's global economy—which has shifted labor demands—huge layoffs are not uncommon in the corporate and public sectors. The act of terminating an employee creates document security and IP challenges while raising the question: How does the organization retrieve and retain its IP and confidential data? An IG program to secure information assets must also deal with everyday resignations of employees who are in possession of sensitive documents and information.[100]

According to Peter Abatan, author of the Enterprise Digital Rights Management blog, "As a general rule *all organizations should classify all their documents with the aim of identifying the ones that need persistent protection"* (emphasis added). That is to say, documents should be protected at all times, regardless of where they travel and who is using them, while the organization still retains control of usage rights. There are two basic technological approaches to this protection:

1. The first, as discussed earlier in this chapter, is combining *IRM with DLP*; DLP is used to conduct deep content inspection and identify all documents that may contain sensitive information, then the DLP agent "notifies the enterprise [information] rights management engine that sensitive information is about to be copied to external media or outside the firewall and therefore needs to be encrypted."

2. The second is using a form of *context-sensitive IRM* "in which all documents that contain sensitive data defined in the [global] data dictionary [are] automatically encrypted."

These two technological approaches must be fostered by an IG program. They can have significant positive impact in protecting sensitive information, no matter where it is located, and can help document owners withdraw access to its sensitive documents at any time.

Organizations must educate their employees to increase awareness of the financial and competitive impact of breaches and to clarify that sensitive documents are the property of the organization. If those handling sensitive documents are informed of the benefits of IRM and related technologies, they will be more vigilant in their efforts to keep information assets secure.

Persistently Protecting Blueprints and CAD Documents

Certain IRM software providers have focused on securing large-format engineering and design documents, and they have made great strides in the protection of computer-aided design files. As much as *95% of CAD files are proprietary designs and represent valuable, proprietary IP of businesses worldwide*. And CAD files are just as vulnerable as any other e-document in that, when unprotected, they "can be e-mailed or transferred to another party without the knowledge of the owner of the content."[101]

> As much as 95% of CAD files are proprietary designs and represent valuable IP.

In today's global economy, it is common to conduct manufacturing operations in markets where labor is inexpensive and regulations are lax. Many designs are sent to China, Indonesia, and India for manufacturing. Although they usually are accompanied by binding confidential disclosure contracts, but these agreements are often difficult to enforce, especially given the disparity in cultures and laws. And what happens if a rogue employee in possession of designs and trade secrets absconds with them and sells them to a competitor? Or starts a competing business? There are a number of examples of this happening.

Owners of valuable proprietary IP must vigilantly protect it; the very survival of the business may depend on it. Monitoring and securing IP wherever it might travel is now a business imperative.

Theft of IP and confidential information represents a clear and present danger to all types of businesses, especially global brands dependent on proprietary designs for a competitive advantage. Immediate IG action by executive management is required to identify possible leaks and plug the holes. Not safeguarding IP and confidential or sensitive documents puts the organization's competitive position, strategic plans, revenue stream, and very future at risk.

Securing Internal Price Lists

In 2010, it was reported that confidential information about the advertising expenditures of some of Google's major accounts was leaked to the public.[102] This may not seem like a significant breach, but, in fact, with this information, Google's customers can determine if they are getting a preferred price schedule, and competitors can easily undercut Google's pricing for major customers. According to Peter Abatan, "[It is clear] why this information is so critical to Google that this information is tightly secured."

Is your company's price list secured at all times? Price lists are confidential information assets, and if they are revealed publicly, major customers could demand steeper discounts and business relationships could suffer irreparable damage, especially if customers find out they are paying more for a product or service than their competitors.

A company's price list is critical to an organization because it impacts all aspects of the business, from the ability to generate revenue to private dealings with customers and suppliers. IRM should be used to protect price lists, and printing of these valuable lists must be monitored and controlled using secure printing methods and document analytics.

Confidential information should be persistently protected throughout their document life cycle in all three states (at rest, in motion, and in use) *so that if they are compromised or stolen, they are still protected and controlled by the owning organization*.

Approaches for Securing Data Once It Leaves the Organization

It is obvious with today's trends that, as Andrew Jaquith of Forrester Research states, "The enterprise security perimeter is quickly dissolving." A lot of valuable information is routed outside the owning organization through unsecured e-mail. A breach can compromise competitive position, especially in cases dealing with personnel files and marketing plans or merger details. Consider for a moment that even proprietary software and company financial statements are sent out. Exposure of this data can have real financial impact. Without additional protections, such as IRM and e-mail encryption, these valuable information assets are often out of the control of the IT department of the owning organization.[103]

Third-party possession or control of enterprise data is a critical point of vulnerability, and many organizations realize that securing data outside the organizational perimeter is a high priority. But a new concept has cropped up of late that bucks unconventional wisdom: "*control does not require ownership.*"

Instead of focusing on securing devices where confidential data is accessed, the new thinking focuses on securing the data and documents directly. With this new mind-set, security can be planned under the assumption that the enterprise owns its data but none of the devices that access it. As Jaquith states, "*Treat all endpoints as hostile*" (emphasis added). Forrester Research refers to this concept as the "zero-trust model of information security." Zero trust, according to Jaquith, is "centered on the idea that security must become ubiquitous" throughout an organization's infrastructure.

Forrester has developed a new network architecture that builds security into the DNA of a network, using a mixture of five data security design patterns:

1. *Thin client.* Access information online only, with no local operations, using a diskless terminal that cannot store data, documents, or programs so confidential information stays stored and secured centrally. For additional security, "IT can restrict host copy-and-paste operations, limit data transfers, and require strong or two-factor authentication using SecurID or other tokens."
2. *Thin device.* Devices such as smartphones, which have limited computing resources, Web surfing, e-mail, and basic Web apps that locally conduct no real information processing are categorized as thin devices. In practice, these devices do not hold original documents but merely copies, so the official business record or master copy cannot be altered or deleted. A nice feature of many smartphones is the ability to erase or wipe data remotely, in the event the device is lost. This is a little added insurance, and it makes smartphones "truly 'disposable,' unlike PCs," according to Jaquith.
3. *Protected process.* This approach allows local processing with a PC where confidential e-documents and data are stored and processed in a partition that is highly secure and controlled. This processing can occur even if the PC is not owned and controlled by the organization. "The protected process pattern has many advantages: local execution, offline operation, central management, and a high degree of granular security control, including remote wipe [erase]." A mitigating factor to consider here is most business PCs today are Windows based, and the world is rapidly moving to other, more nimble platforms.
4. *Protected data.* Deploying IRM and embedding security into the documents (or data) provides complete DLS. The newer wave of more sophisticated, easier-to-use IRM vendors have role-based policy implementation and such features as "contextual" enforcement, where document rights are dependent on the *context*—that is, *where* and *when* a user attempts access. For instance, allow access to documents on workers' desktops but not on their laptops; or provide access to printing confidential documents at the facility during office hours but not after. "*Of all the patterns in the Zero Trust data security strategy, protected data is the most fine-grained and effective because it focuses on the information, not its containers.*"
5. *Eye in the sky.* This design pattern uses technologies such as DLP to scan network traffic content and halt confidential documents or sensitive data at the perimeter. Deployed properly, DLP is "ideal for understanding the velocity and direction of information flow and for detecting potential breaches, outliers, or anomalous transmissions." It should be noted that DLP does not provide complete protection. To do so would mean that many legitimate and sanctioned e-mails and documents would be held up for inspection, thus slowing the business process. As stated earlier, DLP is best for discovering information flows and monitoring network traffic. Another negative is that you cannot always require partner organizations and suppliers to install DLP on their computers. So this is a complementary technology, not a complete solution to securing confidential information assets.

By discarding the "age-old conflation of ownership and control, enterprises will be able to build data protection programs that encompass all possible ownership scenarios, including Tech Populism, offshoring, and outsourcing."

Document Labeling

Document labeling is "an easy way to *increase user awareness about the sensitivity of information* in a document" (emphasis added).[104] What is it? It is the process of attaching a label to classify a document. For instance, who would not know that a document labeled "confidential" is indeed confidential? If the label appears prominently at the top of a document, it is difficult for persons accessing it to claim they did not know it was sensitive.

The challenge is to *standardize and formalize the process* of *getting the label onto the document—enterprise-wide*. This issue would be addressed in an IG effort focused on securing confidential e-documents, or may also be a part of a classification and taxonomy design effort. It cannot simply be left up to users to type in labels themselves, or it will not be sufficiently executed and will end up leaving a mishmash of labeled documents without any formal classification.

Another great challenge is legacy or archived documents, which are the lion's share of an organization's information assets. How do you go back and label those? One by one? Nope. Not practical.

Some content repositories or portals, such as Microsoft SharePoint, provide some functionality toward addressing the document labeling challenge. SharePoint is the most popular platform for sharing documents today.

SharePoint has an information management policy tool called Labels, which can be used to add document labels, such as *Confidential*, to the top of documents.

There are several options available for administrators to customize the labels, including the ability to:

1. Prompt users to add the label when they save or print, rather than relying on the user to click the Label button in the ribbon;
2. Specify labels containing static text and/or variables such as Project Name;
3. Control the appearance of the labels, such as font, size, and justification.[105]

The labels are easily added from within Microsoft Office Word, PowerPoint, and Excel. One method that can be used is for the user to click the Label button on the Insert ribbon group; another method is to add the label through a prompt that appears when a user saves or prints a document (if the administrator has configured this option).

The labeling capabilities in document and content management systems such as Microsoft's SharePoint are a good start for increasing user awareness and improving the handling of sensitive documents. However, *the document labeling capabilities of SharePoint are basic and limited*. These basic capabilities may provide a partial or temporary solution, although organizations aiming for a high level of security and confidentiality for their documents will need to search for supplemental technologies from third-party software providers. For instance, finding the capabilities to label documents in bulk rather than one by one, add watermarks, or force users to save or print documents with a standard document label that cannot be altered may require looking at alternatives. Some are software vendors that have enhanced the SharePoint document labeling capability and may provide the complete solution.

Document Analytics

Some software providers also provide document analytics capabilities that monitor the access, use, and printing of documents and create real-time graphical reports of document use activities. These capabilities are *very* valuable.

Document analytics allow a compliance officer or system administrator to view exactly how many documents a user accesses in a day and how many documents the user accesses *on average*. Using this information, analytics monitors can look for spikes or anomalies in use. It is also possible to establish baselines and compare usage with that of an employee's peers as well as his or her past document usage. If, for instance, a user normally accesses an average of 25 documents a day and that suddenly spikes to 200, the system sends an alert, and perhaps it is time to pay a visit to that person's office. Or, if an employee normally prints 50 pages per day, then one day prints 250 pages, a flag is raised. Document analytics capabilities can go so far as to calculate the average time a user spends reading a document; significant time fluctuations can be flagged as potentially suspicious activity.

Confidential Stream Messaging

E-mail is dangerous. It contains much of an organization's confidential information, and 99% of the time it is sent out unsecured. It has been estimated that as many as 20% of e-mail messages transmitted pose a legal, financial, or regulatory threat to the organization. Specifically, "34 percent of employers investigated a leak of confidential business information via e-mail, and an additional 26 percent of organizations suffered the exposure of embarrassing or sensitive information during the course of a year," according to Nancy Flynn, Executive Director of the ePolicy Institute. These numbers are rising, giving managers and business owners cause to look for confidential messaging solutions.[106]

Since stream messaging (a form of ephemeral messaging) separates the header and identifying information from the message, sends them separately, and leaves no record or trace, it is a good option for executives and managers, particularly when engaged in sensitive negotiations, litigation, or other highly confidential activities. Whereas e-mail leaves behind an indelible fingerprint that lives forever on multiple servers and systems, stream messaging does not.

Business records, IP and trade secrets, and confidential executive communications can be protected by implementing stream messaging. It can be implemented alongside and in concert with a regular e-mail system, but clear rules on the use of stream messaging must be established, and access to it must be tightly restricted to a small circle of key executives and managers.

The ePolicy Institute offers seven steps to controlling stream messaging:

1. Work with your legal counsel to define "business record" for your organization on a companywide basis. Establish written records retention policies, disposition and destruction schedules. And litigation hold rules. Support the e-mail retention policy with a bona fide e-mail archiving solution to facilitate the indexing, preservation, and production of legally authentic records. Implement a formal electronic records management system to manage all records.
2. Work with your legal counsel to determine when, how, why, and with whom confidential stream messaging is the most appropriate, effective—and legally compliant—way to hold recordless, confidential business discussions *when permanent records are not required.*
3. In order to preserve attorney-client privilege, a phone call or confidential electronic messaging may be preferable to e-mail. Have corporate counsel spell out the manner in which executives and employees should communicate with lawyers when discussing business, seeking legal advice, or asking questions related to specific litigation.
4. Define key terms for employees. Don't assume employees understand what management means when using terms like "confidential," "proprietary," "private" or "intellectual property," and so on. Employees must clearly understand definitions if they are to comply with confidentiality rules.
5. Implement written rules and policies governing the use of e-mail and confidential stream messaging. E-policies should be written clearly and should be easy for employees to access and understand. Make them [as] "short and sweet" as possible. Do not leave anything up to interpretation.
6. Distribute a hard copy of the new confidential messaging policy, e-mail policy and other electronic communications (e.g., social media, blogs). Insist that each and every employee signs and dates the policy, acknowledging that they understand and accept it and that disciplinary action including termination may result from violation of the organization's established policies.
7. Educate, educate, educate. Ensure that all employees who need to know are able to discern between e-mail that leaves a potential business record and stream messaging which does not, and what is confidential.[107]

Securing confidential information assets effectively requires an eclectic, multifaceted approach. It takes clear and enforced IG policies, a collection of technologies, and regular testing and audits, both internally and by a trusted third party.

CHAPTER SUMMARY: KEY POINTS

- Information Privacy refers to individuals or corporations controlling what others know about them.
- Privacy came of age in 2018, when the European Union General Data Protection Regulation (GDPR) went into effect. Its impact was felt across the globe.
- In May 2018, the EU GDPR provided the most significant privacy framework, policy, and regulatory overhaul in history.

- Privacy policies are ways for organizations to explain what they do with PII.
- Privacy Notices are typically exclusive to external facing stakeholders or to the public, where privacy policy could be both.
- In 2009, the Madrid Resolution brought 50 countries together to provide further guidance on information privacy.
- HIPAA was enacted in 1996 to improve the delivery of healthcare services and to provide standards on how patient records are handled in data exchanges.
- In 2000, HIPAA was amended to include the "Privacy Rule"; in 2003, the "Security Rule" was added that focused on the protection of EMRs.
- The HITECH Act was enacted as part of the American Recovery and Reinvestment Act of 2009 to promote the adoption of health information technology.
- The protection of personally identifiable information (PII) is a core focus of IG efforts.
- In the United States, the California Consumer Privacy Act (CCPA) goes into effect January 1, 2020. The CCPA is waking up other US states to the need for more robust privacy legislation.
- The average cost of a data breach in 2018 was nearly $4 million.
- Attacks on organizations' networks and theft of their IP continue to increase.
- Attacks can continue in organizations for years before they are uncovered—if they are discovered at all.
- All organizations should classify all their documents with the aim of identifying the ones that need persistent security protection.
- Today's ECM and document management solutions rely mostly on perimeter security and were not designed to allow for secure document sharing and collaboration.
- Businesses are operating in a more distributed model than ever before, and they are increasingly sharing and collaborating—exposing confidential documents.
- Secure document printing reduces the chance that files can be compromised during or after printing. There are various methods to secure the print stream, depending on the print manufacturer. Copies or remnants of large print files often exist unsecured on the hard drives of high-speed printers. These files must be completely wiped to ensure security.
- Identity and access management (IAM) software governs user access to information through an automated, continuous process that addresses access creep, whereby employees move to a different business unit and their access rights are not updated.
- Data governance (DG) software is another tool that looks at who is accessing which documents and creates a matrix of roles and access along behavioral lines.
- Encrypting sensitive e-mail messages is an effective step to securing confidential information assets while in transit. Encryption can be applied to desktop folders and files.
- For e-mail communication with no trace or record, stream messaging is a solution.
- Digital signatures authenticate the identity of the signatory and prove that the signature was, in fact, generated by the claimed signatory. This is known as nonrepudiation.
- Data loss prevention technology performs a "deep content inspection" of all e-documents and e-mails before they leave the organization's perimeter to stop sensitive data from exiting the firewall.
- DLP can be used to discover the flow of information within an organization. Additional security tools can then be applied. This may be its best use.
- Information rights management software enforces and manages use rights of electronic documents. IRM provides a sort of security wrapper around documents and protects sensitive information assets from unauthorized use or copying. IRM is also known as enterprise rights management.
- Persistent security tools like IRM should be enforced on price lists, proprietary blueprints, and CAD designs. Printing these documents should be highly restricted.
- DLP started in the form of network gateways (much like firewalls) that searched e-mails, Web traffic, and other forms of information for data that was defined as internal. When it detected such data, it

blocked it from leaving the perimeter or monitored its use.

- Combining IRM and DLP technologies is the best available approach to securing e-documents and data. Other encryption methods should also be utilized, such as e-mail encryption and FDE.
- The use of thin-client and thin-device architecture can reduce security threats to confidential information assets.
- Document analytics monitor the access, use, and printing of documents and create real-time graphical reports of document use activities.
- Document labeling is an easy way to increase user awareness about the sensitivity of information in a document.
- Stream messaging is a way to conduct sensitive business negotiations and activities without leaving a business record. Legal counsel must be consulted, and clear policies for regular e-mail versus stream messaging must be established and enforced.

Notes

1. Dalvin Brown, "Americans Are More Concerned with Data Privacy Than Job Creation, Study Shows," *USA Today*, November 9, 2018, https://www.USAToday.com/story/money/2018/11/09/Americans-more-concerned-data-privacy-than-healthcare-study-says/1904796002/.
2. F. Bélanger and R. E. Crossler, ""Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS Quarterly* 35, no. 4 (2011): 1017–1041. doi:10.2307/41409971.
3. R. Gradwohl and R. Smorodinsky, "Perception Games and Privacy," *Games and Economic Behavior* 104 (2017): 293–308. doi:10.1016/j.geb.2017.04.006.
4. P. Swire and S. Bermann, *Information Privacy: Official Reference for the Certified Information Privacy Professional (CIPP)* (York, ME: International Association of Privacy Professionals, 2008), 7.
5. J. Camenisch, "Information Privacy?" *Computer Networks* 56, no. 18 (2012): 3834–3848. doi:10.1016/j.comnet.2012.10.012
6. O. Shy and R. Stenbacka, "Customer Privacy and Competition," *Journal of Economics & Management Strategy* 25, no. 3 (2016): 539–562. doi:10.1111/jems.12157
7. https://www.privacyrights.org/data-breaches.
8. Shy and Stenbacka, "Customer Privacy and Competition."
9. S. Wilson et al., "Analyzing Privacy Policies at Scale: From Crowdsourcing to Automated Annotations," *ACM Transactions on the Web* 13, no. 1 (2018): 1–29. doi:10.1145/3230665
10. "Sample Data Protection Policy Template," https://iapp.org/resources/article/sample-data-protection-policy-template-2/ (accessed June 7, 2019).
11. Ibid.
12. Bob Siegel, "Privacy Policy or Privacy Notice: What's the Difference?" IDG Contributor Network, May 4, 2016, https://www.csoonline.com/article/3063601/privacy/privacy-policies-and-privacy-notices-whats-the-difference.html.
13. Pam Dixon, "A Brief Introduction to Fair Information Practices," World Privacy Forum, January 4, 2008, https://www.worldprivacyforum.org/2008/01/report-a-brief-introduction-to-fair-information-practices/.
14. "The HEW Report: Defining the Fair Information Practices," CIPP Guide, August 23, 2012, https://www.cippguide.org/2012/08/23/the-hew-report-defining-the-fair-information-practices/.
15. "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data," www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm (accessed June 7, 2019).
16. "Data Protection Authorities from over 50 Countries Approve the 'Madrid Resolution' on International

Privacy Standards," 31st International Conference of Data Protection and Privacy, Madrid, November 6, 2009, www.privacyconference2009.org/media/notas_prensa/common/pdfs/061109_estandares_internacionales_en.pdf.

17. "Comments by the DHS Privacy Office and the Staff of the U.S. Federal Trade Commission on the Joint Proposal for International Standards on the Protection of Privacy with Regard to the Processing of Personal Data," August 10, 2010, https://www.dhs.gov/xlibrary/assets/privacy/privacy_comments_madrid_resolution_082010.pdf.

18. Abu Bakar Munir, "Madrid Resolution: A Step Towards a Privacy Treaty?," December 29, 2009,http://profabm.blogspot.com/2009/12/another-privacy-standard.html.

19. "The History of the General Data Protection Regulation," European Data Protection Supervisor, https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en (accessed June 7, 2019).

20. Brian Fung, "Why You're Getting Flooded with Privacy Notifications in Your Email," Washington Post, May 25, 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/05/25/why-youre-getting-flooded-with-privacy-notifications-in-your-email/?noredirect=on&utm_term=.56b9d972bfa3.

21. "General Data Protection Regulation," https://gdpr-info.eu/ (accessed June 7, 2019).

22. "GDPR Fines and Penalties," https://www.gdpreu.org/compliance/fines-and-penalties/ (accessed June 7, 2019).

23. Ibid.

24. Russell Brandom, "Facebook and Google Hit with $8.8 Billion in Lawsuits on Day One of GDPR," *The Verge*, May 25, 2018, https://www.theverge.com/2018/5/25/17393766/facebook-google-gdpr-lawsuit-max-schrems-europe.

25. James Sanders, "To Save Thousands on GDPR Compliance, Some Companies Are Blocking All EU Users," *Tech Republic*, May 7, 2018, https://www.techrepublic.com/article/to-save-thousands-on-gdpr-compliance-some-companies-are-blocking-all-eu-users/.

26. Rani Lofstrom, "Securing Your Digital Transformation," March 6, 2019, https://www.microsoft.com/security/blog/2019/03/06/securing-your-digital-transformation/.

27. Alex Scroxton, "Facebook Facing 10 GDPR Investigations in Ireland," *Computer Weekly*, March 1, 2019, https://www.computerweekly.com/news/252458664/Facebook-facing-10-GDPR-investigations-in-Ireland.

28. Kalev Leetaru, "Facebook's Password Breach Suggests the Public Sees Cybersecurity as Obsolete," *Forbes*, March 23, 2019, https://www.forbes.com/sites/kalevleetaru/2019/03/23/facebooks-password-breach-suggests-the-public-sees-cybersecurity-as-obsolete/#7598a3313e24.

29. Interview with Scott Renfro. March 21, 2019. "Facebook Stored Hundreds of Millions of User Passwords in Plain Text for Years." *Krebs on Security*. https://krebsonsecurity.com/2019/03/facebook-stored-hundreds-of-millions-of-user-passwords-in-plain-text-for-years/

30. IAPP, *Certified Information Privacy Study Guide*, Study Guide edition (CreateSpace Independent Publishing Platform, 2015).

31. Swire and Bermann, *Information Privacy*, 7.

32. Fair Credit Reporting Act, 15 U.S.C. §1681, revised September 2018, https://www.ftc.gov/system/files/545a_fair-credit-reporting-act-0918.pdf.

33. https://www.hhs.gov/ocr/index.htm.

34. Wilson et al., "Analyzing Privacy Policies at Scale."

35. Ibid.

36. Ibid.

37. "HITECH Act Enforcement Interim Final Rule," https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html.

38. "What Are the Penalties for HIPAA Violations?" *HIPAA Journal*, June 25, 2015, https://www.hipaajournal.com/what-are-the-penalties-for-hipaa-violations-7096/.

39. Wilson et al., "Analyzing Privacy Policies at Scale."

40. Ibid.

41. 940 CMR 27.00: Safeguard of Personal Information, Office of Attorney General Maura Healey, January 22, 2010, https://www.mass.gov/regulations/940-CMR-2700-safeguard-of-personal-information.

42. Dipayan Ghosh, "What You Need to Know About California's New Data Privacy Law," *Harvard Business Review*, July 11, 2018, https://hbr.org/2018/07/what-you-need-to-know-about-californias-new-data-privacy-law.

43. Heather Kelly, "California Passes Strictest Online Privacy Law in the Country," *CNN Business*, June 28, 2018, https://money.cnn.com/2018/06/28/technology/california-consumer-privacy-act/index.html.

44. Scott Allbert, "CA Consumer Privacy Act," *IG World*, June 4, 2019, https://infogovworld.com/ig-topics/ccpa-new-privacy-law/.

45. Ibid.

46. "APEC Privacy Framework (2015)," Asia-Pacific Economic Cooperation, August 2017, https://www.apec.org/Publications/2017/08/APEC-Privacy-Framework-(2015).

47. Douglas B. Laney, *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage* (New York: Taylor & Francis, 2018).

48. Shira Scheindlin and Daniel Capra, "The Sedona Conference," *Electronic Discovery and Digital Evidence* (Thomson Reuters, 2009), p. 204, www.amazon.com/Scheindlin-Conferences-Electronic-Discovery-Evidence-ebook/dp/B00AUE0LRI.

49. "The Global Cost of a Data Breach Is Up in 2018," https://securityintelligence.com/ponemon-cost-of-a-data-breach-2018/ (accessed December 13, 2018).

50. "The 10 Biggest Data Breaches of 2018 … So Far," July 16, 2018, https://blog.barkly.com/biggest-data-breaches-2018-so-far. .

51. "Marriott Data Breach Is Traced to Chinese Hackers as U.S. Readies Crackdown on Beijing," *New York Times*, December 11, 2018, https://www.nytimes.com/2018/12/11/us/politics/trump-china-trade.html.

52. "Annual Number of Data Breaches and Exposed Records in the United States from 2005 to 2018," https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/ (accessed December 13, 2018).

53. Jim Finkle, "'State Actor' behind Slew of Cyber Attacks," *Reuters*, August 3, 2011, www.reuters.com/article/2011/08/03/us-cyberattacks-idUSTRE7720HU20110803.

54. Ibid.

55. Ibid.

56. Ibid.

57. Peter Abatan, "Persistently Protecting Your Computer Aided Designs," Enterprise Digital Rights Management, http://enterprisedrm.tumblr.com/post/1423979379/persistently-protecting-your-computer-aided-designs (accessed August 18, 2011).

58. Ari Ruppin, e-mail to author, March 20, 2011.

59. Sam Narisi, "IT's Role in Secure Staff Cuts, March 2, 2009, http://www.financetechnews.com/its-role-in-secure-staff-cuts/.

60. Ibid.

61. Oracle White Paper, "Oracle Information Rights Management 11g—Managing Information Everywhere It Is Stored and Used" (March 2010), www.oracle.com/technetwork/middleware/webcenter/content/irm-technical-whitepaper-134345.pdf, p. 4.

62. Ibid.

63. Open Web Application Security Project, "Defense in Depth," https://www.owasp.org/index.php/Defense_in_depth (accessed June 24, 2013).

64. HCL, "Identity and Access Management Services," www.hclisd.com/identity-and-access-management.aspx (accessed September 2, 2011).

65. Ibid.

66. Ibid.

67. Nicola Clark and David Jolly, "Fraud Costs Bank 7.1 Billion," *New York Times*, January 25, 2008, www.nytimes.com/2008/01/25/business/worldbusiness/25bank-web.html?hp.

68. Oracle White Paper, "Oracle Information Rights Management 11g."

69. Robert Smallwood, "E-DRM Plugs ECM Security Gap," *KM World,* April 1, 2008, www.kmworld.com/Articles/News/News-Analysis/E-DRM-plugs-ECM-security-gap-41333.aspx.

70. Adi Ruppin, e-mail to author, March 20, 2011.

71. Annik Stahl, "Secure Printing: No More Mad Dashes to the Copy Room," http://office.microsoft.com/en-us/help/secure-printing-no-more-mad-dashes-to-the-copy-room-HA001227631.aspx (accessed August 22, 2011).

72. William Broddy, telephone interview by author, August 7, 2011.

73. Bill Blake, "WikiLeaks, the Pearl Harbor of the 21st Century," eDocument Sciences LLC, December 6, 2010, http://edocumentsciences.com/wikileaks-the-pearl-harbor-of-the-21st-century.

74. VaporStream, www.vaporstream.com (accessed December 9, 2013).

75. Ibid.

76. Ibid.

77. NIST, "Federal Information Processing Standards Publication," FIPS PUB 186-3, issued June 2009, http://csrc.nist.gov/publications/fips/fips186-3/fips_186-3.pdf (accessed August 15, 2011). FIPS Publication 186-3 (dated June 2009), was superseded on July 19, 2013 and is provided here only for historical purposes. For the most current revision of this publication, see: http://csrc.nist.gov/publications/PubsFIPS.html.

78. Doug Miles, AIIM White Paper, "Digital Signatures – Making the Business Case," https://www.docusign.com/partner/sharepoint-online-for-cosign-central.

79. Computer Desktop Encyclopedia, www.computerlanguage.com (accessed March 30, 2012).

80. Doug Miles, AIIM White Paper, "Digital Signatures—Making the Business Case.

81. Ibid.

82. Ibid.

83. Ibid.

84. Ari Ruppin, e-mail to author, March 20, 2011.

85. "Global Enterprise Data Loss Prevention Market (2018–2023): Industry Trends, Opportunities and Forecasts - CAGR to Grow at 16.28% - ResearchAndMarkets.com," https://www.businesswire.com/news/home/20180305005642/en/Global-Enterprise-Data-Loss-Prevention-Market-2018-2023

86. Fred Donovan, "Gartner: Enterprise Content-Aware Data Loss Prevention Market to Reach $670 Million This Year," February 7, 2013, www.fierceenterprisecommunications.com/story/gartner-enter-prise-content-aware-data-loss-prevention-market-reach-670-mill/2013-02-07.

87. Data Loss Prevention Experts, "DLP Product Guide for RSA Conference Expo 2011," January 17, 2011,

www.dlpexperts.com/dlpxblog/2011/1/17/dlp-product-guide-for-rsa-conference-expo-2011.html.

88. Ibid.

89. Ibid.

90. Ibid.

91. Peter Abatan, "Who Should Be Blamed for a Data Breach?" Enterprise Digital Rights Management, February 2011, http://enterprisedrm.tumblr.com/post/1087100940/who-should-be-blamed-for-a-data-breach.

92. Peter Abatan, "Understanding Enterprise Rights Management," Enterprise Digital Rights Management, www.enterprisedrm.info/page/2 (accessed August 3, 2011).

93. Robert Smallwood, "Securing Documents in the WikiLeaks Era," May 28, 2011, www.kmworld.com/Articles/Editorial/Feature/Securing-documents-in-the-WikiLeaks-era-75642.aspx.

94. Oracle, "Oracle Information Rights Management 11g—Managing Information Everywhere It Is Stored and Used," Oracle White Paper, March 2010, https://www.oracle.com/technetwork/middleware/webcenter/content/irm-technical-whitepaper-134345.pdf.

95. Abatan, "Understanding Enterprise Rights Management," http://enterprisedrm.tumblr.com/page/3 (accessed December 9, 2013).

96. Ibid.

97. Ibid.

98. Ibid.

99. Ibid.

100. This discussion and quotes are from Peter Abatan's blog, "Preparing for Staff Layoffs/Resignations Where Confidential Information Is Concerned," Enterprise Digital Rights Management (which has been deleted).

101. Ibid.

102. This discussion and quotes are from Peter Abatan's blog, "Is Your Price List under Lock and Key?" Enterprise Digital Rights Management (which has been deleted).

103. This discussion and quotes are from Andrew Jaquith, "Own Nothing—Control Everything: Five Patterns for Securing Data on Devices You Don't Own," ComputerWeekly.com, September 8, 2010, www.computerweekly.com/Articles/2010/09/10/242661/Own-nothing-control-everything-five-patterns-for-securing-data-on-devices-you-dont.htm.

104. This discussion and quotes are from Charlie Pulfer, "Document Labeling in SharePoint," September 13, 2009

105. Ibid.

106. Nancy Flynn, *The E-Policy Handbook: Rules and Best Practices to Safely Manage Your Company's E-Mail, Blogs, Social Networking, and Other Electronic Communication Tools*, 2nd ed. (New York: AMACOM, 2009), 57.

107. Ibid., 68–70.

* Portions of this chapter are adapted from Chapters 11 and 12, Robert F. Smallwood, *Safeguarding Critical E-Documents: Implementing a Program for Securing Confidential Information Assets*, © John Wiley & Sons, Inc., 2012. Reproduced with permission of John Wiley & Sons, Inc.

PART FOUR

Information Governance for Delivery Platforms

CHAPTER 12
Information Governance for E-Mail and Instant Messaging*

E-mail is a major area of focus for information governance (IG) efforts: it is the most common business software application and the backbone of business communications today. Also, e-mail is the leading piece of evidence requested during the discovery phase of civil trials, so it is critically important to implement IG measures for e-mail communications.

Employees utilize e-mail all day, including during their personal time, sometimes mixing business and personal use of e-mail. Between 2014 and 2018, the average office worker received an average of 90 e-mail messages daily, and sent 40.[1] Social media use has skyrocketed in recent years and actually has surpassed e-mail for personal use, but the fact remains that in business, knowledge workers rely on e-mail for almost all communications, *including those of a sensitive nature*. In one survey of corporate e-mail users worldwide, nearly two-thirds stated that e-mail was their favorite form of business communication, surpassing not only social media but also telephone and in-person contact.[2]

These e-mail communications may contain discoverable information in litigation, and a percentage of them will be declared formal business records. E-mail often contains records, such as financial spreadsheets and reports, product price lists, marketing plans, competitive analyses, safety data, recruitment and salary details, progressing contract negotiations, and other information that may be considered as constituting a business record.

E-mail systems a point of vulnerability for organizations, as over 90% of cyber attacks start with a **phishing** e-mail.[3] Phishing is a social engineering technique that hackers use to pose as someone the recipient knows or works with, to try to get the recipients to reveal confidential information and passwords. In 2017, Google and Facebook were scammed out of $100 million by a hacker by sending fake invoices (although later he was caught).[4] E-mail systems can be hacked, monitored, and compromised and cause far-reaching damage to a victimized organization. The damage may occur slowly and go undetected while information assets—and business value—are eroded.

Employees Regularly Expose Organizations to E-Mail Risk

One global e-mail survey, commissioned by a leading hosted e-mail services provider, found that nearly 80% of all employees send work e-mail to and from their personal accounts, and 20% do so regularly, which means that critical information assets are exposed to uncontrolled security risks.[5]

"Awareness of the security risks this behavior poses does not act as a deterrent" (emphasis added). Over 70% of people questioned recognize that there is an additional risk in sending work documents outside the corporate e-mail environment, but almost half of "these same respondents feel it is acceptable to send work e-mails and documents to personal e-mail accounts anyway." According to the survey, the reasons for using personal e-mail accounts for work purposes range from working on documents remotely (71%), to sending files that are too big for the company mailbox (21%), taking documents with them when they leave a company (18%), and those who simply do not want to carry a laptop home (9%). The top two frustrations users had with work e-mail were restrictions on mailbox size, which has a negative impact on e-mail management, and the inability to send large attachments. This second issue often forces workers to use a personal account to send and receive necessary files. If size limits are imposed on mailboxes and attachments, companies must provide a secure alternative to file storage and transfer. Otherwise, employees are pushed into risking corporate information assets via personal e-mail. This scenario not only complicates things for e-mail administrators but has serious legal and regulatory implications. Clearly, as stated by tech blogger Paul Mah in his "E-mail Admin" blog, "e-mail retention and archival becomes an impossible task when e-mails are routed in a haphazard manner via personal accounts."[6]

This means that security, privacy, and records management issues must be addressed by first creating IG policies to control and manage the use of e-mail. These policies can utilize the e-mail system's included security features and also employ additional monitoring and security technologies where needed.

The e-mail survey also found an overall lack of clear e-mail policies and weak communication of existing guidelines. *This means a lack of IG*. Nearly half of the respondents stated either that their company had no e-mail policy or that they were unaware of one. Among those aware of a corporate e-mail policy, 4 in 10 think it could be communicated better. Among companies that have a policy, most (88%) deal with the appropriate use of e-mail as a business tool, but less than one-third (30%) address e-mail retention from a security standpoint.

Generally, employees are aware that sending work documents outside of their corporate network is unsafe, yet they continue to do so. It is abundantly clear that *e-mail policies have to be updated and upgraded to accommodate and manage the increasingly sophisticated and computer-savvy generation* of users who are able to find ways to work around corporate e-mail restrictions. (These users have been dubbed *Generation Gmail*.) In addition, new e-mail monitoring and security technologies need to be deployed to counter this risky practice, which exposes information assets to prying eyes or malicious attacks.

E-Mail Polices Should Be Realistic and Technology Agnostic

E-mail policies as part of your IG program must not be too restrictive. It may be tempting to include catchall policies that attempt to tamp down user behavior, but such efforts cannot succeed.[7] An important step is consulting with stakeholders to understand their usage patterns and needs and then going through a series of drafts of the policy, allowing for input. It may be determined that some exceptions and changes in technologies need to be factored in and that some additional technology is needed to accommodate users while keeping information assets safer and meeting compliance and legal demands. Specifics of these policies and tools should be progressively tightened on a regular basis as the process moves forward.

These new IG guidelines and policies need to refer to technology in a generic sense—a "technology-neutral" sense—rather than specifying proprietary software programs or features.[8] That is to say, they should be written so that they are *not* in need of revision as soon as new technologies are deployed.

Developing organization-wide IG policies is time consuming and expensive; they are a defensive measure that does not produce revenue, so managers, pressed for performance, often relegate policy making to the low-priority list. Certainly, it is a tedious, difficult task, so organizations should aim to develop policies that are flexible enough to stand the test of time. But it is also necessary to establish a review process to periodically revise policies to accommodate changes in the business environment, the law, and technology.

Here is an example of a technology-agnostic policy directive:

> All confidential information must be encrypted before being transmitted over the Internet.

This statement does not specify the technology to be used, or the mode of transmission. The policy is neutral enough to cover not only e-mail and instant messaging (IM) but also social media, cloud computing, enterprise synch and share (EFSS), mobile computing, and other means of communication. The policy also does not specify the method or brand of the encryption technology, so the organization can select the best method and technology available in the future without adapting the policy.[9]

E-Record Retention: Fundamentally a Legal Issue

Considering the massive volume of e-mail exchanged in business today, most e-mail messages do not rise to the level of being formal business records. But many of them do and are subject to IG, regulatory compliance, and legal requirements for maintaining and producing business records. And *all* of them are potentially discoverable during litigation.

Although often lumped in with other information technology (IT) concerns, *the retention of e-mail and other e-records is ultimately a legal issue*. Other departments, including records management, compliance, and business units, should certainly have input and should work to assist the legal team to record retention challenges and archiving solutions. But e-mail and e-record retention is "fundamentally a *legal* issue," particularly for public or highly regulated companies. According to Nancy Flynn of the ePolicy Institute, "It is essential for the organization's legal department to take the lead in determining *precisely* which types of e-mail messages will be preserved, *exactly* how and where data will be stored, and *specifically when*—if ever—electronically stored information [ESI] will be deleted."[10]

> Managing e-records is primarily a legal issue, especially for public and heavily regulated companies.

Since they are often shot out in the heat of battle, many times e-mail messages are evidence of a smoking gun in lawsuits and investigations. In fact, they are the most requested type of evidence in civil litigation today. The content and timing of e-mail messages can provide exonerating information, too.

Perhaps the most famous scandal involving use of personal e-mail for business was the controversy over Hillary Clinton's periodic use of private e-mail for official government business during her term as Secretary of State, which became a major issue it the 2016 presidential election.[11] In 2018, Ivanka Trump's use of her personal e-mail account for official government matters also came into question. Also, in January 2010, a US House of Representatives committee probing bailout deals subpoenaed the Federal Reserve Bank of New York for e-mail and other correspondence from Treasury Secretary Timothy Geithner (former president of the New York Federal Reserve Bank) and other officials. The House Oversight and Government Reform Committee was in the process of examining New York Fed decisions that funneled billions of dollars to big banks, including Goldman Sachs Group and Morgan Stanley.[12]

These are examples of how crucial it is for employees to use business e-mail for conducting business, and that e-mail messages can play a major role in legal investigations and in reconstructing events and motives for legal purposes.

Preserve E-Mail Integrity and Admissibility with Automatic Archiving

Most users are not aware that e-mail contents and characteristics can be changed—"and rendered legally invalid"—by anyone with malicious motives, including those who are essentially "covering their tracks." Not only can the content be edited, but metadata that includes such information as the time, date, and total number of characters in the message can also be changed retroactively.[13]

To offset this risk and ensure that **spoliation** (i.e., the loss of proven authenticity of an e-mail) does not occur, *all messages, both inbound and outbound, should be captured and archived automatically and in real time*. This preserves legal validity and forensic compliance. Additionally, e-mail should be indexed to facilitate the searching process, and all messages should be secured in a single location. With these measures, e-mail records can be assured to be authentic and reliable.

E-Mail Archiving Rationale: Compliance, Legal, and Business Reasons

There are good reasons to archive e-mail and retain it according to a specific retention schedule that follows your organization's IG policies. Having a handle on managing voluminous e-mail archives translates to being able to effectively and rapidly search and retrieve exactly the right messages, which can provide a significant legal advantage. It gives your legal team more and better information and more time to figure out how to leverage it in legal strategy sessions. This means the odds are tipped in your organization's favor in the inevitable litigation arena. Your legal opponent may be driven to settle a weak claim when confronted with indisputable e-mail evidence, and, in fact, "e-mail often produces supportive evidence that may help 'save the day' by providing valuable legal proof" of innocence.[14] This evidence may stop frivolous lawsuits in their tracks. Further, reliable e-mail evidence also can curtail lengthy and expensive lawsuits, and prevail. And if your company is public, Sarbanes-Oxley regulations require the archiving of e-mail.

Don't Confuse E-Mail Archiving with Backup

All backups are not created equal. *There is a big difference between traditional system backups and specialized e-mail archiving software.*

Backups are huge dumps to mass storage, where the data is stored sequentially and not compressed or indexed.[15] It is impossible to search backups except by date, and even doing that would mean combing through troves of raw, nonindexed data.

The chief executive may not be aware of it, but without true e-mail archiving, system administrators could spend long nights loading old tapes and churning out volumes of data, and legal teams will bill hourly for manual searches through troves of data. This compromises your enterprise's legal position and not only increases raw costs but also leads to less capable and informed legal representation. According to one study, fully one-third of IT managers state they would have difficulty producing an e-mail that is more than one year old. *"A backup system is no substitute for automatic archiving technology"*[16] (emphasis added).

No Personal Archiving in the Workplace

Employees are naturally going to want to back up their most important files, just as they probably do at home. But for an overall IG information-security program to be effective, personal archiving at work must be prohibited. This underground archiving results in hidden shadow files and is time consuming and risky. According to Flynn, *"Self-managed e-mail can result in the deletion of electronic records, alteration of e-mail evidence, time-consuming searches for back-up tapes, and failure to comply with legal discovery demands"* (emphasis added). Also, users may compromise formal electronic records, or they may work from unofficial records, which therefore by definition might be inaccurate or out-of-date, posing compliance and legal ramifications.[17]

Are All E-Mails Records?

Are e-mail messages records? This question has been debated for years. The short answer is no, not all e-mail messages constitute a record. But how do you determine whether certain messages are a business record or not? The general answer is that a record documents a transaction or business-related event that may have legal ramifications or historic value. Most important are business activities that may relate to compliance requirements or those that could possibly come into dispute in litigation. Particular consideration should be given to financial transactions of any type.

Certainly evidence that required governance oversight or compliance activities have been completed needs to be documented and becomes a business record. Also, business transactions, where there is an exchange of money or the equivalent in goods or services are also business records. Today, these transactions are often documented by a quick e-mail. And, of course, any contracts (and any progressively developed or edited versions) that are exchanged through e-mail become business records.

The form or format of a potential record is irrelevant in determining whether it should be classified as a business record. For instance, if a meeting of the board of directors is recorded by a digital video recorder and saved to DVD, it constitutes a record. If photographs are taken of a groundbreaking ceremony for a new manufacturing plant, the photos are records, too. If the company's founders tape-recorded a message to future generations of management on reel-to-reel tape, it is a record also, since it has historical value. But most records are going to be in the form of paper, microfilm, or an electronic document.

Here are three guidelines for determining whether an e-mail message should be considered a business record:

1. The e-mail documents a transaction or the progress toward an ultimate transaction where anything of value is exchanged between two or more parties. All parts or characteristics of the transaction, including who (the parties to it), what, when, how much, and the composition of its components are parts of the transaction. Often seemingly minor

parts of a transaction are found buried within an e-mail message. One example would be a last-minute discount offered by a supplier based on an order being placed or delivery being made within a specified time frame.

2. The e-mail documents or provides support of a business activity occurring that pertains to internal corporate governance policies or compliance to externally mandated regulations.
3. The e-mail message documents other business activities that may possibly be disputed in the future, whether it ultimately involves litigation or not. (Most business disputes actually are resolved without litigation, provided that proof of your organization's position can be shown.) For instance, your supplier may dispute the discount you take that was offered in an e-mail message and, once you forward the e-mail thread to the supplier, it acquiesces.[18]

Destructive Retention of E-Mail

Destructive retention is an approach to e-mail archiving where e-mail messages are retained for a limited time (say, 90 days or six months), followed by their permanent manual or automatic deletion of messages from the company's network, so long as there is no litigation hold or the e-mail has not been declared a record, in accordance with IG and records management policies. Implementing this as a policy may shield the enterprise from retaining potentially libelous or litigious e-mail that is not a formal business record (e.g. off-color jokes or other personnel violations).

For heavily regulated industries, such as health care, energy, and financial services, organizations may need to archive e-mail for longer periods of time.

Bucket Approach

Some companies attempt to enforce retention periods by giving users guidelines and the option of dragging and dropping e-mail messages into retention buckets or varying lengths, such as one, three, five, and seven years. The problem with this approach is that, in reality, users do not want to spend time considering retention periods and classifying messages, so although many employees make good faith efforts, others simply drag the messages they want to keep into the seven-year bucket to simplify the process and make sure the e-mail will be there if they need it. The issue with this approach is consistency, and with inconsistent retention of e-mail messages, proving that the organization is adhering to its retention schedule becomes challenging and these practices may be ruled legally indefensible in court.

Instant Messaging

Instant messaging (IM) use in enterprises has proliferated—despite the fact that frequently proper policies, controls, and security measures are not in place to prevent e-document and data loss. There are a variety of threats to IM use that enterprises must defend against to keep their information assets secure.

The first basic IM systems, which came into use in the mid-1960s, had real-time text capabilities for routing messages to users logged on to the same mainframe computer. Early chat systems, such as AOL Instant Messenger, have been in use since the late 1980s, but true IM systems that included buddy list features appeared on the scene in the mid-1990s, followed by the release of Yahoo! and Microsoft IM systems. The use of these personal IM products in the workplace has created new security risks.[19]

More secure enterprise instant messaging (EIM) products can be deployed. Leading EIM installed systems include IBM Sametime, Slack, Workplace by Facebook, Jabber, SkyHistory, and Zoho Cliq/Chat. In the financial sector, Bloomberg Messaging and Reuters Messaging are leading platforms.

By the year 2000, it was estimated that nearly 250 million people worldwide were making use of IM, and today estimates are that more than 2 billion people use IM, with the addition of hundreds of millions of users in China.

As with many technologies, IM became popular first for personal use, then crept into the workplace—and exploded. IM is seen as a quicker and more efficient way to communicate short messages than engaging in a telephone conversation or going through rounds of sending and receiving endless e-mail messages. *The problem with IM is that many organizations are blind to the fact that their employees are going to use it one way or another*, sometimes for short personal conversations outside the organization. If unchecked, such messaging exposes the organization to a myriad of risks and gives hackers another way to compromise confidential information assets.

Best Practices for Business IM Use

Employing best practices for enterprise IM use can help mitigate its security risks while helping to capitalize on the business agility and velocity benefits IM can provide. Best practices must be built into IG policies governing the use of IM, although “the specifics of these best practices must be tailored for each organization's unique needs.”

A methodology for forming IM-specific IG policies and implementing more secure use of IM must begin with surveying and documenting the proliferation of IM use in the organization. It should also discover how and why users are relying on IM—perhaps there is a shortcoming with their available IT tools and IM is a work-around.

Documenting IM use in the organization is the first step in building IG policies to govern its use. Those policies must be tailored to the organization and its IM use.

Typically, executives will deny there is much use of IM and that if it is being used, its impact is not worth worrying about. Also, getting users to come clean about their IM use may be difficult, since this may involve personal conversations and violations of corporate policy. A survey is a good place to start, but more sophisticated network monitoring tools need to be used to factually discover what IM systems are actually in use.

Once this discovery process has concluded and the use of IM is mapped out, the IG team or steering committee must create or update policies to: decide which IM systems it will allow to be used, how, when, and by whom; decide what restrictions or safeguards must be imposed; and create guidelines as to appropriate use and content. As a part of an overall IG effort, Quest Software determined that a successful IM policy will:

- *Clearly and explicitly explain the organization's instant messaging objectives*. Users should know why the organization permits IM and how it is expected to be used.
- *Define expectations of privacy*. Users should be made aware that the organization has the right to monitor and log all IM sessions for corporate compliance, safety, and security reasons.
- *Detail acceptable and unacceptable uses*. An exhaustive list of permitted and forbidden activities may not be necessary, but specific examples are helpful in establishing a framework of IM behaviors for users.
- *Detail content and contact restrictions (if any)*. Most organizations will want to limit the amount of idle IM chat that may occur with family, friends, and other nonbusiness related contacts. There may also be additional issues related to information confidentiality and privacy. Some businesses may choose to block the distribution of certain types of information via live IM chat session or file transfer.
- *Define consequences for violations of the policy*. Users should be advised of the consequences of policy violations. Generally these should be aligned with the company's personnel and acceptable use policies.

The use of a standard disclaimer, to be inserted into all users' IM sessions, can remind employees of appropriate IM use and that all chat sessions are being monitored and archived, and can be used in court or compliance hearings.

The next major step is to work with the IT staff to find the best and most appropriate security and network monitoring tools, given the computing environment. Alternatives must be researched, selected, and deployed. In this research and selection process, it is best to start with at least an informal survey of enterprises within the same industry to attempt to learn what has worked best for them.

The key to any compliance effort or legal action will be ensuring that IM records are true and authentic, so the exact, unaltered archiving of IM messages along with associated metadata should be implemented in real time. This is the only way to preserve business records that may be needed in the future. But in addition, a policy for deleting IM messages after a period of time, so long as they are not declared business records, must be formulated.

Records of IM use must be captured in real time and preserved to ensure they are reliable and accurate.

IG requires that these policies and practices not be static; rather, they must be regularly revisited and updated to reflect changes in technology and legal requirements and to address any shortcoming or failure of the IG policies or technologies deployed.

Technology to Monitor IM

Today, it has been estimated that as much as 80% of all IM used by corporate employees comes from free IM providers like Yahoo!, MSN, or AOL. These programs are also the least secure. Messages using these IM platforms can fly around the Internet unprotected. Any monitoring technology implemented must have the capability to apply and enforce established IM use policies by constantly monitoring Internet traffic to discover IM conversations. Traffic containing certain keywords can be monitored or blocked, and chat sessions between forbidden users (e.g. those who are party to a lawsuit) can be stopped before they start. But this all necessarily starts with IG and policy formulation.

Tips for Safer IM

Organizations should assume that IM is being used, whether they have sanctioned it or not. And that may not be a bad thing—employees may have found a reasonable business use for which IM is expedient and effective. So management should not rush to ban its use in a knee-jerk reaction. Here are some tips for safer use of corporate IM:

- Just as e-mail attachments and embedded links are suspect and can contain malicious executable files, *beware of IM attachments* too. The same rules governing e-mail use apply to IM, in that employees should never open attachments from people they do not know. Even if they do know them, with phishing and social engineering scams, these

attachments should first be scanned for malware using antivirus tools.

- *Do not divulge any more personal information than is necessary*. This comes into play even when creating screen names—so the naming convention for IM screen names must be standardized for the enterprise. Microsoft advises, "Your screen name should not provide or allude to personal information. For example, use a nickname such as SoccerFan instead of BaltimoreJenny."[20]
- *Keep IM screen names private*; treat them as another information asset that needs to be protected to reduce unwanted IM requests, phishing, or spam (actually *spim*, in IM parlance).
- *Prohibit transmission of confidential corporate information*. It is fine to set up a meeting with auditors, but do not attach and route the latest financial report through unsecured IM.
- *Restrict IM contacts to known business colleagues*. If personal contacts are allowed for emergencies, limit personal use for everyday communication. In other words, do not get into a long personal IM conversation with a spouse or teenager while at work. Remember, these conversations are going to be monitored and archived.
- *Use caution when displaying default messages when you are unavailable or away*. Details such as where an employee is going to have lunch or where their child is being picked up from school may expose the organization to liability if a hacker takes the information and uses it for criminal purposes. Employees may be unknowingly putting themselves in harm's way by giving out too much personal information.
- *Ensure that IM policies are being enforced by utilizing IM monitoring and filtering tools and by archiving messages in real time* for a future verifiable record, should it be needed.
- *Conduct an IM usage policy review at least annually*; more often in the early stages of policy development.

Team and Channel Messaging Solutions Emerge

In August, 2013, Slack technologies released a cloud-based business team messaging and collaboration tool which helps project teams keep track of progress and allows for communication in teams, subteams, and one-on-one, both real-time and asynchronously. Slack became a very popular alternative to traditional e-mail and IM technologies. Other competitors soon followed.

Slack and other business messaging applications such as Flock, Zoho Cliq, Ryver, Workplace by Facebook, and Microsoft Teams quickly gained adoption and became the de facto way that business teams communicate on a daily basis. Most of these offer a free entry-level version.[21]

> New collaborative business messaging solutions have emerged that are more efficient than e-mail, categorize messages into channels, allow for audio and video calls, and have robust search capabilities.

These business messaging solutions address the overload of e-mail and IM by keeping the communications in segregated channels for ease of organizing and tracking project communications. In addition to messaging, most of these messaging solutions offer screen sharing. audio calls, video calls, and robust search capabilities. All content, including conversations, files, and users, is searchable in most of this class of messaging solution, When implemented and used properly, these new business messaging applications can actually impact and transform the corporate culture while improving productivity.

Some of the advantages of this new feature-rich business messaging approach include: brevity of messages, opt-in discussions so time is not wasted, persistent ongoing conversations (in contrast to e-mail threads), and even notifications based on user-defined keywords, collaborative document editing, team calendars, and task management tools.

CHAPTER SUMMARY: KEY POINTS

- E-mail is a critical area for IG implementation, as it is a ubiquitous business communication tool and the leading piece of evidence requested at civil trials.
- Nearly 80% of all employees send work e-mail messages to and from their personal e-mail accounts, which exposes critical information assets to uncontrolled security risks.
- Meeting e-mail retention and archival requirements becomes an impossible task when e-mail messages are routed in a haphazard manner via personal accounts.
- In developing e-mail policies, an important step is consulting with stakeholders.
- E-mail policies must not be too restrictive or tied to a specific technology. They should be flexible enough to accommodate changes in technology and should be reviewed and updated regularly.
- Not all e-mail messages constitute a business record.
- Not all e-mail rises to the level of admissible legal evidence. Certain conditions must be met.
- Automatic archiving protects the integrity of e-mail for legal purposes.
- Instant messaging use in business and the public sector has become widespread, despite the fact that often few controls or security measures are in place.
- Typically as much as 80% of all IM use in corporations today is over free public networks, which heightens security concerns.
- IM monitoring and management technology provides the crucial components that enable the organization to fully implement best practices for business IM.
- Enterprise IM systems provide a greater level of security than IM from free services.
- Regular analysis and modification (if necessary) of business IM policies and practices will help organizations leverage the maximum benefit from the technology.
- Records of IM use must be captured in real time and preserved to ensure they are reliable and accurate.
- New collaborative business messaging solutions have emerged that are more efficient than e-mail, categorize messages into channels, allow for audio and video calls, and have robust search capabilities.

Notes

1. "How Many Emails Does the Average Person Receive Per Day?" https://www.templafy.com/blog/how-many-emails-are-sent-every-day-top-email-statistics-your-business-needs-to-know/ (accessed December 14, 2018).
2. "Research Finds That Restrictive Email Policies Are Creating Hidden Security Risks for Businesses," *BusinessWire*, March 9, 2011, www.businesswire.com/news/home/20110309005960/en/Research-Finds-Restrictive-Email-Policies-Creating-Hidden.
3. "91% of Cyber Attacks Stat with a Phishing Email," https://www.darkreading.com/endpoint/91-of-cyberattacks-start-with-a-phishing-email/d/d-id/1327704 (accessed December 14, 2018).
4. "This $100 Million Email Scam Tripped Up Two Big U.S. Companies," https://www.cnet.com/uk/news/100-million-email-scam-phishing-cybercrime-lithuania/#pt0-9210 (accessed December 14, 2018).
5. Quotes from this survey are from "Research Finds That Restrictive Email Policies Are Creating Hidden Security Risks for Businesses."
6. Paul Mah, "How to Reduce the Email Security Risks to Your Business," *EmailAdmin*, March 10, 2011, www.theemailadmin.com/2011/03/how-to-reduce-the-email-security-risks-to-your-business/.
7. Blair Kahn, *Information Nation: Seven Keys to Information Management Compliance* (Silver Spring, MD: AIIM International, 2004), 98–99.
8. Ibid, 95–96.
9. Ibid.
10. Nancy Flynn, *The E-Policy Handbook: Rules and Best Practices to Safely Manage Your Company's E-Mail, Blogs, Social Networking, and Other Electronic Communication Tools*, 2nd ed. (New York: AMACOM, 2009), 20.
11. "Hillary Clinton Emails: What's It All About?" BBC News, November 6, 2016, https://www.bbc.com/news/world-us-canada-31806907.

12. Hugh Son and Andrew Frye, "Geithner's E-mails, Phone Logs Subpoenaed by House (update3)," January 13, 2010, www.bloomberg.com/apps/news?pid=newsarchive&sid=aGzbhrSxFlXw.

13. Flynn, *E-Policy Handbook*, 37.

14. Flynn, *E-Policy Handbook*, 41.

15. Nancy Flynn and Randolph Kahn, *Email Rules, A Business Guide to Managing Policies, Security, and Legal Issues for E-Mail and Digital Communication* (New York: AMACOM, 2003), 81–82.

16. Flynn, *The E-Policy Handbook*, 41.

17. Ibid., 43.

18. Robert F. Smallwood, *Taming the Email Tiger: Email Management for Compliance, Governance, & Litigation Readiness* (New Orleans, LA: Bacchus Business Books, 2008).

19. This discussion is based on Quest Software White Paper, "Best Practices in Instant Messaging Management," October 2008, http://media.govtech.net/Digital_Communities/Quest%20Software/Best_Practices_in_Instant_Messaging_Management.pdf, p. 5.

20. M. Adeel Ansari, "10 Tips for Safer IM Instant Messaging," July 6, 2008, http://adeelansari.wordpress.com/tag/safer-im-instant-messaging/ (accessed December 17, 2018).

21. "The Best Business Messaging Apps of 2018," *PC Magazine,* July 27, 2018, https://www.pcmag.com/roundup/355674/the-best-team-messaging-apps.

* Portions of this chapter are adapted from Chapter 11, Robert F. Smallwood, *Managing Electronic Records: Methods, Best Practices, and Technologies*, © John Wiley & Sons, Inc., 2013. Reproduced with permission of John Wiley & Sons, Inc.

CHAPTER 13
Information Governance for Social Media*

By Dr. Patricia Franks and Robert Smallwood

Information is the lifeblood of every organization, and an increasing volume of information today is created and exchanged through the use of social networks and Web 2.0 tools like blogs, microblogs, and wikis.

Corporations use public social media technology to create a visible brand, strengthen relations with current customers while attracting new connections and clients, highlight their products and services, and gather intelligence that can be used in decision making.

Governments use public social media technologies to consult with and engage citizens, provide services, and keep pace with fast-moving events (e.g. natural disasters).

Both types of enterprises also benefit from the use of internal social media solutions that facilitate communication and collaboration, improve employee engagement, and boost productivity and efficiency.

Content created through or posted to these new social media platforms must be managed, monitored, and, quite often, archived. Content that meets the organization's definition of a record (i.e. documents business activities) must be retained in accordance with the organization's records retention and disposition policy.

Too often social media content is not managed by information governance (IG) policies or monitored with controls that ensure protection of the brand and critical information assets and preservation of business records.

Types of Social Media in Web 2.0

The term "Web 2.0" was coined to characterize the move from static Web sites that passively provided information to consumers to more participative, interactive, collaborative, and user-oriented Web sites and Web applications that allow for input, discussion, and sharing. Users can add content, increasing the value of the Web site or service. Examples include blogs and Web pages containing podcasts (digital media, usually audio) where readers can post comments or pose questions; wikis that hyperlink to related information to create a knowledge base that shows interrelationships and allow users to add content; and RSS (really simple syndication) feeds that provide a stream of fresh content to the user or consumer.

Web 2.0 is the term used to describe the second generation of the World Wide Web, which is comprised of a combination of technologies that allow consumers of Web content to participate, collaborate, and share information online. The improved functionality reflects consumer needs and preferences that surfaced as a result of increased use of the Web for daily information and communications.

Social media sites like LinkedIn, Twitter, and Facebook encourage social interactions by allowing users to create their own close network of business associates or friends—essentially a hand-picked audience—and to post their own content in the form of comments, links, photos, videos, and so forth. Others in their social network may view, forward, share, organize, and comment on this content.[1]

Web 2.0 and social media platforms began as outward-facing, public Web services that could link users from around the world. Subsequently, businesses discovered that social media technology could also be leveraged for internal use in various ways, such as by creating a directory and network of subject matter experts that users can search when working on special projects or by sending out microblog messages to keep their workforce informed. These internal social networks may be extended to include external stakeholders, such as suppliers and customers, in a controlled environment. A number of platform and software options exist for enterprise social media development and use.

Today one would be hard pressed to find a business that does not incorporate social media into its marketing strategy. Because not all social media platforms are created equal, careful consideration must be given to the social media platform—whether the most popular or most obscure—that will bring the right message to the intended target audience. Social media tools can be categorized in a number of ways. Five types of social media platforms to be considered for any marketing campaign are: social networking sites, image sharing and messaging sites, video sharing sites, social blogging, and social community and discussion sites. Examples of each type are shown in Table 13.1 based on an article published on SproutSocial based on an article written by Brent Barnhart.[2]

Table 13.1 Social Media Categories and Examples of Tools Commonly Used by Marketing Professionals

Social Networking Sites provide a platform for users to connect with others.	
Facebook	Facebook ads reach a diverse global population comprised of almost any age group and level of income. Both personal and professional connections are possible.
LinkedIn	LinkedIn is considered the social network of choice for industry professionals. Members can connect with other professionals and market their talents to CEOs and executives.
Twitter	Twitter users connect with almost anyone in 280 characters or less instantaneously. Twitter is the best social media channel for customer engagement according to 47% of marketers surveyed.[3] Mobile marketers find this a useful tool to promote their brand.
Image Sharing and Messaging Sites provide platforms that allow users to curate and promote content	
Instagram	Instagram can be used to create connections with customers by encouraging user-generated visual content. An estimated 71% of US businesses use Instagram,[4] and 80% of users say they follow at least one business on the app.
Snapchat	Snapchat is popular with millennials; 71% of users are under 34 years of age, 45% are between 18 and 24. To date, bigger brands primarily buy advertising on Snapchat; however, smaller firms can gain attention by highlighting industry connections and interactions, co-marketing, and previewing products or demo releases.
Pinterest	Pinterest is a social bookmarking site that acts as a digital pinboard for users. An advantage of developing high-quality Pins is the fact they last forever. Two-thirds of all content on Pinterest comes from businesses, and Pinterest, at 5%, is second to only Facebook, at 25%, when it comes to driving "referral" traffic to websites.[5]
Video Sharing Sites are optimal for educating an audience, require little effort for viewers to grasp, and have a low barrier to entry for business.	
YouTube	YouTube provides a platform for businesses to educate, entertain and promote their brand. While commercials are becoming more common, marketers can create videos to promote products and provide tutorials.
Vimeo	Vimeo was the first video site to support HD video uploads and is the platform favored by professionals—filmmakers, editors, and marketers—to share their work.
Social Blogging sites allow marketers to engage audiences via written content	
Medium	Medium is a publishing platform used by businesses seeking to build their brand and expand the reach of their written content. It's a good way to become recognized as an authority on a particular subject and to reach new readers.
Tumblr	Tumblr supports image, video, and text content. It's not the social networking and blogging platform of choice by most marketers today but can be used to target a niche audience. As of January 2018, 43% of internet users between 18 and 24 years of age used Tumblr.[6]
Social Community and Discussion Sites allow marketers to keep a pulse on the industry and establish themselves as trusted resources for their audience.	
Reddit	Reddit, curated and moderated by users, is considered one of the most popular news sources in the world. It serves as a hub for online discussions and a platform for niche communities in the form of subreddits.[7]
Quora	Quora is a community centered on answering questions. Users customize question streams and profiles based on interests and then can upvote and downvote answers as they wish.
Yahoo! Answers	Yahoo! Answers is an informal version of Quora, allowing those asking and answering questions to rate answers. Rather than foster community, Yahoo! Answers provides fast, simple answers.

A blog post published by HootSuite,[8] a social media management system, provides 10 categories—five similar to the previous categories and five that introduce additional types of tools:

- *Consumer review networks* that help people find, review, and share information about brands, products, and services. Examples of customer review networks are Yelp, Zomato, and Trip Advisor. Positive reviews on these sites bring credibility to business claims.
- *Social shopping networks* that help people spot trends, follow brands, share finds, and make purchases. Examples include Polyvore, Etsy, and Fancy. Brands offered on these sites help the business build awareness, increase engagement, and sell products to new customers.
- *Interest-based networks* that help people connect with others with the same interests or hobbies. Examples are Goodreads, Houzz, and Last.fm. Businesses that can identify interest-based networks that attract their customers can design experiences for their niche audience and keep up with current trends.
- *Sharing economy networks* (also called collaborative economy networks) allow people to advertise, share, find, buy, sell, and trade products and services. Examples are Airbnb, Uber, and Taskrabbit.
- *Anonymous social networks* that allow users to gossip, vent, snoop, and unfortunately even bully. Examples are Whisper, Ask.fm, and After School. Businesses should *avoid these social networks*.

Government agencies on the federal, state, and local levels also use social media technology to help them interact with their audience. Although their options are similar to those available to business and industry, their strategies differ.

The US National Archives and Records Administration launched their first National Archives social media strategy in 2010. By 2018 more than 200 National Archives staff contributed to 130 social media accounts on 14 different platforms. The social media strategy

has evolved as well. The main focus of the use of social media today is to tell great stories, deepen engagement with the public, grow their audience, and cultivate a community of practice.[9]

In order to accomplish its goals, the US National Archives (NARA) employs social media tools categorized as shown in Table 13.2.

Table 13.2 Social Media Tools Used by NARA

Web Publishing Platforms to create, publish, and reuse content.	
Blog	*Narrations*, the National Blog of the US National Archives, run on WordPress.
Microblogs	Tumblr: NARA shares news and current events through Tumblr. Twitter: NARA employs Twitter for live tweeting of events, two-way conversations, and help with questions.
Mashups	NARA shares photos, videos, and audio recordings through 74 collections and 11 tours on HistoryPin. Each of the 2360 objects is pinned to a specific location on Google Maps.
Storify	NARA used this curation tool to chronologically arrange social media posts into 131 different stories. Storify closed May 16, 2018, underscoring the need to constantly evaluate and update your social media strategy.
Social Networking Platforms to provide interactions and collaboration among users.	
Facebook	NARA shares public news and events through Facebook. However, the National Archives website is considered the official source of information about the National Archives.
Google+	NARA uses this social networking service to share information about current and past events and to encourage comments from the public.
Image/Video/Idea Sharing Platforms to host and share content.	
Flickr	NARA uses Flickr to share photos aggregated into albums, complete with catalog descriptions that include creator(s), series, production date, access restrictions, contact(s), National Archives Identifier, Local Identifier, Persistent URL, and Scope and Content Note.
Giphy	NARA uses its Giphy channel to share animated gifs, which can be downloaded by the public.
Instagram	NARA uses Instagram to share photos and descriptions. Comments from the public are welcome.

Additional Social Media Categories

The categories used by business and industry and government will increase and fluctuate as the market matures and the companies providing the social media technologies and services expand, merge, are acquired, or die off. Additional categories and tools not included in the previous sections are show in Table 13.3.

There are certainly additional categories, and the categories will continue to grow. In addition, social media companies do not always fit neatly into one category. Applications (apps) for smartphones and tablets offer instant gratification and combine several functions. For example, Snapchat allows the sender to share an experience by snapping an image or video, adding a caption, and sending it to a friend.[10] The image, unless saved by the recipient, is visible only for the number of seconds set by the sender. The goal is to share a moment in time by sending a fleeting message. In 2017, Facebook's Instagram redesigned its messaging feature to also allow users to send disappearing photos and videos. The videos and photos can be viewed once before disappearing, and the sender will receive a message if the recipient takes a screenshot.[11]

Table 13.3 Additional Social Media Tools by Application Type

Category	Examples
Content curation	Curatq, Digg, Flipboard
Content sharing	LinkedIn Slideshare, Digg, Topix
Genealogy social platform	Family.me, Geneanet, MyHeritage
Photo sharing	Imgur, SmugMug, Photobucket
Social ad networks	Lifestreet, Adknowledge, Media6degrees, BurstMedia
Social analytics	Facebook Analytics, Mixpanel, Twitter Analytics, Webtrends
Social bookmarking	BibSonomy, Dribble, Folkd, StumbleUpon
Social business software	Disqus, Lithium, inSided, Hoop.la, IdeaScale, Jive, Mutt, Ning, Socious, Telligent, TWikI
Social brand engagement	Socialvibe, Post Intelligence (PI), Adly, Sharethrough
Social commerce platforms	Ecwid, Moontoast, Shop Tab (Facebook Store), Shopify Apps, StoreYa.com, Storenvy
Social community platforms	Hivebrite, Ning, Mixxt, Vanilla
Social data	GNIP, DataSift, RavenPack, Ushahidi
Social intelligence software	Netbase, PostRank, Google Analytics, Nuvi, Trendrr, Trackur, Visible Intelligence
Social promotion platforms	Ampsocial, Extole, Fanzila, PROMOJAM, SEMrush, Strutta, Votigo, Wishpond, Zuberance
Social marketing management suites	Falcon.io, Hootsuite, Spredfast, Hearsay Systems, SproutSocial, Sysmos
Social referral	Ambassador, friendbuy, LinkTrust, Mention Me, SocialReferral, Talkable
Social search and browsing	StumbleUpon, Tagboard, Talkwalker Social Search
Social scoring	Klout, Meltwater, PeerIndex

Social Media in the Enterprise

Public-facing social media integrates Internet-based applications, technology, social interaction, and content creation to enable communication, collaboration, and content sharing within and across subnetworks of millions of public users. Implementing tight security on these types of mass networks would likely slow response time and inhibit the user experience, and it may not provide a sufficient level of security to warrant the investment on the part of the social media provider.

While popular consumer-based technologies (Facebook, LinkedIn, YouTube, and Twitter)[12] top the list of most valuable social media technologies used in enterprises, *these services were not designed with the business in mind*. Enterprises that need tight security but wish to take advantage of the many benefits of social media use are increasingly implementing enterprise-wide social media solutions in addition to or in place of public-facing social media.

In the business world, Facebook-like social networking software is offered for private, closed networks with a finite number of users. In this computing environment, implementing security is more manageable and practical. Some services are cloud based; others operate internally behind the enterprise firewall; and some operate either way or in conjunction as hybrid architecture. From a technical standpoint, organizations are urged to position Enterprise Social Media as the hub that integrates with IT systems, business applications, collaborative tools, and digital platforms.[13]

> **Implementing security is more manageable and practical with enterprise social networking software.**

Enterprise social networking is being adopted by business and public-sector entities at a rapid rate. With the entry of *Generation Gmail* into the workforce, many of these initiatives took on an experimental, "cool" image. However, it is crucial to establish social media business objectives, to define time-limited metrics, and to measure progress.

Measuring the effectiveness of enterprise social network implementation (ROI) can be accomplished through either a qualitative (e.g. achievement of specific priorities and outcomes), quantitative (e.g. use of tools such as Stat Insight for Yammer), or blended approach (e.g. sentiment analysis revealing change in positive or negative attitudes over time). Promising work is underway by the UK firm ContentandCode in partnership with the Warwick Business School to test a maturity model based on the measurement of employee attitudes, beliefs, perception of corporate culture, and view of organizational structure and process in relation to five productivity factors: business agility; collaboration; employee engagement; innovation; and security, risk, and governance. The results of the survey can be used to measure the ESM level of maturity and serve as a benchmark against which to measure progress.[14]

Competitive value is the intangible benefit of being a market leader or industry innovator. But to keep that edge, companies need to continually scan the horizon for new technologies and services. Engaging in online conversations with customers and other stakeholders is the norm rather than the exception. One sign of a progressive-thinking organization is its ability to leverage social media technology to refine operations, improve customer services, and make employees' lives easier. An organization with a strong social media reputation likely will be better able to attract, recruit, and retain qualified, high-achieving employees.

Key Ways Social Media Is Different from E-Mail and Instant Messaging

Social media offers some of the same functionality as other communication and collaboration systems like e-mail and instant messaging (IM), yet its architecture and underlying assumptions are quite different.

When implementing enterprise versions of social media applications, a company may exert more control over the computing and networking environment through in-house implementation rather than outsourcing. Consumer-oriented social media applications, such as Facebook and Twitter, reside on application servers outside the enterprise controlled by third-party providers. This creates IG and records management (RM) challenges and poses legal risks.[15]

Some would say that social media is no longer an emerging technology. However, it is definitely a developing technology with standards, design, and architecture continually in flux. E-mail, on the other hand, has been stable and established for 15 to 20 years. E-mail is a mature technology set, meaning it is unlikely to change much. There are standard e-mail communications protocols, and the technology's use is pervasive and constant. So when e-mail IG policies are formed, less updating and fine-tuning is required over time. With social media, new features are being added, standards are nonexistent, privacy settings change overnight, and the legalese in terms of service agreements is continually modified to include new features and settings, which means that your social media policy must be more closely monitored and frequently fine-tuned.

> **Social media differs greatly from e-mail use. E-mail is mature and stable. Social media is not. These distinctions have important ramifications for IG policy development.**

E-mail, IM, and social media are all communication tools used to share content and collaborate, but social media also offers user interaction features, such as “Like” on Facebook or “retweet” (copying and posting a 280-character tweet) on Twitter, that bring attention to the content in the user's network and can be construed as an endorsement or rejection of content based on user opinions expressed and associated with the content.

Further confounding the organization's ability to control the social media environment is the fact that the social media sites are dynamic, with comments and opinions being published in real time. This is not true with e-mail and IM systems, which are more structured and stable.

Biggest Risks of Social Media

“According to IASCA, the second highest risk from the use of social media is to the firm when employees inadvertently distribute work-related information.”[16] This is especially true when working in highly regulated environments like financial institutions and banks, energy, schools and universities, and hospitals; in industries that engage in proprietary research and development; and even in the military.

Organizations that believe they can ban social media in order to avoid risks are mistaken. Prohibition of social media can result in social media use being driven underground. Employees accustomed to the ease of communicating and collaborating through social networks may turn to the use of personal devices and accounts outside the control of the organization. Even strict adherence to a nonuse policy can harm the organization's reputation, finances, ability to gather information that can be used to improve operations, and ability to remain competitive.

Once an organization decides it will engage in social media initiatives, it must identify different types of risks to initiate its IG effort in this area. In 2011, Chris Nerney of *Network World* cautioned enterprises to take precautions to prevent employees from putting the organization at risk. Two of the greatest social media security threats he identified remain problematic today. They are:

1. *Lack of an up-to-date social media policy*. Many organizations have social media policies in place, but they often fail to update them as the organization's goals and social media strategies change. They may believe that their e-mail and communications policy—or existing social media policy—will pretty much cover social media use and that it is not worth the time and expense to update IG policies to include social media.

   This invites complexities, vagaries, and potential disaster. A simple Twitter comment could invite litigation: “Our new project is almost ready, but I'm not sure about the widget assembly.” It's out there. There is a record of it. *Instant potential liability in 280 characters or less*.

   Social media can add value to an organization's efforts to reach out to customers and other stakeholders, but this must be weighed carefully against the accompanying risks.

   The objectives of a social media initiative must be spelled out, and metrics must be in place to measure progress. But more than that, *who can utilize social media on behalf of the company and what they can state needs to be established with clarity in the IG policy*. If not, employees are essentially flying blindly without controls, and they are more likely to put the enterprise at risk.[17]

   More than policy development is needed. If your organization is embarking on a social media program, it needs an executive sponsor to champion and drive the program, communicating policy to key leaders. As your social media program grows, you need to keep the executive team aware of the status of the social media program (including ROI) and reaffirm the commitment from the executive sponsor. You will also need to conduct training—on a consistent basis. *Training is key, since social media is a moving target*.

2. *Employees—the accidental and intentional insider threat*. This may be in part due to lack of social media policy or due to lack of training, monitoring, and enforcement. Sometimes an employee harms an organization intentionally. You may remember three of the most high-profile leaks of information from the US federal government: Private Bradley (now Chelsea) Manning, Edward Snowden, and Reality Leigh Winner.[18] But *most times* employees do not realize the negative impact of their behavior in posting to social media sites. Such was the case for Twitter's Chief Financial Officer, Anthony Noto, when he accidentally sent

a private message discussing a potential deal or acquisition publicly resulting in speculation by others in the industry.[19]

This incident shows that high-level executives must be just as careful as lower-level employees. Noto's revelation was unintentional. Consider what a rogue employee intent on damaging the company might do. The impact could be much worse. For instance, what if a chief executive's assistant were to release details of strategic plans, litigation, or ethics investigations to the public? Or embarrassing details of the CEO's private life? The impact could be quite costly.

People might use social media to vent about a bad day at work, but the underlying message can damage the company's reputation and alienate coworkers and clients. Other times a post that is seemingly unrelated to work can backfire and take a toll on business. We're all human. Emotion sometimes gets the better of us, and we act before we have rationally thought out the consequences. That is especially true in the world of social media, where it may be unclear exactly who can see a comment.

Two of the biggest threats of social media use for organizations come from the lack of a social media policy and threats presented by employee use.

Legal Risks of Social Media Posts

With over 330 million active Twitter users, 81% of whom check Twitter at least once a day,[20] surely some employees in your organization are tweeting. As of the first quarter of 2018, more than 450 million professionals in over 200 countries and territories were members of the LinkedIn network, and the network continues to expand, with 94% of B2B (business-to-business) organizations relying on LinkedIn for content marketing and distribution.[21]

The casual use of public comments can easily create liability for a company. *With no IG policy, guidelines, monitoring, or governance, legal risks of using social media increase significantly. This is an avoidable risk.*

Many people are posting birthday wishes and pictures of what they had for dinner, but others may be venting about specific companies and individuals within those companies. There's a difference between "I can't stand Wall Street," and "Goldman is run by Satan, and his name is John Smith. We're going to sue his butt off." *Instant liability.*

With no IG policy, guidelines, monitoring, or governance, legal risks of using social media increase significantly. This is an avoidable risk.

The specifics of where and how an employee posted or tweeted a message may mean whether a lawsuit against your company is successful or not. If a personal LinkedIn or Twitter account is used, and it was posted after hours using a PC from home, the company *may* be off the hook. But if it was done using a company computer or network, or from a company-authorized account, a defense will be difficult. Opposing counsel likely will ask questions about the policy for posting first. One thing is true: social media will be a target of discovery demands for the foreseeable future.

Just when compliance and records managers thought they had nailed down IG for e-mail, IM, and electronic records, social media came on the scene creating new, dynamic challenges!

Even though not all social media content will rise to the level of a record, according to the definition in use, the organization still may be responsible for managing the nonrecord content. For example, an organization may consider a social networking profile a record but consider comments nonrecords. That decision will have an impact on what must be retained according to the records retentions schedule. It does not, however, absolve the organization from monitoring and evaluating the comments.[22]

"Tweets are no different from letters, e-mail, or text messages—they can be damaging and discoverable, which is especially problematic for companies that are required to preserve electronic records, such as the securities industry and federal contractors. Yet another compliance headache is born."

Blogs are simply Web logs, a sort of online journal that is focused on a particular topic. Blog readers can become followers and receive notices when new content is posted as well as add their own comments, which may be moderated or restricted. It seems confounding, but with the explosion in the use of blogs, there have been actual incidents where employees have "disclosed trade secrets and insider trading information on their blogs. Blogs have also led to wrongful termination and harassment suits."

So the liability and potential for leakage or erosion of information assets is not theoretical; it is *real.*

To safeguard the enterprise that sanctions and supports blog use, *IG policies must be clear, and real-time capture and management of blog posts should be implemented.* Remember, these can be business records that are subject to legal holds, and authenticity and accuracy are crucial in supporting a legal case. So a true and original copy must be retained. This may, in fact, be a legal or regulatory requirement, depending on the industry.

The informal nature of social media posts potentially can be damaging to an organization. The usual fact checking and vetting that is done for traditional press releases and advertising may not be conducted, so social media posts can be unscreened and unfiltered, which poses problems when IG policies are not clear and fully enforced. Beyond that, the consequences of violating policy should be severe and clearly stated in policies, as should the penalties imposed, a message that should be reinforced consistently over time.

Tools to Archive Social Media

New approaches to capture, manage, and archive social media are emerging. Some are free or inexpensive and appropriate for personal and small business use. Others require a more substantial investment of resources but better meet the needs of midsize

and large organizations.

Public Social Media Solutions

In 2017, 81% of US citizens had a social media profile. It's not surprising, then, that most businesses believe social networking can influence their revenue and sales[23] and that a growing number of citizens believe "it is a priority for government to integrate their digital services with social media."[24]

The increasing use of social media by business and government brings with it a growing expectation that the social media communications will be planned, managed, stored, and archived based upon the content of the communications.

Social media networks—including Facebook, Twitter, and LinkedIn—provide options to archive (download) data that individuals and small businesses may use. Facebook users can download and archive their Facebook data from their account settings page. However, you cannot select the data you'd like to archive, so all Facebook data would have to be downloaded repeatedly to capture new posts. Twitter provides a similar feature to download a zip file of your Twitter archive. Both send an e-mail to the account holder when the download file is available. LinkedIn allows users to download their data as well. Two e-mail messages are currently sent to users: (1) Within minutes a message will be sent allowing download of certain categories of personal information that can be compiled quickly (e.g. messages, connections, and contacts); and (2) within 24 hours a second e-mail will be sent with a link to allow the download of the full archive, including activity and account history.

Another option is to convert records to a standard format for use outside of the social media application. Products exist that can be used to create PDF documents out of social media posts, including PDF995 and PDFCreate.

Because Facebook, Twitter, and LinkedIn initially did not provide archiving tools, some third-party applications popped up to perform the task. One example still available is TwInbox, a free MS Outlook plug-in that archives Twitter postings and allows users to install a (Twitter) menu option to send tweets directly from Outlook; these tweets are archived into a standard Outlook folder. The folder can be configured to capture tweets that a user sends outside of Outlook, so that everything is stored in one folder.[25] However, other options to back up a specific social media network, including a free plug-in (ArchiveFacebook) for Mozilla's Firefox browser, have been discontinued.

The tools we've talked about previously may not provide a legally defensible audit trail in court. The trend today is to seek an automated, integrated approach to archiving social media that meets all (or most) of your social media archiving needs and that can help the organization meet its compliance obligations.

PageFreezer is a software-as-a-service (SaaS) solution for small and medium businesses as well as large enterprises that wish to preserve website and social media content for compliance with regulations such as FDA, FINRA, and SEC. It is also suitable for government agencies that must preserve public records and make them available in response to Freedom of Information requests. PageFreezer provides multiple, redundant cloud storage. All content is available through a user-friendly dashboard. The monthly fee is based on the number of users.[26]

Additional archiving solutions may be available as the social media archiving market matures.

Government and Industry Solutions

Most of the products and methods that could be of use for personal or small business archiving of social media content involves manual intervention, which can be time consuming. All organizations must focus on their core business and would benefit from tools and services that streamline and automate the archiving process as much as possible—however, there is a cost. Midsize and large organizations, often using both public and enterprise social media technologies, may find the investment in commercial products and services worth the additional cost, especially those products that integrate and manage social media content with other enterprise content. Capture and management of social media content is an area that must be addressed as part of an overall IG strategy. Some of the solutions available at this time are described in Table 13.4; however, because of the recent increased focus on archiving solutions for public and enterprise social media content, the landscape will continue to become more efficient, effective, and *unified*.

The two social media archiving solutions suitable for mid-sized and large organizations cited in Table 13.4 are Smarsh and ArchivesSocial. Because their features are similar to those offered by PageFreezer, suitable for small businesses, a brief description of each is warranted.

Smarsh is a cloud-based archiving solution that archives e-mail, social content, instant messages, mobile messages, and website content. Not only does Smarsh archive content from Facebook, Twitter, and LinkedIn but from a number of other social media networks including Instagram, Google+, Flickr, Vimeo, Jive, and Pinterest. In 2018, Smarsh announced partnering with NextRequest, a solution provider that enables government agencies to quickly respond to public record requests.[27] This end-to-end solution allows for the capture of electronic communications, including social media and instant messaging, ingest from the archiving platform into the Next Request portal, and a public-facing portal for citizens to make, track, and pay for public records requests.

Table 13.4 Social Media Archiving and Management Software

Type of Solution	Description	Examples
Archiving solution	Services that capture, protect, and retain social media for compliance, e-discovery, digital preservation, and records management	Archives Social; Smarsh
Unified solutions	Services and software that facilitate the management of various file types across the enterprise (e.g. social media, legacy data, word files, SharePoint files) for storage, optimization, e-discovery, compliance, and records management	Unified Archive® by ZL Technologies; Retain Social Archiving (Micro Focus)
Integrated solutions	Services that integrate various types of systems (e.g. customer relationship management in the cloud with social media tools, enterprise content management (ECM), and/or records management) to manage records and information for business operations and compliance.	Office 365/SharePoint Online (integrates with Yammer and contains social and collaboration features as well as RM and compliance features); Salesforce and Chatter (integrates social collaboration technology and potential to integrate with ECM content repository and ECM Documentum Records Manager).

Another option is ArchiveSocial, a cloud-based solution focusing on social media archiving for government, education, and law enforcement. Records can be viewed in their original context, changes made over time can be viewed using a version history feature, and PDF exports can accurately reconstruct the social media conversation that surrounds the keywords relevant to a records request. This plan can be extended to provide an Open Archive for public access to social media records.[28]

The trend toward providing an integrated solution that allows the user to manage content via a dashboard, search for and retrieve content in context in response to requests, and provide a public-facing interface to share information is certain to continue.

In addition to providing archiving functions, unified and integrated solutions provide business intelligence applications and tools to enable the enterprise to better achieve its organizational goals, processes, and performance requirements.

IG Considerations for Social Media

An early report on social media use, "How Federal Agencies Can Effectively Manage Records Created Using Social Media Tools," recommends building an IG framework for social media that resonates today. An IG model provides the overarching policies, guidelines, and boundaries for social media initiatives.[29]

An IG framework for social media should incorporate social media policy, controls, and operational guidelines as well as spell out consequences for violations. Best practices for social media are continually evolving. In addition to establishing policies to govern the use of social media across the organization, best practices should include industry-specific, vertical market considerations. A cross-section of functional groups within the enterprise should provide input into the policy-making process. At the very minimum, internal audit, marketing, finance, information technology (IT), legal, human resources, and RM must be consulted, and all business units should be represented. Clear roles and responsibilities must be spelled out, and controls must be implemented to govern acceptable use—essentially what is allowed and what is not. Even writing style, logo format, branding, and other marketing considerations should be weighed. The enterprise's image and brand are at risk, and prudent steps must be taken to protect this valuable, intangible asset. And most important, all legal and regulatory considerations must be folded into the new IG policy governing the use of social media.

> **An IG framework for social media should incorporate social media policy, controls, and operational guidelines, and spell out consequences for violations.**

Key Social Media Policy Guidelines

Your social media policy development process can begin by examining the published policies of major organizations in your industry or closely related industries. It should also be based on changes in the workplace as well as established standards, such as guidance developed as the result of an August 2016 ruling by the National Labor Relations Board.[30] *More important, social media policies must be hand-crafted and customized for each organization*. And further, they should be divided into two sections (or perhaps separate policies) to address two different purposes: (1) social media policy for company's official accounts and (2) social media policy for employees.[31]

A prudent and properly crafted social media policy:

- Specifies who is authorized to create social media accounts for the organization.
- Authorizes specifically who can speak on the organization's behalf and who cannot (by role/responsibility).
- Outlines the types of negative impact on the company's brand and reputation that unscreened, poorly considered posts may have.[32]
- Draws clear distinctions between business and personal use of social media and specifies whether personal access is allowed during work hours.

- Underscores the fact that employees should not have any expectation of privacy when using social media for corporate purposes, just as in using other forms of communications such as e-mail, IM, and voicemail, which may be monitored.
- Clearly states what is proper and allowed on the organization's behalf and what is forbidden in social media posts or using organization resources.
- Instructs employees to always avoid engaging in company-confidential or even controversial discussions.
- Encourages/requires employees to include a standard disclaimer when publishing content that makes clear the views shared are representative of the employee and not the organization.
- Strictly forbids the use of profanity and uses a professional business tone, albeit more informal than in other corporate communications.
- Strictly forbids any statements that could be construed as defamatory, discriminative, or inflammatory.
- Outlines clear punishments and negative actions that will occur to enforce social media policy.
- Draws clear rules on the use of the company name and logo.[33]

The policy need not be long but should be clear. Best Buy's social media policy, for example, uses the slogan, "Be smart. Be respectful. Be human."[34] It then breaks the guidance into two major sections: what you should do and what you should never disclose. A word of caution contained in the Best Buy Social Media Policy explains the rationale for the employee to abide by the social media policy: *protect the brand, protect yourself*.

To ensure compliance with the organization's IG strategy, it is also necessary to include a reference to the organization's related policies, including the records and information management policy.

Records Management and Litigation Considerations for Social Media

Legal requirements and demands trump all others when making decisions about capturing and preserving social media records. Social media is no different from other forms of **electronically stored information** (ESI) in that it is potentially discoverable during litigation.[35] Not all ESI residing in social media are records, but all are discoverable. If an organization employs social media and makes a conscious decision *not* to archive all or some portion of that data, it is taking risks. A legally defensible records retention schedule must be in place. The schedule must be based on specific laws that identify the records that must be retained and related to a records retention policy that explains the process for identifying, categorizing, and managing information and records.

US corporations that utilize social media are compelled to preserve those records, including metadata and associated linked content, according to Rule 34 of the **Federal Rules of Civil Procedure** (FRCP), which states that opposing parties in litigation may request "any designated documents or ESI—including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations—stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a usable form." Guidance on complying with Rule 34 of the RCP as amended in 2015 is provided by The Sedona Conference, in their publication, Federal Rule of Civil Procedure 34(b)(2) Primer: Practice Pointers for Responding to Discovery Requests.[36] In addition, Rule 26 of the FRCP requires that any and all information that might be discoverable or "potentially responsive" must be preserved and produced if requested by the opposing party. So it is clear that there is a legal duty to preserve social media records.

> U.S. corporations must archive social media records under Rule 34 of the FRCP.

From a RM perspective, it is critical to consider that social media posts are more than the posts themselves; for legal or compliance purposes, they include metadata and hyperlinks to external content—*and that external content in its native format*—that must also be preserved, preferably in real time. That external content may be a PDF document, a PowerPoint presentation, Web site content, or even a video on YouTube, which would require that archiving ESI, along with associated metadata, is in place.

> Social media policy must be unique to each particular organization.

To truly capture the necessary content required by law, records and compliance managers must understand how software programs communicate with each other in order to recommend possible solutions to the IT department. One way to preserve the Web-based data of social media applications is to use the application programming interfaces (APIs) that social media providers offer. APIs offer standard "hooks" into an application. Another way, perhaps preferable, is to enlist a service that can capture and archive information from multiple social networks for the organization.

Content found in social media networks can be static or dynamic. Profiles in Facebook and blog posts are examples of static content. They can be captured before being posted to the Web. Blog comments and endorsements through "liking" or "favoriting" a post are examples of dynamic content. The ideal method from a RM standpoint is to capture all dynamic social media content and its metadata *in real time* in order to be able to prove authenticity and fight claims of records **spoliation** (corruption or adulteration of evidence) in the event of a discovery request.

Regardless of method of capture, social media content that meets record status criteria should be moved to a repository under the control of an **electronic records management** (ERM) system or application. Then business rules for retention should be applied to those records. Typical functions of an ERM system or application include these:

- Marking an electronic document as a read-only electronic record

- Protecting the record against modification or tampering
- Filing a record against an organizational file plan or taxonomy for categorization
- Marking records as essential (vital) records
- Assigning disposal (archival or destruction rules) to records
- Freezing and unfreezing disposal rules
- Applying access and security controls (Security rules may differ from the source electronic document in an electronic document management system or **enterprise content management** [ECM] software.)
- Executing disposal processing (usually an administrative function)
- Maintaining organizational/historical metadata that preserves the business context of the record in the case of organizational change
- Providing a history/audit trail[37]

Robust search capabilities are perhaps the most crucial component of a social media ERM or archiving solution. It is fine to preserve the records and their associated metadata perfectly, but if you cannot easily *find and produce* the information, compliance and e-discovery efforts will fall short and may cost the organization dearly.

Social media policy will be unique to each particular organization. It is fine to start with a social media policy example or template, but it must be tailored to the needs of the organization for it to be effective and legally defensible.[38]

Records Retention Guidelines

Here are some basic records retention guidelines:

- *Make records threshold determinations*. Examine the content to see if it in fact constitutes a record *by your own organization's definition of a record*, which should be contained in your IG policies. This records determination process likely also will require consultation with your legal counsel. If the social media site has not been kept operating, or it was used for a specific project that has been completed (and all pertinent records for that project have been retained), then its content may not require retention.[39]
- *Use existing retention schedules if they apply*. If your organization already has retention policies for, say, e-mail, then any e-mail sent by social media should adhere to that same scheduling guideline, unless there is some legal reason to change it.
- *Apply basic content management principles*. Focus on capturing all related content for social media posts, including conversation threads, and associated metadata that may be required in legal discovery to provide context and maintain the completeness, authenticity, and integrity of the records.
- *Risk avoidance in content creation*. Instruct and reinforce the message to employees participating in corporate social media that content on the Web stays there indefinitely and that it carries potential legal risks. In addition, once something is posted on the Web, completely erasing and destroying the content at the end of its retention period is nearly impossible.

Content Control Models

There are several basic ways to manage social media content, ranging from tightly controlling it through one single, accountable person, to delegating control to the business unit level, all the way to letting the social media participants post their thoughts, unmoderated and unfettered, to encourage spontaneity and enthusiastic use of the tool. The approach your organization takes will depend on the specified business objectives you have for utilizing social media and your organization's appetite for risk.

Emerging Best Practices for Managing Social Media Records

Best practices for managing social media business records are evolving, and will continue to develop as records and information practitioners gain more experience with social media records. Here are some emerging best practices:

- *Establish a social media publication process*. Organizations now consider social media outlets integral to their communications strategy. Document the steps that lead up to sharing information through social media, including who is involved in creating and publishing the post.
- *Identify records during the social media planning stage*. If a new social media initiative is under consideration, both a social media policy and the records and information policy should refer to a request completed by the person or unit proposing the new initiative and indicating if records will be created and, if so, how they will be managed.
- *Promote cross-functional communications*. A social media team of representatives from various departments, such as IT, social media, legal, compliance, records management, and other stakeholders, is formed, and communication and collaboration is encouraged and supported.
- *Require consultation in policy development*. Extending beyond the social media team, input and advice from multiple stakeholder groups is essential for creating IG policies that cover social media records management.
- *Establish clear roles and responsibilities*. The cross-functional social media team must lay out clear expectations and responsibilities and draw lines of accountability so that stakeholders understand what is expected of them.
- *Utilize content management principles*. Management of social media content should fall under an ECM software implementation, which can capture and track content, including associated metadata and external content, and manage that social media content through its life cycle.

- *Implement RM functionality*. Establish a retention schedule and identify a centralized repository to store social media records if they are to be retained in-house. An alternative is to contract with a service that will capture, store, and provide RM functionality for social media content. Features that enable records retention and disposition, implementation of legal holds, and lifting of legal holds are essential.
- *Control the content*. Clear guidelines and monitoring mechanisms must be in place to control and manage content *before* it gets published on the Web, when possible (e.g. static content on blogs and profiles in social networks) if there is any potential legal risk at all.
- *Capture content in real time*. By implementing a real-time content capture solution for content posted directly to social media (e.g. comments on blogs and posting of someone else's content or retweets), organizations will begin their control and management of the content at soonest point and can more easily prove it is authentic and reliable from a legal perspective.
- *Champion search capabilities*. After capture and preservation of records and associated metadata, search capabilities are the single most important feature that the technology must provide.
- *Train, train, train*. Social media is a developing technology that changes rapidly. Users must be trained, and that training must be updated and reinforced on a regular basis so that employees have clear guidelines, understand the technology, and understand the business objectives for its use.

CHAPTER SUMMARY: KEY POINTS

- Organizations are increasingly using social media and Web 2.0 platforms to connect people to companies and government.
- Social media use presents unique challenges because of key differences with other electronic communications systems, such as e-mail and IM.
- Two of the biggest risks that social networking poses to organizations are (1) not having a social media policy; and (2) employees may be—intentionally or not—exposing information that is not meant for public consumption.
- Enterprise social networking software has many of the features of consumer social applications such as Facebook, but with more oversight and control, and they come with analytics features to measure adoption and use.
- Various software tools and services have become available in recent years for archiving social media posts and followers for RM purposes.
- An IG framework provides the overarching policies, guidelines, and boundaries for social media initiatives, so that they may be controlled, monitored, and archived.
- Social media posts are more than the post itself; they include metadata and also include hyperlinks to external content—and that external content must be preserved in its native format to meet legal standards.
- Robust search capabilities are the most crucial component of a social media ERM or archiving solution.
- Social media policy will be unique to each particular organization.
- Best practices for managing social media business records are still evolving but include forming cross-functional social media teams with clear responsibilities, encouraging communication, and capturing complete content and associated metadata in real time.

Notes

1. U.S. National Archives and Records Administration, NARA Bulletin 2014-02, "Guidance on Managing Social Media Records," October 25, 2013, https://www.archives.gov/records-mgmt/bulletins/2014/2014-02.html.
2. See: https://sproutsocial.com/insights/types-of-social-media/ (accessed April 17, 2018).
3. Ibid.
4. Ibid.
5. Emma Dunbar, "10 Reasons Why Your BusinessNeeds to Be on Pinterest," September 16, 2015, https://business.pinterest.com/en/blog/10-reasons-why-your-business-needs-to-be-on-pinterest.
6. Statistica,"Percentage of Internet Users Who Use Tumblr as of January 2018 by Age," 2018, https://www.statista.com/statistics/246214/share-of-us-internet-users-who-use-tumblr-by-age-group/ (accessed April 17, 2018).
7. Irfan Ahmad, "109 Facts and Stats 'bout Reddit" [InfoGraphic], November 4, 2017, https://www.socialmediatoday.com/news/109-facts-and-stats-about-reddit-infographic/508523/.
8. Curtis Foreman, June 20, 2017, https://blog.hootsuite.com/types-of-social-media/.
9. NARA. "Social Media and Digital Engagement" (page last reviewed January 8, 2018), https://www.archives.gov/social-media (accessed April 17, 2018).
10. See: http://www.snapchat.com/ (accessed April 14, 2018).

11. Alex Heath, Instagram continues its attack on Snapchat with disappearing messages, April 11, 2017, http://www.businessinsider.com/instagram-attacks-snapchat-with-disappearing-messages-2017-4.

12. Kristen Herhold, "How Businesses Use Social Media: 2017 Survey, "*Clutch*, September 14, 2017, https://clutch.co/agencies/social-media-marketing/resources/social-media-survey-2017.

13. M. Charki, N. Boukef, N., and S. Harrison, "Maximizing the Impact of Enterprise Social Media," *MIT Sloan Management Review*, February 20, 2018, https://sloanreview.mit.edu/article/maximizing-the-impact-of-enterprise-social-media/.

14. Steve Crompton, "How Do I Measure the Value of My Enterprise Social Network?" *Contentand Code*, (n.d.), https://www.contentandcode.com/blog/measuring-value-enterprise-social-network/ (accessed April 18, 2018).

15. Patricia C. Franks, "How Federal Agencies Can Effectively Manage Records Created Using New Social Media Tools," IBM Center for the Business of Government, San Jose State University, 2010, www.businessofgovernment.org/sites/default/files/How%20Federal%20Agencies%20Can%20Effectively%20Manage%20Records%20C pp. 20–21 (accessed April 15, 2018).

16. Guy Bunker, "Social Media Threats," *SCMedia*, March 1, 2016, https://www.scmagazine.com/social-media-threats/article/530308/.

17. Chris Nerney, "5 Top Social Media Security Threats," *Network World*, May 31, 2011, https://www.networkworld.com/article/2177520/collaboration-social/5-top-social-media-security-threats.html.

18. Malcolm Harris, "Why Do Millennials Keep Leaking Government Secrets?" *The Washington Post*, June 9, 2017, https://www.washingtonpost.com/posteverything/wp/2017/06/09/why-do-millennials-keep-leaking-government-secrets/?noredirect=on&utm_term=.33b72329eb4e.

19. Alexis Kelinman, "Twitter Exec. Accidentally Tweets Private Direct Message," November 25, 2014, *Huffington Post*, https://www.huffingtonpost.com/2014/11/25/twitter-cfo-direct-message_n_6218488.html.

20. Alex York, "61 Social Media Statistics to Bookmark for 2018," 2018, *SproutSocial*, https://sproutsocial.com/insights/social-media-statistics/#twitter (accessed April 18, 2018).

21. Ibid.

22. This discussion and the quotes in this section are from Patricia C. Franks, *Records and Information Management,* 2nd edition (Chicago: American Library Association Neal-Schuman, 2018), 179.

23. Statistica, "Percentage of U.S. Population with a Social Media Profile from 2008 to 2017," 2018, *Statistica*, https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/ (accessed April 18, 2018).

24. Elizabeth Zima, "Report: Effective Government Outreach Requires Social Media," *Government Technology*, February 26, 2018, www.govtech.com/social/Report-Effective-Government-Outreach-Requires-Social-Media.html.

25. See: https://www.techhit.com/TwInbox/twitter_plugin_outlook.html.

26. See: https://reviews.financesonline.com/p/pagefreezer/#what-is.

27. Smarsh, ("Smarsh and NextRequest Team to Streamline Records Requests for Public Agencies," *Smarsh*, March 15, 2018, https://www.smarsh.com/press-release/smarsh-nextrequest-team-streamline-records-requests-public-agencies/.

28. See: https://archivesocial.com/.

29. The next discussion is based on Franks, "How Federal Agencies Can Effectively Manage Records."

30. Philip L. Gordon and Kwabena A. Appenteng, "NRB Ruling in Social Media Case Provides Useful Guidance for Employers," Littler, August 29, 2016, https://www.littler.com/publication-press/publication/nlrb-ruling-social-media-case-provides-useful-guidance-employers.

31. Jylian Russell, "What Is a Social Media Policy?" *HootSuite*, (n.d.), https://blog.hootsuite.com/social-media-policy-for-employees/ (accessed April 18, 2018).

32. Nelson and Simek, "Mitigating Legal Risks of Using Social Media," https://www.questia.com/read/1P3-2557235721/mitigating-legal-risks-of-using-social-media (accessed April 18, 2018).

33. Ibid.

34. Best Buy Social Media Policy, (last edited July 21, 2016), http://forums.bestbuy.com/t5/Welcome-News/Best-Buy-Social-Media-Policy/td-p/20492 (accessed April 18, 2018).

35. The next discussion is based on Rakesh Madhava, "10 Things to Know about Preserving Social Media," *Information Management* (September/October 2011): 34–35, 37. ARMA International, http://content.arma.org/IMM/September-October2011/http://content.arma.org/IMM/September-October2011/10thingstoknowaboutpreservingsocialmedia.aspx.

36. The Sedona Conference. (2018). *Federal Rule of Civil Procedure 34(b)(2) Primer: Practice Pointers for Responding to Discovery Requests*, https://thesedonaconference.org/ (accessed April 17, 2018).

37. Franks, *Records and Information Management*, 2nd edition, 151.

38. Ibid., 36–37.

39. Guidelines here and in the next section are from New York State Archives, "Records Advisory: Preliminary Guidance on Social Media," May 24, 2010, www.archives.nysed.gov/records/mr_social_media.shtml.

Portions of this chapter are adapted from Chapter 13, Robert F. Smallwood, *Managing Electronic Records: Methods, Best Practices, and Technologies*, © John Wiley & Sons, Inc., 2013. Reproduced with permission of John Wiley & Sons, Inc.

CHAPTER 14
Information Governance for Mobile Devices

According to CTIA (The Wireless Association), "There are more than 400 million connections in America, equal to 1.2 wireless devices for every person in the country."[1] This is a more than 100% penetration rate, since many users have more than one mobile device. Citizens of China, India, and the European Union (EU) have even greater mobile phone usage than the United States.

Mobile computing has vastly accelerated in popularity over the last decade. Several factors have contributed to this: improved network coverage, physically smaller devices, improved processing power, better price points, a move to next-generation operating systems (OS) such as Google's Android and Apple's iOS, and a more mobile workforce have fueled the proliferation of mobile devices.

Mobile devices include laptops, netbooks, tablet PCs, personal digital assistants (PDAs) like BlackBerry, and smartphones, such as Apple's iPhone and those based on Google's Android platform. What used to be simple cell phones are now small computers with nearly complete functionality, and some unique communications capabilities. These devices all link to an entire spectrum of public and private networks.

A report by IDC noted that "By 2020 mobile workers will account for nearly three-quarters (72.3 percent) of the US workforce." This significant shift to mobile workers has gained momentum as mobile computing capabilities have improved.[2]

With these new types of devices and operating environments come new demands for information governance (IG) policies and unknown security risks.[3] "The plethora of mobile computing devices flooding into the market will be one of the biggest ongoing security challenges [moving forward] ..." the Digital Systems Knowledge Transfer Network, a UK think tank, found. "With mobile devices connecting to Wi-Fi and Bluetooth networks, there are suddenly many more opportunities [for hackers] to get in and steal personal information."[4]

> There are 1.2 mobile devices for each person in the United States, and the rate of mobile device adoption is even greater in countries like China and India.

The rapid shift toward mobile computing means that companies with mobile personnel like salespeople and service technicians need to be aware of, and vigilant toward, these impending security threats, which can compromise confidential information.

Securing mobile devices is critical: a 2018 study conducted by Ponemon Institute found that data breaches average $3.86 million in direct costs, a more than 6%

increase from 2017.[5]

The reality is that most mobile devices *are not designed with security in mind*; in fact, some compromises have been made to enable new smartphone operating systems to run on a variety of hardware, such as the Android O/S from Google. This is analogous to the trade-offs Microsoft made when developing the Windows operating system to run across a variety of hardware designs from many PC manufacturers.

New techniques by rogue hackers include **smishing,** which is sending phishing text messages (Short Message Service [SMS], thus "smish") to unwitting smartphone users, in an attempt to get them to reveal login credentials or to install malware. Smartphone virus infections are particularly difficult to detect and thorny to remove. A user may be unaware that all their data is being monitored and captured and that a hacker may be waiting for just the right time to use it. Businesses can suffer economic and other damage, such as erosion of information assets, or even negative goodwill from a damaged image.

> By 2020 mobile workers will account for nearly three-quarters of the US workforce.

The smartphone market is rapidly expanding with new developments almost daily, each providing criminals with a new opportunity. An International Data Corporation (IDC) report indicated that smartphone sales have outpaced PC sales since 2010, and in another report that nearly 1.5 billion smartphone devices were shipped in 2017.[6] The growth in smartphone sales and new services from financial institutions—such as making deposits remotely by snapping a picture of a check—means that there are new and growing opportunities for fraud and identity theft.

> **Smishing** is sending phishing text messages to unwitting smartphone users.

Awareness and education are key. *The first line of defense is for users to better understand cybercriminal techniques and to become savvier in their use of information and communications technologies*. Holding regular security awareness training (SAT) sessions helps to reduce this risk. Using an entertaining, "gamified" approach, as some SAT suppliers have released, makes SAT training more engaging and fun, and helps employees to retain core concepts.

Biometric authentication technologies (those that use retina, voice, and fingerprint recognition) are mature enough to positively identify a user to ensure the correct person is accessing financial or confidential accounts. Biometric technologies for account access are much harder to hack than traditional passwords, but it is possible to, using sophisticated techniques.

Application suppliers for mobile devices are first concerned about functionality

and widespread adoption, so security is not their top priority. Users must be aware and vigilant to protect themselves from theft and fraud. On a corporate level, organizations must step up their training efforts in addition to adding layers of security technology to safeguard critical electronic documents and data and to protect information assets.

Social engineering—using various ways of fooling the user into providing private data—*is the most common approach criminal hackers use*, and it is on the rise. Machines do their job, and software performs exactly as it is programmed to do, but human beings are the weakest link in the security chain, and as usage trends in the direction of a more mobile and remote workforce, people need to be trained as to what threats exist, and constantly updated on new criminal schemes and approaches. This training is all part of an overall information governance (IG) effort, controlling *who* has access to *what* information, *when,* and from *where*.

> Holding regular security awareness training (SAT) sessions helps to reduce the risk of data loss on mobile devices.

With more and more sensitive business information being pushed out to mobile devices (e.g. financial spreadsheets, medical test results, business contracts, strategic plans, and the like) and advancing and evolving threats to the mobile realm, *IG becomes an imperative; and the most important part of IG is that it is done on an ongoing basis, consistently and regularly*. Policies must be reviewed when a new device starts to be utilized, when new threats are uncovered, as employees use unsecured public WiFi networks more and more, and as business operations change to include more and more mobile strategies. IT divisions must ensure their mobile devices are protected from the latest security risks, but users must regularly be apprised of changing security threats and new criminal approaches by hackers.

> Social engineering is the most common approach criminal hackers use.

Mobile device management (MDM) is critical to secure confidential information assets and managing mobile devices. Some available MDM technologies can wipe devices free of confidential documents and data remotely, even after they are lost or stolen. They can also provide mass security patch updates. These types of utilities need to be deployed to protect an enterprise's information assets in the mobile environment.

Gartner expanded their definition of MDM, placing it under the umbrella of **enterprise mobility management** (EMM). While most IG and IT professionals focus mostly on finding robust MDM solutions, the supplier marketplace has started using the broader term of EMM.[7] MDM solutions are more limited in scope; most EMM solutions include not only MDM but also Mobile App

Management, Mobile Content Management, App Wrapping, Containerization, and other features. The focus is more on the information being secured rather than the device itself, as more and more information is stored in cloud services.

> Enterprise mobility management includes mobile device management but is broader in scope. EMM includes app management and wrapping, content management, Containerization, and more.

Current Trends in Mobile Computing

With the rapid pace of change in mobile computing, it is crucial to convey an understanding of trends, to better know what developments to anticipate and how to plan for them. When a new mobile device or operating system is released, the best thing may be to first wait to see what security threats pop up. It is important to understand the direction mobile computing usage and deployment are taking, in order to plan and develop IG policies to protect information assets.

From CIOZone.com, here are some top trends in mobile computing:

1. *Long-term evolution (LTE)*. The so-called fourth generation of mobile computing (4G) is expected to be continue to be rolled out across North America over the next several years, making it possible for corporate users to run business applications on their devices simultaneously with Voice over IP (VoIP) capabilities. Many areas offer full 4G capabilities.
2. *WiMax* [Worldwide Interoperability for Microwave Access]. As LTE and WiMax networks are deployed in the US, expect to see more netbooks and laptops equipped with built-in radio frequency identification (RFID) and wireless support. [The Microsoft Surface tablet, for instance, does not allow for direct connection.] (WiMax is [a communications] protocol ... that provides ... much faster speeds than WiFi for fixed and mobile Internet access. The [2011] IEEE 802.16m update pushed the speed to up to 1 Giga bit/second fixed speeds.)
3. *3G and 4G interoperability*. [Various wireless providers have] developed a dual mode card that will enable mobile device users to work on both 3G and 4G networks.
4. *Smartphone applications*. Third-party software vendors will increasingly make enterprise applications available for smartphones, including inventory management, electronic medical records management, warehousing, distribution, and even architectural and building inspection data for the construction industry.
5. *GPS*. Global positioning systems (GPS) will increasingly be used to identify end users by their whereabouts and also to analyze route optimization for delivery workers and service technicians. [Without GPS, apps like Uber and

Waze would not exist.]

6. *Security*. As new and different types of mobile devices are introduced, corporate IT departments will find it increasingly challenging to identify and authenticate individual end users. As such, expect to see a combination of improvements in both virtual private network (VPN) software and hardware-based VPNs to support multiple device types.
7. *Anti-virus*. As more third-party business applications are made available on smartphones and other mobile devices, CIOs will also have to be [more] cognizant about the potential for viruses and worms.
8. *Push-button applications*. Let's say a waste disposal truck arrives at an industrial site and is unable to empty a dumpster because a vehicle is blocking its path. Smartphones will increasingly have applications built into them that would make it possible for the disposal truck driver to photograph the impeding object and route the picture to a dispatcher to document and time-stamp the obstruction.
9. *Supplemental broadband*. As carriers implement LTE and WiMax networks, companies such as Sprint and Verizon are looking at potentially extending wireless broadband capabilities to small businesses that don't have fiber optic or copper connections on the ground....
10. *Solid state drives (SSD)*. Corporate customers should expect to see continued improvements in the controllers and firmware built into SSDs in order to improve the longevity of the write cycles in notebooks.[8]

Security Risks of Mobile Computing

Considering their small size, mobile computing devices store a tremendous amount of data, and storage capacities are increasing with the continued shrinking of circuits and advancement in SSD technologies. Add to that the fact that they are highly portable and often unsecured and you have a vulnerable mix that criminals can target. Considering how often people lose or misplace their mobile devices daily, and what valuable targets they are for physical theft (this author had a laptop stolen in the Barcelona airport, right from under his nose!), and the use of mobile devices represents an inherent security risk.

But they don't have to be lost or stolen to be compromised, according to Stanford University's guidelines, intended to help mobile computing device users protect the information the devices contain: "*... intruders can sometimes gain all the access they need if the device is left alone and unprotected, or if data is 'sniffed out of the air' during wireless communications*" (italics added).[9] The devices can be compromised with the use of keystroke loggers, which capture every single entry a user makes. This can be done without the user having any knowledge of it. That means company passwords, confidential databases, and financial data (including personal and corporate credit card numbers) are all at risk.

Mobile devices are vulnerable to man-in-the-middle attacks that can intercept data during wireless communications, especially over unsecured WiFi.

Securing Mobile Data

The first and best way to protect confidential information assets is to remove confidential, unnecessary, or unneeded data from the mobile device. Confidential data should not be stored on the device unless explicit permission is given by the IT department, business unit head, or the IG Steering Committee to do so. This includes price lists, strategic plans, competitive information, photo images of corporate buildings or coworkers, protected heath information (PHI), and financial data such as tax identification numbers, company credit card or banking details, and other confidential information.

If it is necessary for confidential or sensitive data to be stored on mobile devices, there are options to secure the data more tightly, like USB drives, flash drives, and hard drives that have integrated digital identity and cryptographic (encryption) capabilities.

The best way to protect confidential information in a mobile environment is to avoid pushing it to mobile devices or remove it from the mobile device.

Mobile Device Management (MDM)

Mobile device management (MDM), now sometimes conflated with the broader term *enterprise mobility management*, is software that helps organizations to remotely monitor, secure, and manage devices such as smartphones and tablet PCs.[10] MDM improves security and streamlines enterprise management of mobile devices by providing ways to contact the remote devices individually or *en masse* to add, upgrade, or delete software, change configuration settings, and "wipe" or erase data, and make other security-related changes and updates. More sophisticated MDM offerings can manage not only homogenous company-owned mobile devices, but also those that employees use in the workplace in a **bring-your-own-device (BYOD)** environment.

The ability to control configuration settings and secure data remotely allows organizations to better manage and control mobile devices, which reduces the risk of data leakage, and reduces support costs by providing more uniformity and the ability to monitor enforce company-dictated IG policy for mobile devices.

Key vendors in the MDM marketplace include VMWare AirWatch, Apple (Profile Manager) Ivanti, BoxTone, Centrify, Citrix, Good Technology (acquired by BlackBerry), IBM (Endpoint Manager for Mobile Devices), Microsoft, MobileIron,

SAP Afaria, Sophos, SOTI, and Symantec (Mobile Suite).

Rapid growth is expected in the MDM marketplace, and broader EMM market, which is estimated to grow to $2.2 billion by 2022, according to a 2017 study by Strategy Analytics.[11] According to Gina Luk, author of the report, "The two leaders in this space are VMware AirWatch, with 19 percent, and BlackBerry/Good Technology, with 18 percent. However, MobileIron, Citrix and Microsoft all displayed strong signs of growth.... Even SAP, IBM, SOTI, Sophos, and Symantec are challenging the top players in this space for market share." Luk went on to say, "Mobile security and growth in BYOD (bring your own devices) are primary drivers behind EMM adoption."

EMM platforms are transitioning from "tactical device management tools to broader unified end-user computing management (UEM) platforms, crossing mobile devices, apps, and data, as well as traditional computing platforms such as laptops and PCs," according to the reports.

Trends in Enterprise Mobility Management

Some major trends are clearly emerging in mobility management, including:

- *IoT explosion*—as more and more types of devices are connected to the Internet, the challenge to manage and secure the content on those devices will also grow.[12] IoT devices can provide hackers an entryway into an organization, especially if default passwords are not changed and security updates are not applied. EMM software can assist in managing and controlling these new IoT devices;
- *Unified endpoint management*—will help enterprises secure smartphones, laptops, remote printers, and IoT devices. IBM is a leader in this area and new players are entering the market;
- *Deploying AI*—Artificial intelligence is increasingly being used to detect and quickly counter sophisticated malware attacks;
- *Improved BYOD management capabilities*—a heterogeneous mobile environment presents greater challenges, and with the advent of new IoT devices, even more complicated security threats. New capabilities will have more privacy and security capabilities.[13]

EMM trends include managing the variety of IoT devices, deploying AI and unified endpoint management for security, and improving BYOD capabilities.

IG for Mobile Computing

Stanford University's guidelines are a helpful foundation for IG of mobile devices. They are "relatively easy to implement and use and can protect your privacy" and safeguard data "in the event that the device becomes compromised, lost or stolen."

Smartphones and Tablets

- *Encrypt communications*. For phones that support encrypted communication (**secure sockets layer** [SSL], virtual private network [VPN], hypertext transfer protocol secure [https]), *always configure defaults to use encryption*.
- *Encrypt storage*. Phones approved to access confidential information assets must encrypt their bulk storage with hardware encryption.
- *Password protection*. Configure a password to gain access and or use the device. Passwords for devices that access confidential information assets should be at least seven characters in length, and use upper- and lowercase letters, as well as some numerical characters. Passcodes should be changed every 30 days.
- *Timeout*. Set the device so that it is locked after a period of idleness or timeout, perhaps as short as a few minutes.
- *Update*. Keep all system and application patches up to date, including mobile OS and installed applications. This allows for the latest security measures and patches to be installed to counter ongoing threats.
- *Protect from hacking*. Phones approved to access confidential and restricted data must not be jailbroken (hacked to gain privileged access on a smartphone using the Apple iOS) or rooted (typically refers to jailbreaking on a smartphone running the Android operating system). The process of rooting varies widely by device. It usually includes exploiting a security weakness in the firmware shipped from the factory. "'Jailbreaking' and 'rooting' removes the manufacturer's protection against malware."
- *Manage*. Phones approved to gain access to confidential information assets must be operating in a managed environment to maintain the most current security and privacy settings and monitor use for possible attacks.[14]

Portable Storage Devices

- These include thumb drives or memory sticks, removable hard drives, and even devices like iPods that are essentially mobile disk storage units with extra bells and whistles.
- Create a user name and password to protect the device from unauthorized access—especially if lost or stolen.
- Utilize encryption to protect data on devices used to store and/or transport confidential information assets.
- Use additional levels of authentication and management for accessing the device, where possible.
- Use biometric identification to authenticate users, where possible.

Laptops, Netbooks, Tablets, and Portable Computers

- *Password protect*: This is the most basic protection, yet it is often not used. Create a user name and password to protect the device from unauthorized access; require that they are entered each time the computer is used.
- *Timeout*: Require that the password is re-entered after a timeout period for the screensaver.
- *Encryption*: Laptops, notebooks, or tablets used to access confidential information assets should be required to be encrypted with whole disk encryption (WDE).
- *Secure physically*: Physical locks should be used "*whenever the system is in a stationary location for extended periods of time.*"[15]

Building Security into Mobile Applications

While it is a relatively new channel, mobile e-commerce is growing rapidly, and new software applications or *apps* are emerging for consumers as well as business and public sector enterprises. These apps are reducing business process cycle times and making the organizations more agile, more efficient, and more productive. There are some key strategies that can be used to build secure apps.

As is the case with any new online delivery channel, security is at the forefront for organizations as they rush to deploy or enhance mobile business apps in the fast-growing smartphone market. Their priorities are different from those of the software developers churning out apps.

In the banking sector, initially many mobile apps limited customers to a walled off set of basic functions—checking account balances and transaction histories, finding a branch or ATM location, and initiating transfers—but "a new wave of apps is bringing person-to-person payments, remote deposit capture and bill pay to the mobile channel. Simply, the apps are getting smarter and more capable. *But with those capabilities comes the potential for greater threats"* (italics added).[16]

Security experts state that most of the challenges that could result from mobile fraud have not been seen before. Mobile e-commerce is relatively new and has not been heavily targeted—yet. But also industrial espionage and the theft of trade secrets by targeting mobile devices is going to be on the rise and the focus of rogue competitive intelligence-gathering organizations. So user organizations have to be even more proactive, systematic, and diligent in designing and deploying mobile apps than they did with web-based apps.

> Mobile e-commerce is relatively new and presents new security threats.

Software developers of mobile apps necessarily seek the widest audience possible, so they often deploy them across multiple platforms (e.g. Apple's iOS, Google's Android, TCL Communication's BlackBerry, and others) and this forces some

security trade-offs: enterprises *have to build apps for the "strengths and weaknesses intrinsic to every device, which adds to the security challenges*" (italics added).[17]

A side effect of mobile app development efforts from the user perspective is that it can reshape the way they interact with core information management (IM) applications within the enterprise.

The back-office IM systems such as accounting, customer relationship management (CRM), human resources, and other enterprise apps that are driving online and mobile are the same as before, but the big difference comes in how stakeholders including employees, customers, and suppliers, are interacting with the enterprise. In the past, when deploying basic online applications for browser access, there was much more control over the operating environment; whereas with newer mobile applications running on smartphones and tablets, that functionality has been pushed out to the end-user device.

Real Threats Are Poorly Understood

The list of threats to mobile apps is growing, and existing threats are poorly understood, in general. They are just too new, because mobile commerce by downloadable app is still a relatively new phenomenon. So the current list of threats is not complete or well understood. This does not mean the threat is not real because it could be other aspects related to the app. For example, it could be the unsecure network users are on, or a device infection of some sort.

For mobile apps, antivirus protection is not the focus as it is in the PC world; the security effort mostly focuses on keeping malware off the device itself by addressing secure software development methods and network vulnerabilities. Surely, new types of attacks on mobile devices will continue to be introduced (like smishing). That is the one thing that can be counted on.

There have been some high-profile examples of mobile devices being compromised. In 2017, it was reported that White House Chief of Staff John Kelly had been using a comprised smartphone for several months. *Wired* magazine reported that, "The breach was apparently discovered over the summer, when Kelly gave the smartphone to White House tech support after having problems with it and struggling to successfully run software updates."[18] This is more common than one would think. We do not know our smartphones have been comprised until the device does not work anymore because of the hack.

For mobile apps, antivirus protection is not the focus as it is in the PC world; the security effort mostly focuses on keeping malware off the device itself.

Incidents like this and many others make it imperative to understand the mobile app marketplace itself in order that effective IG policies and controls may be developed, deployed, and enforced. Simply knowing how Google has approached

soliciting app development is key to developing an IG strategy for Android devices. Their relative open-door approach initially meant that almost anyone could develop and deploy an app for Google Android. While the open-door policy has evolved somewhat to protect Android users, it is still quite easy for any app developer—well-intentioned or malicious—to release an app to the Android Marketplace. This can pose a risk to end users, who sometimes cannot tell the difference between a real app released by a bank and a banking app built by a third party, which may be fraudulent. Apple has taken a more prudent and measured approach by enforcing a quality-controlled approval process for all apps released to its iTunes App Store. Sure, it slows development, but it also means apps will be more thoroughly tested and secure.

Both approaches, Android and Apple, have their positives and negatives for Google and Apple, and for their users. But clearly, Apple's more curated and quality-controlled approach is better from a security risk standpoint.

Understanding the inherent strengths and, perhaps more important, weaknesses of specific mobile hardware devices and operating systems—and their interaction with each other—is key when entering the software design phase for mobile apps.

It's a different development environment altogether. Windows programmers will experience a learning curve. Mobile apps under Android or Apple operating systems operate in a more restricted and less transparent file management environment.

Bearing that in mind—regardless of the mobile OS—*first ensure that data is secured, and then check the security of the application itself.* That is, practice good information technology (IT) governance to ensure that the software source code is also secure. Malicious code can be inserted into the program and once it is deployed the hackers will have an easy time stealing confidential data or documents.

Innovation Versus Security: Choices and Trade-offs

As organizations deploy mobile apps, they must make choices, given the limited or confined software development environment and the need to make agile, intuitive apps that run fast so that users will adopt them. To ensure that a mobile offering is secure, many businesses are limiting their apps' functionality. So stakeholder users get mobile access that they didn't have before, and a new interface with new functionality, but it is not possible to offer as much functionality as in web apps. And more security means some sacrifices and choices will need to be made versus speed and innovative new features.

Some of the lessons learned in the deployment of online Web apps still apply to mobile apps. Hackers are going to try social engineering like phishing (duping the user into providing access or private information), smishing, and assuming the identity of an account holder, bank, or business. They will also attempt man-in-the-middle attacks where data is "sniffed" out the air during wireless transmission.

With mobile applications, the most used mode of operation is operating the app

directly on a mobile device such as a smartphone. *This is a key difference between apps and traditional PC-based interfaces that rely on browser access or using basic mobile phone text messaging*. Connecting to a business via app can be more secure than relying on a browser or texting platform, which require an additional layer of software (e.g. the browser, texting platform, or WiFi connection) to execute sensitive tasks. These security vulnerabilities can compromise the safety of information transmitted to a secure site. Thankfully, *if the app is developed in a secure environment, it can be entirely self-contained, and the opportunity to keep mobile data secure is greatest when using the app as opposed to a browser-based platform*.

This is because a mobile app provides a direct connection between the user's device and the business, governmental agency, or e-commerce provider. Some security experts believe that mobile apps potentially could be more secure because they can communicate on an app-to-app (or computer-to-computer) level, as opposed to browser-based access on a PC.

> For mobile apps, connecting to a business via app can be more secure than relying on a browser or texting platform.

In fact, "a customer using a bank app on a mobile network might just be safer than a customer accessing online banking on a PC using an open Wi-Fi connection" that anyone can monitor.[19]

How do you combat this browser-based vulnerability if it is required to access an online interface? *The most effective and simplest way to counter security threats in the PC-based browser environment and to eliminate* man-in-the-browser *or* man-in-the-middle *attacks is to use two different devices*, rather than communicate over a standard Internet connection. This approach can be built into IG guidelines.

Consider this: mobile apps can *render greater security*. For example, do you receive alerts from your bank when hitting a low balance threshold? Or a courtesy e-mail when a transaction is posted? Just by utilizing these types of alerts—and they can be applied to any type of software application beyond banking—tech-savvy users can serve as an added layer of protection themselves. If they receive an alert of account activity regularly, they may be able to identify fraudulent activity immediately and act to counter it and stop it in its tracks, limiting the damage, and potential exposure of additional private data or confidential information assets.

Best Practices to Secure Mobile Applications

Mobile computing is not going away; it is only going to increase in the future. Most businesses and governments are going to be forced to deploy mobile apps to compete and provide services customers will require. There is the potential for exposure of confidential data and e-documents, but this does not mean that

organizations must shy away from deploying mobile apps.[20] There are some proven best practice approaches, which can help to ensure that mobile apps are secure.

There are some steps that can be taken to improve security—although there can never be any guarantees—and some of these should be folded into IG guidelines in the policy development process. *BankTech* magazine identified six best practices that can shape an organization's app development process:

1. Make sure your organization or outside development firm uses seasoned application developers who have had secure-coding training and use a *secure software development life cycle* (SDLC).
2. [Developed for banking apps, this approach can be applied to other vertical apps, too.] Follow the guidance suggested by the Federal Deposit Insurance Corp. (FDIC FIL-103-2005) regarding authentication in an Internet banking environment. The guidance describes *enhanced authentication methods,* such as multifactor authentication, that regulators expect banks to use when authenticating the identity of customers using the bank's online products and services.
3. Make sure that the customer (or employee) is *required to re-enter his or her credentials after a certain time period* to prevent someone other than the mobile device's owner from obtaining access to private account information.
4. *Hire an information security expert* to assess the security around your mobile application servers. Unfortunately, *an organization's servers are often overlooked* during a risk assessment, as they require a specialized skill set to test them.
5. *Encrypt sensitive data* that is stored on a mobile device and account data that travels from the handset across the Internet. Ensure that the encryption is implemented properly.
6. Hire a security expert *to test the security of a mobile application* before you implement it across your customer base. (all italics added)[21]

Use mobile app development and deployment Best Practices to reduce mobile computing risks.

Developing Mobile Device Policies

Where do you start? Developing a comprehensive mobile strategy is key before you craft your mobile device policies. You will need input from a variety of stakeholders, and you will need to understand where mobile devices fit in your overall technology infrastructure and strategy. Here are some best practices for developing your mobile device policies:[22]

1. *Form a cross-functional mobility strategy team*—you will need the input of primary stakeholder groups including IT, field business units, and human resources (HR, for policy creation and distribution). Your strategy development process should also tap into the expertise of your risk management, compliance, records management, and legal departments. The aim will be to balance risks and benefits to improve employee productivity and guard against risk while focusing on the goals and business objectives of the organization.
2. *Clarify goals for your mobile strategy*—start your discussion with the big picture, the broader view of the business drivers, challenges, threats, and opportunities that mobile computing provides in today's technology context and your business context. Draw a direct line from your mobile business needs to your planned mobile support strategy and infrastructure. Keep your business goals in mind and link them to the discussion.
3. *Drill down into policy requirement details*—you may want to survey other existing mobile device policies to inform your mobility strategy team. Those from peer organizations and competitors will be most relevant. Then start with the basics: which types of devices and operating systems make sense for your organization to support, what changes and trends are occurring in the technology marketplace, which sensitive e-documents and data you must protect (or disallow) on mobile devices, and what available security technologies (e.g. EMM, MDM, mobile virtual private networks, or VPNs, encryption, information rights management, or IRM) you might deploy. It may be helpful to segment your mobile users into broad categories and break out a list of their specific business needs related to mobile computing. Your strategy and policies for executives will be somewhat different than for users in field business units. And you will need BYOD policies if your organization opts to go this route.
4. *Budgeting and expense control*—Is the organization going to buy devices and pay all mobile expenses through direct billing each month? What cost controls need to be in place? Or will mobile device use expenses be reimbursed by a flat rate or by processing expense reports? What about BYOD? Roaming charges limits? Decisions on the financial and cost control aspects of mobile computing use must be made by your mobility policy team, under the guidance of an executive sponsor.
5. *Consider legal aspects and liability issues*—Consult your legal counsel on this. What key laws and regulations apply to mobile use? Where could users run afoul? What privacy and security issues are most prominent to consider? What about the private data that users may hold on their own (BYOD) devices? An overarching consideration is to maintain security for private information, and to have a policy in place for data leaks and lost or stolen devices. That includes your policy on remote "wipes" of sensitive data or perhaps *all* data.
6. *Weigh device and data security issues*—since most mobile devices—especially smartphones—were not designed with security as a foremost consideration,

you must take steps to protect your sensitive data, and to secure the devices themselves, without impeding business or making operation too difficult for the end user. The world of mobile computing presents new challenges that were not present when IT had full control of endpoint devices and your internal network. So clear mobile security policies and controls must be in place.

7. *Develop your communications and training plan*—users must be apprised and reminded of your mobile device policy if they are going to adhere to it. They also need to know the consequences of violating your policies. Your communications and training plan should be creative—from wall posters to text and e-mail messages, from corporate newsletters to group training sessions. You may want to first pilot your new policy with a small group of users. But communication and training are key: a perfect mobile device policy won't work if it is not communicated properly and users are not trained properly.
8. *Update and fine-tune*—There will be some misses, some places where after you deploy your mobile policy that you find room for improvement. You will receive user feedback which should be considered too. And there will be changes in the technology marketplace and user trends. A program must be in place to periodically (every six months, perhaps) review your mobile device policy and any audit information to make improvements in the policy.

CHAPTER SUMMARY: KEY POINTS

- The plethora of mobile computing devices flooding into the market and the IoT trend will be some of the biggest ongoing security challenges moving forward.
- There are 1.2 mobile devices for each person in the United States, and the rate of mobile device adoption is even greater in countries like China and India.
- It is estimated that by 2020 mobile workers will account for nearly three-quarters of the US workforce.
- **Smishing** is sending phishing text messages to unwitting smartphone users.
- Human beings remain the weakest link in security, particularly with the increasing use of mobile devices. Information governance policies must be established, and employees must be trained to be aware of security and privacy risks.
- Social engineering is the most common approach criminal hackers use.
- Holding regular security awareness training (SAT) sessions helps to reduce the risk of data loss on mobile devices.
- Mobile devices are vulnerable to man-in-the-middle attacks that can

intercept data during wireless communications, especially over unsecured WiFi.

- Mobile e-commerce is relatively new and presents new security threats. For mobile apps, antivirus protection is not the focus as it is in the PC world; the security effort mostly focuses on keeping malware off the device itself.
- Connecting to a business directly via an app can be more secure than relying on a browser or texting platform, which require an additional layer of software.
- Mobile device management (MDM) software helps organizations to remotely monitor, secure, and manage devices such as smartphones and tablet PCs.[23]
- Enterprise mobility management includes mobile device management but is broader in scope. EMM includes app management and wrapping, content management, containerization, and more.
- EMM trends include managing the variety of IoT devices, deploying AI and unified endpoint management for security, and improving BYOD capabilities.
- Mobile computing security challenges require that organizations follow best practices when developing and deploying apps. Some keys are: encrypting sensitive data, using the secure software development life cycle (SDLC) methodology and enhanced authentication methods, and hiring a security expert to test new apps.
- Developing a comprehensive mobile strategy is key before you craft your mobile device policies. You will need input from a variety of stakeholders, and you will need to understand where mobile devices fit in your overall technology infrastructure and strategy.

Notes

1. CTIA, "Industry Data," https://www.ctia.org/the-wireless-industry/infographics-library (accessed September 16, 2018).
2. Andrew Burger, "IDC: Mobile Workers Will Make Up Nearly 75 Percent of U.S. Workforce," June 23, 2015, https://www.telecompetitor.com/idc-mobile-workers-will-make-up-nearly-75-percent-of-u-s-workforce.
3. Stacy Collett, "Five New Threats to Your Mobile Security," *CSO,* August 1, 2017, https://www.csoonline.com/article/2157785/data-protection/five-new-threats-to-your-mobile-security.html.
4. Warwick Ashford, "Mobility among the Top IT Security Threats in 2011, Says

UK Think Tank," *Computer Weekly*, January 7, 2011, www.computerweekly.com/Articles/2011/01/07/244797/Mobility-among-the-top-IT-security-threats-in-2011-says-UK-think.htm.

5. "IBM Study: Hidden Costs of Data Breaches Increase Expenses for Businesses," July 11, 2018, https://newsroom.ibm.com/2018-07-11-IBM-Study-Hidden-Costs-of-Data-Breaches-Increase-Expenses-for-Businesses.

6. "Smartphone Vendor," https://www.idc.com/promo/smartphone-market-share/vendor (accessed September 16, 2018).

7. Tess Hanna, "What's the Difference Between EMM and MDM Anyway?" *Solutions Review,* May 7, 2018, https://solutionsreview.com/mobile-device-management/whats-the-difference-between-emm-and-mdm-anyway/.

8. Bill Gerneglia, "Top Ten Trends in Mobile Computing," CIO Zone, http://mycioview.com/entry/top-ten-trends-in-mobile-computing (accessed September 16, 2018).

9. Stanford University, "Guidelines for Securing Mobile Computing Devices," www.stanford.edu/group/security/securecomputing/mobile_devices.html#Risks (accessed September 16, 2018).

10. Markus Pierer, "Mobile Device Management (MDM)." In: *Mobile Device Management* (Wiesbaden: Springer Vieweg), 10.1007/978-3-658-15046-4_2

11. Ashley Troutman, "Enterprise Mobility Management Market to Reach $2.2B by 2022," *Solutions Review,* November 15, 2017, https://solutionsreview.com/mobile-device-management/enterprise-mobility-management-market-reach-2-2b-2022/.

12. Rahul Sharma, "Enterprise Mobility Management: Know These Key Trends or Be Left Behind," TechGenix, June 4, 2018, http://techgenix.com/enterprise-mobility-management-trends/.

13. Ibid.

14. Stanford University, "Guidelines for Securing Mobile Computing Devices."

15. Ibid.

16. Ibid.

17. Ibid.

18. Lily Hay Newman, "The Worst Case Scenario for John Kelly's Hacked Phone," *Wired,* October 6, 2017, https://www.wired.com/story/john-kelly-hacked-phone/.

19. AU: Please provide text for note 19

20. Beau Woods, "6 Ways to Secure Mobile Apps," Bank Systems and Technology,

May 26, 2011, www.banktech.com/architecture-infrastructure/229700033..

21. Ibid.

22. Alan Joch, "How to Create an Effective Mobile Device Policy," *BizTech*, March 26, 2013, http://www.biztechmagazine.com/article/2013/03/how-create-effective-mobile-device-policy.

23. Markus Pierer, "Mobile Device Management (MDM)."

CHAPTER 15
Information Governance for Cloud Computing*

By Monica Crocker and Robert Smallwood

Cloud computing represents one of the most significant paradigm shifts in information technology (IT) history. It may have evolved as an extension of sharing an application-hosting provider, which has been around for a half century and was common in highly regulated vertical industries, such as banks and healthcare institutions. But cloud computing is a very different computing resource, utilizing advances in IT architecture, system software, improved hardware speeds, and lower storage costs.

The impetus behind cloud computing is that it provides economies of scale by spreading costs across many client organizations and pooling computing resources while matching client computing needs to consumption in a flexible, (nearly) real-time way. Cloud computing can be treated as a utility that is vastly scalable and can be readily modulated, just as the temperature control on your furnace regulates your energy consumption. This approach has great potential, promising on-demand computing power, off-site backups, strong security, and "innovations we cannot yet imagine."[1]

When executives hear of the potential cost savings and elimination of capital outlays associated with cloud computing, their ears perk up. Cloud deployments can give users some autonomy and independence from their IT department, and IT departments are enthused to have instant resources at their disposal and to shed some of the responsibilities for infrastructure so they can focus on business applications. Most of all, they are excited by the agility offered by the on-demand provisioning of computing and the ability to align IT with business strategies more nimbly and readily.

But for all the hoopla and excitement, *there are also grave concerns about security risks and loss of direct IT control*, which call for strict information governance (IG) policies and processes. Managers and IT leaders who are customers of cloud computing services are ultimately responsible for IT performance. A number of critical IG challenges associated with cloud computing must be addressed. These include privacy and security issues, records management (RM) issues, and compliance issues, such as the ability to respond to legal discovery orders. In addition, there are metadata management and custody challenges to consider. An investigation and analysis of how the cloud services provider(s) will deliver RM capability is crucial to supporting IG functions, such as archiving and e-discovery, and meeting IG policy requirements.

Organizations need to understand the security risks of cloud computing, and they must have IG policies and controls in place for leveraging cloud technology to

manage electronic information before moving forward with a cloud computing strategy.

Defining Cloud Computing

The definition of **cloud computing** is, rather, well, *cloudy*, if you will. The flurry of developments in cloud computing makes it difficult for managers and policy makers to define it clearly and succinctly, and to evaluate available options. Many misconceptions and vagaries surround cloud computing. Some misconceptions and questions include:

- "That hosting thing is like SaaS."
- "Cloud, SaaS, all the same, we don't own anything."
- "OnDemand is Cloud Computing."
- "ASP, Hosting, SaaS seems all the same."
- "It all costs the same, so what does it matter to me?"
- "Why should I care if it's multitenant or not?"
- "What's this private cloud versus public cloud?"[2]

Cloud computing is a shared resource that provides dynamic access to computing services that may range from raw computing power, to basic infrastructure, to fully operational and supported applications.

It is a set of newer information technologies that provides for on-demand, modulated, shared use of computing services remotely. This is accomplished by telecommunications via the Internet or a virtual private network (which may provide more security). It eliminates the need to purchase server hardware and deploy IT infrastructure to support computing resources and gives users access to applications, data, and storage within their own business unit environments or networks.[3] Perhaps the best feature of all is that services can be turned on or off, increased or decreased, depending on user needs.

> "Cloud computing encompasses any subscription-based or pay-per-use service that, in (near) real time over the Internet, extends IT's existing capabilities."

There are a range of interpretations and definitions of cloud computing, some of which are not completely accurate. Some merely define it as renting storage space or applications on a host organization's servers; others center definitions around Web-based applications like social media and hosted application services.

Someone has to be the official referee, especially in the public sector. The National Institute of Standards and Technology (NIST) is the official federal arbiter of definitions, standards, and guidelines for cloud computing. NIST defines cloud computing as:

a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.[4]

> Cloud computing enables convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned.

NIST has offered its official definition, but "the problem is that (as with Web 2.0) everyone seems to have a different definition."[5] The phrase "the cloud" has entered the mainstream—it is promoted on prime-time TV—but its meaning and description are in flux: that is, if you ask 10 different people to define it, you will likely get 10 different answers. According to Eric Knorr and Galen Gruman in *InfoWorld,* it's really just "a metaphor for the Internet," but when you throw in "computing" alongside it, "the meaning gets bigger and fuzzier." Cloud computing provides "a way to increase capacity [e.g. computing power, network connections, storage] or add capabilities dynamically on the fly without investing in new infrastructure, training new personnel, or licensing new software. Cloud computing encompasses any subscription-based or pay-per-use service that, in (near) real time over the Internet, extends IT's existing capabilities."[6]

Given the changing nature of IT, especially for newer developments, NIST has stated that the definition of cloud computing "is evolving." People looking for the latest official definition should consult the most current definition available from NIST's Web site at www.nist.gov (and other resources).

Key Characteristics of Cloud Computing

NIST also identifies five essential characteristics of cloud computing:

1. *On-demand self-service*. A [computing] consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.
2. *Broad network access*. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g. mobile phones, laptops, and PDAs [personal digital assistants]).
3. *Resource pooling*. The [hosting] provider's computing resources are pooled to serve multiple consumers using a multitenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of

abstraction (e.g. country, state, or datacenter). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

4. *Rapid elasticity*. Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
5. *Measured service*. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g. storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.[7]

What Cloud Computing Really Means

Cloud computing growth is expected to continue to climb dramatically. A recent Gartner study shows that the United States is the leader in adopting cloud computing, and the market is expanding rapidly.[8] The cloud computing market is expected to grow 21% annually from 2012 to 2016, exceeding $16 billion in 2014 and growing to over $22 billion in 2016.[9]

The use of **service-oriented architecture**—which separates infrastructure, applications, and data into layers—permeates enterprise applications, and the idea of loosely coupled services running on an agile, scalable infrastructure may eventually "make every enterprise a node in the cloud." That is the direction the trend is headed. "*It's a long-running trend with a far-out horizon. But among big metatrends, cloud computing is the hardest one to argue with in the long term*"[10] (emphasis added).

> Among metatrends, "Cloud computing is the hardest one to argue with in the long term."

A common misconception is that an organization "moves to the cloud." In reality, the organization may decide to transition some specific business applications to the cloud. Those specific business applications are selected because a cloud architecture may offer crucial functions that the internally hosted solution does not or because the internal solution is burdensome to maintain. Some examples of business applications that frequently are moved to the cloud include advertising, collaboration, e-mail, office productivity applications, sales support solutions, customer response systems, file storage, and system backups.

Another common misconception is that if your organization does not decide to migrate to a cloud solution, you are protected from all the dangers of cloud computing. The hard facts are that, for the vast majority of organizations, users are

already putting information in the cloud. They are simply using cloud solutions to compensate for limitations of the current environment. They may be using Box to get at information when working remotely or Dropbox to share information with an outside business partner. Or they are using OneDrive to get to documents from their iPad. They may not even realize they just posted company information to a cloud environment, so they do not realize they violated any policy against doing that. To complicate matters, they probably also left a copy of the information within your organization's firewall. Internal users might not realize they are not using the current version, and your records manager does not know another copy is floating around out there. This is completely ungoverned information in the cloud. The best defense against it is to deliver solutions for those business needs so that users do not have to find their own.

> The idea of loosely coupled services running on an agile, scalable infrastructure should eventually "make every enterprise a node in the cloud."

Cloud Deployment Models

Depending on user needs and other considerations, cloud computing services typically are deployed using one of four models, as defined by NIST:

1. *Private cloud*. This is dedicated to and operated by a single enterprise. This is a particularly prudent approach when privacy and security are key issues, such as in the health care and financial services industries and also for sensitive government or military applications and data. A private cloud may be managed by the organization or a third party and may exist on or off premises.
2. *Community cloud*. Think co-ops, nonprofit organizations, and nongovernmental organizations. In this deployment, the cloud infrastructure is *shared by several organizations* and supports a specific community that has shared concerns (e.g. mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on or off premises.
3. *Public cloud*. Open to the public, this cloud can be maintained by a user group or even a fan club. In this case, "the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services."
4. *Hybrid cloud*. This utilizes a combined approach, using parts of the aforementioned deployment models: private, community, and/or public. The cloud infrastructure is a "*composition of two or more clouds,* (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g. cloud bursting for load-balancing between clouds)" (emphasis added).[11]

There are four basic cloud computing models: private, public, community, and hybrid (which is a combined approach).

Benefits of the Cloud

The risks and security vulnerabilities of cloud computing have been reviewed in this chapter—so much so that perhaps some readers wonder whether cloud computing is truly worth it. The answer is a *qualified* yes—it can be, based on your organization's business needs and computing resource capabilities. Besides the obvious benefit of getting your company out of the IT infrastructure business and back to focusing on its real business goals, there are many benefits to be gained from cloud computing solutions.

Some of the specific benefits offered by cloud computing solution are listed next:

- Cloud computing solutions provide a means to support bring-your-own-device (BYOD) initiatives. As long as users have an Internet browser and Internet connectivity, they can use any device to access an application deployed in the cloud.
- Your workers need to be able to access corporate information via a mobile device. Some cloud solutions allow them to access information stored in a secure location that only requires a smart phone and a login. Some of these solutions can even ensure that the information is not actually stored on the device itself. Entire applications, such as expense reporting, can be deployed this way and incorporate mobile capture technology as well.
- Cloud computing solutions provide a mechanism to support collaboration with external business partners. You need to exchange information with an outside business partner in a manner that e-mail just will not support. For instance, you want to create one copy of the information that anyone on your team or on a business partner's team can access and that reflects any updates or changes on an ongoing basis. Or you need to exchange files that are large or in a format that is prohibited by your e-mail servers. And you do not want to grant partners access to information within your firewall and they do not want to grant you access to information within theirs. A third-party cloud-based file-sharing solution may provide the answer. You can post files there, partners can access them, you can update them as necessary, and everyone always has access to the most current version of the information without compromising security to your network.
- A cloud file storage solution provides a better alternative to remote information access than having users copy information to unsecured removable media or send an e-mail to their personal e-mail account. Again, it prevents duplication of information, provides access to the most current version of information, and stores information in an environment that only

authenticated users can access.

- Cloud computing solutions also can form a key part of your organization's disaster recovery/business continuity strategy. If your data center is rendered inoperable, users still can access applications and information hosted by cloud providers. Most cloud providers have redundant data centers so that, even if one of their data centers was affected by the same incident that rendered your data center inaccessible, all your information is available. Many organizations deploy solutions to back up their in-house applications to a cloud-based storage provider for just this reason. It is a way to provide geographic diversification.

The business benefits of cloud computing may largely outweigh the security threats for the vast majority of enterprises, so long as they are anticipated and the preventive actions described are taken.

Security Threats with Cloud Computing

Cloud computing comes with serious security risks—some of which have not yet been uncovered. In planning your cloud deployment, these risks must be borne in mind and dealt with through controls and countermeasures. Controls must be tested and audited, and the actual enforcement must be carried out by management. Key cloud computing security threats are discussed next, along with specific examples and remedial measures that can be taken (fixes). The majority of this information and quotations are from the Cloud Security Alliance.[12]

> Cloud computing carries serious security risks—some of which have not yet been uncovered.

Information Loss

When information is deleted or altered without a backup, it may be lost forever. Information also can be lost by unlinking it from its indices, deleting its identifying metadata, or losing its encoding key, which may render it unrecoverable. Another way data/document loss can occur is by storing it on unreliable media. And as with any architecture—not just cloud computing—unauthorized parties must be prevented from hacking into the system and gaining access to sensitive data. In general, providers of cloud services have more resources at their disposal than their individual clients typically have.

Examples

- Basic operational failures, such as server or disk drive crashes.
- Data center reliability, backup, and disaster recovery/business continuity issues.

- Implementation of information purging without your approval (e.g. purging all data over three years old without regard to your retention schedule or existing legal holds).

The Fixes

- Agreement by cloud provider to follow standard operating procedures for data backup, archiving, and retention.
- Standard procedures for information purges that require your sign-off before they are completed.
- Check your insurance coverage. Are you covered for the costs or liability associated with a breach or loss of information that is stored in the cloud?
- Clear delineation of the process for notifying the client of a security breach or data loss.

Information Breaches

Many times damage to information is malicious, while other times damage is unintentional. *Lack of training and awareness, for example, can cause an information user to accidentally compromise sensitive data.* Organizations must have proactive IG policies that combat either type of breach. The loss of data, documents, and records is always a threat and can occur whether cloud computing is utilized or not.

But the threat of data compromise inherently increases when using cloud computing, due to "the number of and interactions between risks and challenges which are either unique to cloud, or more dangerous because of the architectural or operational characteristics of the cloud environment."

> Lack of training on cloud use can lead to users compromising sensitive data.

Examples

- Lack of **document life cycle security** (DLS) technologies, such as data loss prevention (DLP) and information rights management (IRM) technologies.
- Insufficient **authentication, authorization, and audit** (AAA) controls to govern login access.
- Ineffective encryption and software keys, including lost keys or inconsistent encryption.
- Security challenges related to persistent data or ineffective disposal methods.
- Inability to verify disposal at the end of information life cycle.

The Fixes

- DLS implementation where needed to protect information from creation to their final disposition.
- Strong **encryption** to protect sensitive data at rest, in use, and in transit.
- IG policies for data and document security during the software application design phase as well as testing and auditing the controls for those policies during live operation.
- Secure storage, management, and document destruction practices.
- Contractual agreement by cloud service providers to completely delete data before storage media are reused by other clients.
- Check your insurance coverage. Are you covered for the costs or liability associated with a breach or loss of information that is stored in the cloud?
- Clear delineation of the process for notifying the client of a security breach or data loss.

The Enemy Within: Insider Threats

Since the advent of the National Security Agency controversy and the slew of examples in the corporate world, the threat of the malicious insider is well known. "*This threat is amplified for consumers of cloud services by the convergence of IT services and customers under a single management domain, combined with a general lack of transparency into provider process and procedure*" (emphasis added). It is important to understand your cloud provider's security procedures for its employees: How are they screened? Are background checks performed? How is physical access to the building and data center granted and monitored? What are its remedial procedures for noncompliance?

> It is prudent to investigate the security and personnel screening processes of a potential cloud provider.

When these security, privacy, and support issues are not fully investigated, it creates an opportunity for identity thieves, industrial spies, and even "nation-state sponsored intrusion. The level of access granted could enable such an adversary to harvest confidential data or gain complete control over the cloud services with little or no risk of detection."

Examples

- A cloud provider's employee steals information to give or sell to one of your company's competitors.
- Inadequate screening processes (by your company or a cloud provider) can result in the hiring of people with criminal records, granting them access to sensitive information.

- A cloud provider's subcontractor steals information to give or sell to one of your company's competitors.
- A cloud provider's employee allows unauthorized access to data that your company believes is secure in the cloud.
- The physical cloud storage facility lacks security, so anyone can enter the building and access information.

The Fixes

- Implementation of DLP and IRM technologies and related technology sets at all stages of DLS.
- Assessment of suppliers' practices and complete supply chain, especially those services that are subcontracted.
- Screening and hiring requirements (e.g. background checks) for employees as part of contract with cloud provider.
- Transparent policies regarding information security, data management, compliance, and reporting, as approved by the client.
- Clear delineation of the process for notifying the client of a security breach or data loss.

Hacking and Rogue Intrusions

Although cloud computing providers, as a rule, invest heavily in security, they also can be the target of attacks, and those attacks can affect many client enterprises. Providers of cloud infrastructure service (e.g. network management, computing power, databases, storage) offer their customers the illusion of unlimited infrastructure expansion in the form of computing, network resources, and storage capacity. Often this is coupled with a very easy sign-up process, free trials (even for anonymous users), and simple activation with a credit card. This is a boon to hackers who can assume multiple identities. Using these anonymous accounts to their advantage, hackers and spammers can engage in criminal operations while remaining elusive.

> Easy sign-up procedures for cloud services mean that hackers can easily assume multiple identities and carry out malicious attacks.

Examples

- Cloud services providers have often unknowingly hosted malicious code, including Trojan horses, keystroke loggers, bot applications, and other programs that facilitate data theft. Recent examples include the Zeus botnet and InfoStealer.
- Malware can masquerade as downloads for Microsoft Office, Adobe PDFs, or

other innocuous files.

- Botnets can infect a cloud provider to gain access to a wide range of data, while leveraging the cloud provider's control capabilities.
- Spam is a perennial problem—each new countermeasure is met with new ways to sneak spam through filters to phish for sensitive data.

The Fixes

- IG policies and monitoring controls must require tighter initial registration and thorough user verification processes.
- IG policies and technologies to combat credit card fraud.
- Total network monitoring, including deep content inspection.
- Requirement that the cloud provider regularly monitor public blacklists to check for exploitation.

Insecure Points of Cloud Connection

By their very nature, cloud computing solutions involve the movement of information. Information moves from a workstation in your network to the cloud, from the cloud to a mobile device user, from an external partner to the cloud and then to one of your workstations, and so on. Further, information may be moved automatically from an application in the cloud to an application you host internally and vice versa. The movement of information complicates the process of securing it, as it now must be protected at the point of origin, the point of receipt, on the device that transmits it, on the device that receives it, and at all times when it is in transit.

An **application programming interface** (API) is a way of standardizing the connection between two software applications. APIs are essentially standard hooks that an application uses to connect to another software application—in this case, a system in the cloud. System actions like provisioning, management, orchestration, and monitoring can be performed using these API interfaces.

> APIs must be thoroughly tested to ensure they are secure and abide by policy.

It comes down to this: a chain is only as strong as its weakest link, so *APIs must be thoroughly tested to ensure that all connections abide by established policy*. Doing this will thwart hackers seeking work-arounds for ill intent as well as valid users who have made a mistake. It is possible for third parties to piggyback value-added services on APIs, resulting in a layered interface that is more vulnerable to security breaches.

Examples

- Anonymous logins and reusable passwords can undermine the security of an

entire cloud community.

- Unencrypted transmission or storage and unencrypted verification allow successful man-in-the-middle data theft.
- Rigid basic access controls or false authorizations pose a threat.
- Poor management, monitoring, and recording of cloud logins and activity make it difficult to detect malicious behavior.
- Weak APIs provide opportunities for data compromise.
- Dependency on unregulated API interfaces, especially third-party add-ons, can allow critical information to be stolen as necessary connections are made.

The Fixes

- Utilization of multiple logon authentication steps and strong access controls.
- Encryption of sensitive data during transmission.
- More robust and secure API access control.
- An understanding of the security model of cloud provider APIs and interfaces, including any third-party or organization-created dependencies.
- Understanding how the API impacts associated cloud usage.

Issues with Multitenancy and Technology Sharing

Basic cloud infrastructure is designed to leverage scale through the sharing of components. Despite this, many component manufacturers have not designed their products to function in a multitenant system. Newer architectures will evolve to address this issue.

In the meantime, virtual computing is often used, allowing for multiple instances of an operating system (OS) (and applications) to be walled off from others that are running on the same computer. Essentially, each instance of the OS runs independently, as if it were the only one on the computer. A "virtualization hypervisor mediates access between guest operating systems and the physical compute resources" (like central processing unit processing power). Yet flaws have been found in these hypervisors "that have enabled guest operating systems to gain inappropriate levels of control or influence on the underlying platform"—and therefore indirectly impact the other guest OSs running on the machine. To combat this, "security enforcement and monitoring" of all shared computing resources must be employed. Solid partitions between the guest OSs—known as compartmentalization—should be employed to ensure that one client's activities do not interfere with others running on the same cloud provider. Customers should *never* have access to any other tenant's "actual or residual data, network traffic" or other proprietary data.

Cloud providers use virtualization heavily, and hypervisors may allow intrusions.

Examples

- Joanna Rutkowska's Blue Pill root technique, which describes how an unauthorized user could intercept data by using virtual hardware called a hypervisor. The Blue Pill would be undetectable as long as the host system was functioning properly. Rutkowska also developed a Red Pill, which could detect a Blue Pill hypervisor, allowing the owner to eliminate it.
- Kostya Kortchinksy's CloudBurst is another example of hypervisor exploitation.

The Fixes

- Security IG that leverages best practices for installation, configuration, monitoring, testing, and auditing of cloud computing resources.
- Requirements for monitoring the computing environment for any rogue intrusions or misuse of cloud resources.
- Control and verify access. Promote a more secure two-factor authentication procedure.
- Enforceable **service-level agreements (SLAs)** for patching software bugs, addressing data breaches, and fixing vulnerabilities.
- An IG policy that requires regular audits and evaluations to detect weaknesses in cloud security and configuration.

Hacking, Hijacking, and Unauthorized Access

Hacking into accounts to assume the identity of an authorized user has been happening almost since personal e-mail existed. It can be as simple as stealing passwords with a keystroke logger. Attack methods such as social engineering (e.g. phishing), fraud by identity theft, and exploitation of software vulnerabilities are still effective at compromising systems. Most people recycle a few passwords and reuse them for multiple accounts, so once one is breached, criminals can gain access to additional accounts. If login credentials are compromised, a hacker can monitor nearly everything your organization is doing: a less passive hacker might alter or destroy sensitive documents, create false information, or replace your links with fraudulent ones that direct users to sites harboring malware or phishing scams. Once they have control, it can look like *your organization* is the origin of the malicious downloads or information capture. From here, the attackers can assume the good name and reputation of an organization to further their attacks.

Examples

- Examples are widespread in the general population; however, no clear instances of this occurring with cloud services providers are known (as this book goes to press).

The Fixes

- IG policies should clearly state that users and providers should never reveal their account information to anyone.
- An IG policy should require more secure two-factor authentication techniques to verify login identity, where possible.
- Require your cloud services provider to actively monitor and log all activity in order to quickly identify users engaging in fraudulent actions or those that otherwise fail to comply with the client's IG policy.
- Understand, analyze, and evaluate the cloud provider's contract, especially regarding security protocols. Negotiate improved terms in SLAs to improve or enhance security and privacy.

Who Are Your Neighbors?

> It is important to know what other clients are being hosted with your cloud services provider, as they may represent a threat. Moving to a private cloud architecture is a solution.

Knowing your neighbors—those who are sharing the same infrastructure with you—is also important, and, as we all know, good fences make good neighbors. If the cloud services provider will not or cannot be forthcoming about who else is sharing its infrastructure services with your organization and this becomes a significant issue, you may want to insert contract language that forbids any direct competitor from sharing your servers. These types of terms are always difficult to verify and enforce, so moving to a private cloud architecture may be the best option.

Examples

- The Internal Revenue Service (IRS) utilized Amazon's Elastic Compute Cloud service. When the IRS asked Amazon for a certification and accreditation (C&A) report, Amazon declined. (Note: The C&A process was developed to help ensure compliance with NIST standards and mandated by the Office of Management and Budget, which oversees Federal Information Security Management Act of 2002 compliance.)
- Heartland, a payment processing corporation, suffered a data breach in 2008. Hackers stole account details for over 100 million credit and debit cards. This data was stored on Heartland's network, which the hackers broke into using information (pertaining to employees, corporate structure, company networks, and related systems) it had stolen in the weeks leading up to the

major breach.

The Fixes

- An IG policy that requires full disclosure of activity and usage logs, and related information. Audit the policy for compliance.
- Investigate the architecture of your cloud services provider (e.g. version levels, network OSs, firewalls, etc.).
- Robust and vigilant supervision, logs, and reporting of all system activity, particularly that requesting expansive and detailed reports on the handling of sensitive information.

Additional IG Threats and Concerns

A primary selling point of cloud computing is that enterprises are freed up to focus on their core business rather than being focused on providing IT services. Modulating computer hardware and software resources without making capital expenditures is another key advantage. Both of these business benefits allow companies to invest more heavily in line-of-business activities and focus on their core products, services, and operations. However, the security risks must be weighed against the financial and operational advantages. Further complicating things is the fact that cloud deployments often are enthusiastically driven by advocates who focus inordinately on potential benefits and do not factor in risk and security issues. Additional examples of IG concerns are listed next:

- Lack of clarity about who owns the information (and if that changes at any point).
- Risk of association with any larger failures of the cloud provider.
- Inability of the cloud services provider to manage records *at the file level*.
- Inability to closely *follow the user's retention schedule* and produce certificates of destruction at the end of the information life cycle. This may result in information that is held for too long and ends up costing the client unnecessary expense if it is deemed to be responsive to litigation or other legal action.
- Lack of RM functionality in many cloud-based applications. This problem is not unique to cloud platforms, but the key difference is that internal storage resource systems may have functionality that supports integration with a RM solution. It is unlikely that a cloud provider will provide the option of integrating your in-house RM system with its system. Too many potential security, access control, and performance issues may result.
- Inability to *implement legal holds* when litigation is pending or anticipated.
- *Poor response time*—inability to deliver files quickly and in line with user expectations.
- *Limited ability to ensure that your cloud provider meets your duties to follow*

regulations related to the governance of your information.

- Jurisdiction and political issues that may arise due to the fact that the cloud provider resides outside of the client's geographic region.
- Storage of personally identifiable information (PII) on servers in Europe or other locales that *prohibit or restrict the release of PII back to the United States* (or home country of the cloud services client organization).[13]

An analysis of an organization's exposure to risk *must* include checking on software versions and revision levels, overall security design, and general IG practices. This includes updating software, tools, and policy, as needed.

Finally, for each of these challenges, "IG policies and controls to secure information assets" and "IG policies and controls to protect the most sensitive documents and data" are a key part of the solution.

New CIS Controls for Mobile Guide[14]

Consistent Security Is the Goal

In March 2019, the Center for Internet Security (CIS) released the Mobile Companion Guide to help organizations map the CIS controls and their implementation in mobile environments.[15] In the companion guide, the focus is on a consistent approach to applying the security recommendations in both Google Android and Apple iOS environments. Factors such as who owns the data and who owns the device affect how the device should be secured. The Mobile Companion Guide explores bring-your-own device (BYOD), corporate-owned, personallyenabled (COPE), fully managed, and unmanaged devices.

- **BYOD** (bring your own device): Devices are owned by the end user but occasionally are used for work purposes. Access from BYOD devices to organizational resources should be strictly controlled and limited.
- **COPE** (corporate-owned, personally enabled): COPE devices work in a fashion similar to BYOD. Restrictions will be applied to the device, but generally don't prevent most of what the user intends to do with the device.
- **Fully managed**: Devices within this deployment scenario are typically locked down and only permitted to perform business functions. This means that employees have a second device for personal use.
- **Unmanaged**: A popular model for small companies and startups, this is the most dangerous scenario to the enterprise and should be avoided, if possible.

The Guide also looks at systems that administer and monitor devices, such as enterprise mobility management (EMM), mobile device management (MDM), mobile application vetting (MAV), and mobile threat defense (MTD). The CIS Mobile Companion Guide includes this checklist to track implementation of the 20 controls on mobile devices.

All organizations operate mobile devices and need to adopt a security mindset and

harden the devices to protect against the unique challenges of on-the-go mobile computing environments. The CIS Mobility Guide provides an excellent overview of how to address this challenge. The complete guide can be downloaded from https://www.cisecurity.org/blog/new-release-cis-controls-mobile-companion-guide/.

Applicability Overview

- More than 60% of CIS Sub-Controls Apply
- Between 60% and 0% of CIS Sub-Controls Apply
- 0%

Control	CIS Control Title	Applicability
1	Inventory and Control of Hardware Assets	
2	Inventory and Control of Software Assets	
3	Continuous Vulnerability Management	
4	Controlled Use of Administrative Privileges	
5	Secure Configuration for Hardware and Software on Mobile Devices, Laptops, Workstations and Servers	
6	Maintenance, Monitoring and Analysis of Audit Logs	
7	E-mail and Web Browser Protections	
8	Malware Defenses	
9	Limitation and Control of Network Ports, Protocols, and Services	
10	Data Recovery Capabilities	
11	Secure Configuration for Network Devices, such as Firewalls, Routers and Switches	
12	Boundary Defense	
13	Data Protection	
14	Controlled Access Based on the Need to Know	
15	Wireless Access Control	
16	Account Monitoring and Control	
17	Implement a Security Awareness and Training Program	
18	Application Software Security	
19	Incident Response and Management	
20	Penetration Tests and Red Team Exercises	

CIS Controls™ v7 Mobile Companion Checklist

Managing Documents and Records in the Cloud

The National Archives and Records Administration has established guidelines for creating standards and policies for managing an organization's e-documents

records that are created, used, or stored in cloud computing environments.

1. Include the Chief Records Management Officer and/or lead RM staff in the planning, development, deployment, and use of cloud computing solutions.
2. Define which copy of records will be declared as the organization's record copy and manage these in accordance with information governance policies and regulations.... Remember, the value of records in the cloud may be greater than the value of any other set because of indexing or other reasons. In such instances, this added value may require designation of the copies as records.
3. Include instructions for determining if records in a cloud environment are covered under an existing records retention schedule.
4. Include instructions on how all records will be captured, managed, retained, made available to authorized users, and retention periods applied.
5. Include instructions on conducting a records analysis, developing and submitting records retention schedules to an organization's central records department for unscheduled records in a cloud environment. These instructions should include scheduling system documentation, metadata, and related records.
6. Include instructions to periodically test transfers of records to other environments, including departmental servers, to ensure the records remain portable.
7. Include instructions on how data will be migrated to new formats, operating systems, and so on, so that records are readable throughout their entire life cycles. Include in your migration planning provisions for transferring permanent records in the cloud to central records.
8. Resolve portability and accessibility issues through good records management policies and other data governance practices. Data governance typically addresses interoperability of computing systems, portability of data (able to move from one system to another), and information security and access. However, such policies by themselves will not address an organization's compliance and information governance demands and requirements.[16]

IG Guidelines for Cloud Computing Solutions

A set of guidelines aimed at helping you leverage cloud computing in a way that meets your business objectives without compromising your IG profile is presented next:

1. As with any technology implementation, it is critical that you define your business objectives first, then select the provider that best meets your business objectives—provided, of course, it can meet your IG requirements. This is consistent with applying a proven IT project management methodology to the initiative. Even though the solution may reside outside your environment, the same basic phases for your project approach still apply, especially for those

tasks related to documentation.

2. As part of the project documentation, make sure to identify roles and responsibilities related to the system in at *least* the same level of detail you do for internally supported systems (preferably in more detail).
3. The biggest deviation from your standard approach is the need to incorporate the investigation and application of the appropriate fixes described in the "Security Threats with Cloud Computing" section into your project plan. Again, as with any service contract, it is helpful to involve a good contract negotiator. The contract negotiation phase is when you have the most influence with your provider. Therefore, you have the greatest chance of mitigating potential risks and optimizing the benefits if you can incorporate specific requirements into the contract language.
4. If the cloud computing paradigm is relatively new to your organization, try to figure out approaches to issues and high-level processes that can be reused in subsequent cloud computing projects. For instance, during the course of your project, you need to figure out:
    - How to migrate information including metadata to the cloud solution.
    - How to get your information including metadata back if you quit using that solution.
    - How to implement a legal hold.

Utilizing cloud computing resources provides an economic way to scale IT resources which allows more focus on core business operations. It can render significant business benefits but its risks must be carefully weighed, and specific threats must be countered, in the context of a long-range cloud deployment plan.

> Most cloud services providers do not have mass content migration or RM capabilities.

IG for SharePoint and Office365

By Robert Bogue

Information Governance on SharePoint and Office 365 requires awareness of the capabilities offered by the platform itself and a basic understanding of the layers underlying the platform. In this section, we'll first cover the capabilities of SharePoint on-premises deployments, then the Office 365 infrastructure, and finally information governance in Office 365.

SharePoint IG Features

SharePoint as a product family has been available since late 2000. In that time many things have changed, including the underlying development technologies

and platforms. During the changes the product developed a set of rich capabilities to support information governance. From a basic information management perspective SharePoint supports file versions, approvals, metadata, workflows, and a host of other expected capabilities. Since 2010, SharePoint has supported not just records but also basic eDiscovery capabilities including holds. SharePoint 2016 introduced data loss prevention support as well.

SharePoint's most basic unit of control is a content type. The content type wraps up a set of properties and behaviors including what metadata columns are allowed and which ones are required, retention policies, available workflows, retention policies, and more. Content types are not defined at a farm (installation) level. Nor are they defined at a web application level (fully qualified name). Instead, content types are defined at a site collection level or a site level. A site collection is—as the name suggests—a collection of sites. The fact that content types are defined at such a low level reduces the consistency across different areas of the business.

SharePoint does offer a content type hub which can publish content types to every site collection—minimizing the potential impact of having multiple definitions for the same type of content; however, the out-of-the-box functionality leaves opportunities for third parties to come in to offer a complete solution that can audit when individual site collection owners have modified the corporate published types.

Storage in SharePoint exists in either a list or a library which itself is located in a site. A list is simply a collection of rows which can have attachments and support versioning. A library is a collection of files and folders. Both lists and libraries use the same content type approach and therefore each item can have its own workflows, retention policies, can be declared as a record, and so on. While most of the considerations for information governance occur at a content type level, versioning is implemented in either the list or library.

Some options for information governance can be applied to a list or library. Most of the time the functions are under-the-covers being implemented as information governance controls on the default content type rather than on the list or library itself.

Lists and libraries support two different mechanisms for records management. The first method declares a record by sending it to a records center. Each implementation can have one or more records centers. Once the record is sent to a records center it can be removed from the originating location, replaced with a link to the location in the records center, or left intact. Records can be declared manually or through the use of workflows.

The second records management implementation is referred to as in-place records management and the declaration of a record marks the information so that even users with permission to the item can't take prohibited actions, such as deleting the record. In-place records management resolves some of the concerns with findability of records. However, in-place records management does expose a large retention problem.

SharePoint, out-of-the-box, provides no mechanism for site or site collection lifecycle management. The result is that when an entire site should be destroyed because it's reached its expiration the process must be done manually or via an automated mechanism not built into SharePoint. This is particularly problematic when the records inside the site have different retention schedules where some should be deleted at one interval and others at another interval. Managing this process is left to third parties or organizations to solve themselves.

In addition to records management, SharePoint supports in-place holds. The holds can be triggered through the eDiscovery mechanisms or done manually. Starting with SharePoint 2013, a document on-hold can be modified, though the version that was placed on hold may not be destroyed. Management of holds is performed through an eDiscovery center. eDiscovery in SharePoint is SharePoint only-scoped and therefore represents one more repository to be managed when responding to a request.

Office 365 Infrastructure

It's important to understand that Office 365 is built on top of the Microsoft Azure services and is delivered from Microsoft Azure datacenters. Microsoft maintains numerous certifications for overall compliance and specific compliance with various industry regulations. This means that the physical and data security of the Microsoft data centers which service Office 365 have been thoroughly evaluated.

Additionally, Office 365 is built on top of the Azure Active Directory service, which allows for corporations to synchronize their internally managed active directories to an Azure hosted replica. This replica can be used only as a directory or, with password synchronization, for authentication. Passwords synchronized to Azure Active Directory go through an additional SHA128 hashing process to ensure their safety.

Microsoft offers a variety of authentication security options—some of which are not included in all Office 365 licenses—that allow for multifactor authentication as well as other limitations and controls including rules based on where the login attempt is coming from. For organizations that do not want to accept Microsoft's safeguards for authentication or have additional requirements, authentication can be performed through a federated authentication provider including third parties or organization hosted Active Directory Federation Services (ADFS) which is an included part of your Windows server license. ADFS servers allow for even more fine-grained control of who can login at what times from what locations and what they must do to prove their identity.

Office 365 IG

IG in Office 365 starts with all of the features in SharePoint for SharePoint and OneDrive content and all of the features and capabilities conveyed by nature of the base infrastructure as well as additional capabilities that are unique to Office 365. Features like customer key allows organizations to bring their own encryption keys so that Microsoft isn't able to provide decrypted information even if they're

required by a court or government to turn over customer information. Though organizations would presumably be required to provide their keys to lawful authorities, having the request go directly to the organization allows them to exercise their legal rights to appeal the request.

More broadly, Office 365 has a security and compliance center which provides a platform view of many information governance concerns. Data governance and data loss prevention are both across-service features that apply to Exchange and SharePoint. This provides a single approach that functions across the service regardless of whether the data is stored or transmitted. These features are, at the time of this writing, integrating Azure information protection labels and experiences in Outlook and SharePoint including mobile clients.

While these information governance capabilities do not use the historical SharePoint approaches for data loss prevention nor records management, the fact that they can be applied across the entire offering make them a compelling solution for addressing the multiple repository problem that plagues all large organizations. While the scope extends only to the Microsoft offerings, this can represent a substantial portion of an organization's information governance needs.

CHAPTER SUMMARY: KEY POINTS

- Cloud computing represents a paradigm shift in computing capabilities. It can streamline operations and cut costs but, because it also has inherent risks, a well-researched and documented IG policy is needed.
- Organizations need to understand cloud computing's security risks and formulate IG policies and controls before deploying it.
- Organizations are rapidly moving applications and storage to the cloud. Cloud computing allows users to access and use shared data and computing services via the Internet or a VPN.
- Five key characteristics of cloud computing are: (1) *on-demand self-service,* (2) *broad network access,* (3) *resource pooling,* (4) *rapid elasticity,* and (5) *measured service*.
- Cloud computing services typically are deployed using one of four models: (1) private cloud, (2) public cloud, (3) community cloud, and (4) hybrid cloud.
- Utilizing cloud computing carries significant security risks, which can be offset by establishing IG policies and preventive measures so that the business benefits of agility and reduced cost may be exploited.
- Cloud application services may have weaknesses related to supporting RM functions, such as: the inability to manage records at the file level; the inability to closely follow the user's RM retention schedule, the inability to migrate data and documents to other platforms for preservation, and the inability to enforce legal holds when litigation is pending or anticipated.

Notes

1. Cloud Security Alliance, "Top Threats to Cloud Computing V1.0," March 2010, https://cloudsecurity-alliance.org/topthreats/csathreats.v1.0.pdf, p. 6.
2. R. "Ray" Wang, "Tuesday's Tip: Understanding the Many Flavors of Cloud Computing and SaaS," March 22, 2010, http://blog.softwareinsider.org/2010/03/22/tuesdays-tip-understanding-the-many-flavors-of-cloud-computing-and-saas/.
3. NARA Bulletin 2010-05, "Guidance on Managing Records in Cloud Computing Environments," September 8, 2010, www.archives.gov/records-mgmt/bulletins/2010/2010-05.html.
4. Peter Mell and Tim Grance, "NIST Definition of Cloud Computing," Version 15, 10-07-09, http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf (accessed December 12, 2013).

5. Eric Knorr and Galen Gruman, "What Cloud Computing Really Means," *New York Times*, April 7, 2008.

6. Ibid.

7. Peter Mell and Tim Grance, "NIST Definition of Cloud Computing," Version 15, October 7, 2009, www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf.

8. Gartner Press Release, "Gartner Says Worldwide Public Cloud Services Market to Total $131 Billion," February 28, 2013, www.gartner.com/newsroom/id/2352816.

9. This and the next quotes in this section are from Louis Columbus, "451 Research: Cloud-Enabling Technologies Revenue Will Reach $22.6B by 2016," September 26, 2013, http://softwarestrategiesblog.com/2013/09/26/451-research-cloud-enabling-technologies-revenue-will-reach-22-6b-by-2016/.

10. Ibid.

11. All definitions are from Mell and Grance, "NIST Definition of Cloud Computing."

12. Cloud Security Alliance, "Top Threats to Cloud Computing V1.0."

13. Gordon E. J. Hoke, CRM, e-mail to author, June 10, 2012.

14. Source: https://www.cisecurity.org.

15. https://www.cisecurity.org/blog/new-release-cis-controls-mobile-companion-guide/.

16. NARA Bulletin 2010-05, "Guidance on Managing Records in Cloud Computing Environments."

* Portions of this chapter are adapted from Chapter 12, Robert F. Smallwood, *Managing Electronic Records: Methods, Best Practices, and Technologies*, © John Wiley & Sons, Inc., 2013. Reproduced with permission of John Wiley & Sons, Inc.

CHAPTER 16
Leveraging and Governing Emerging Technologies

This chapter covers some key emerging and advanced information technologies that can be leveraged in information governance (IG) programs: **data analytics**, **artificial intelligence (AI)**, **blockchain**, and the **Internet of Things (IoT)**.

Data Analytics

Analytics are playing an increasing role in IG programs. File analytics software, also referred to as content analytics, is used to discover duplicates, ownership, content classifications, the age of unstructured information, when it was last accessed, and other characteristics. Predictive analytics are used to help search and classify e-documents during litigation. Security analytics are used to monitor and respond to cybersecurity attacks. Analytics are also used to find new value in information by combining various internal and often external data points. *The use of analytics is key to the success of IG programs*.

Big Data analytics has a lot of uses (real-time fraud detection, complex competitive analysis, call-center optimization, consumer sentiment analysis, intelligent traffic management, etc.). Big Data has three defining characteristics, the "three Vs" as Doug Laney of Gartner proposed: *high volume, high velocity, and high variety of data*. Analysis of the appropriately named Big Data can provide the kind of insight into relationships and business patterns that can improve a business' bottom line.

There are several classes of analytics tools that can be applied to render insights into large data collections and can help organizations find new value in information.

The four main types of analytics are:

1. *Descriptive*: Real-time analysis of incoming data
2. *Diagnostic*: Understanding past performance
3. *Predictive*: Forecast of what might happen
4. *Prescriptive*: Formation of rules and recommendations

Descriptive analytics tells you about information as it enters the system. Diagnostic analytics investigates larger data sets to understand past performance and tell you what has happened. Predictive analytics is used to compare year-over-year, or month-to-month, to determine what might happen in the future. Prescriptive analytics helps companies to determine what actions to take on these predictions based on a variety of potential futures.[1]

In order of increasing complexity, and value added, the sequence is: descriptive, diagnostic, predictive, and prescriptive. These analytics tools will become more

important—and more difficult—as we continue to produce unprecedented amounts of data every year.

> The use of analytics is key to the success of IG programs.

Descriptive Analytics

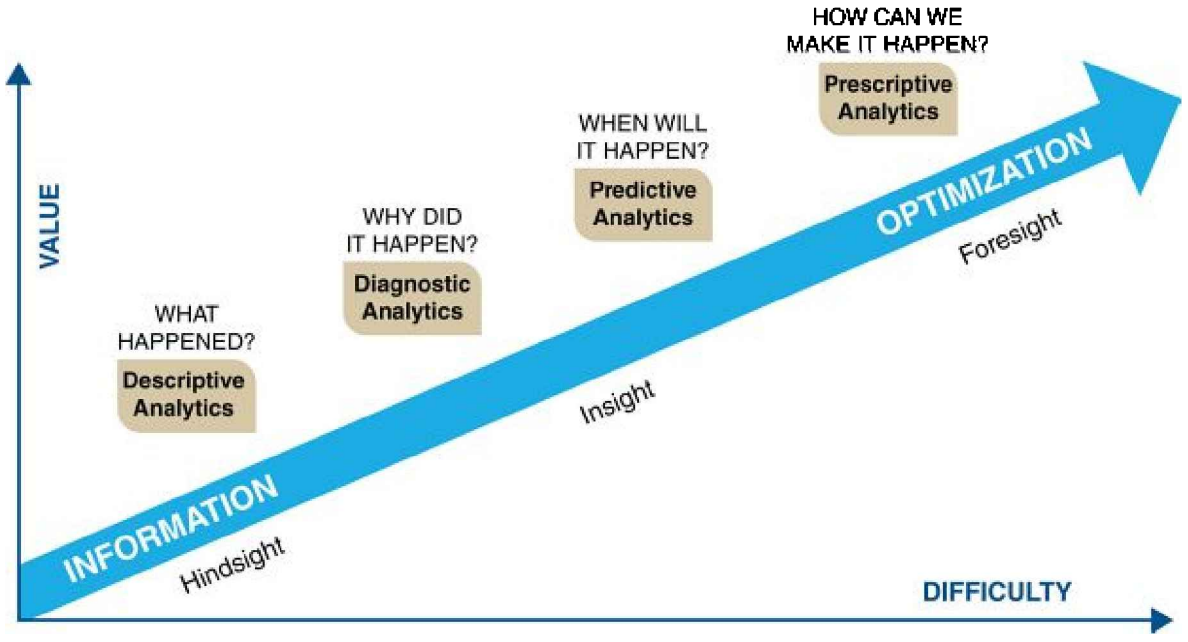
In many ways, **descriptive analytics** (or data mining) is the least sophisticated type of analytics. This doesn't mean that it can't help provide valuable insight into patterns. *With descriptive analytics, we are trying to answer the question: what happened?* Not nearly as exciting as predictive or prescriptive analytics, but useful nevertheless. Just as with other types of analytics, raw data is collated from multiple sources in order to provide insight into past behavior. However, the difference here is that the output is binary: Was it wrong or right? There is little depth to the data; it provides no explanation. Most data-driven companies go well beyond descriptive analytics, utilizing others to understand the data in order to effect change within their business.

Diagnostic Analytics

Diagnostic analytics might be the most familiar to the modern marketer. Often, social media managers track success or failure based on number of posts, new followers (or unfollows), page-views, likes, shares, and so on. Using this methodology, *we are trying to discover why something happened*. Comparative statistics are what drives diagnostic analytics, as we are comparing historical data with another set, in order to understand why data looks the way it does. This is why people love Google Analytics so much: it lets them drill down and identify patterns in how visitors interact with their site.

Predictive Analytics

Predictive analytics utilizes Big Data *to illustrate how patterns in the past can be used to predict the future*. Sales is usually a beneficiary of predictive analytics, as lead generation and scoring, along with the sales process itself, is built out of a series of patterns and data points over time. Predictive analytics tells us *what is likely to happen*, based on tendencies and trends, which makes this forecasting tool so valuable. Bear in mind that statistical models are just estimates. Accuracy is achieved through the continued accruement of clean, accurate data and the refinement of the model based on new information.



Source: Information Governance World magazine (infogovworld.com).

Prescriptive Analytics

Despite its clear value and sophistication, **prescriptive analytics** are not used as often and as widely as it should be, although that is changing. Perhaps it is an insistence on larger, systemwide analytics that makes this narrowly focused body of data overlooked so often. Its purpose is quite simple: *What action should be taken in the future to correct a problem or take advantage of a trend*? By looking at historical data and external information, a statistical algorithm is created that adds value to a company regardless of its industry.

Which Type of Analytics Is Best?

There is no easy conclusion about which one to use. It is dependent on your business scenario. *Deploying a combination of analytics types is best to fit the needs of your company*. Given the options, companies should choose that blend that provides the greatest return on their investment. *Descriptive and diagnostic approaches are reactive; predictive and prescriptive are proactive*. The real take-home message is to utilize data-driven analytics to first learn what is happening using descriptive and diagnostic analytics, and to move toward and understanding of what might happen, using predictive and prescriptive analytics. Leveraging analytics can provide dividends (pun intended) for companies by giving context to their business story and arming decision makers with better information, helping to provide a sustainable competitive advantage.

There are four main type of analytics: descriptive, diagnostics, predictive, and prescriptive. Descriptive and diagnostic approaches are reactive; predictive and prescriptive are proactive.

The Role of Analytics in IG Programs

For nearly four decades, data analytics has been used by leading organizations to gain new insights and track emerging market trends. Today, in the era of Big Data, increasingly sophisticated analytics capabilities are being used to help guide and monitor IG programs.

Structured Data Versus Unstructured Information

Data analytics relies on structured data, which is stored in relational databases. When computers are fed data, it fits into a defined model. All the data within the model is structured. Unstructured information, on the other hand, is basically everything else—e-mail messages, word processing and spreadsheet files, presentation files, scanned images, PDFs, and so on. Unstructured information lacks detailed and organized metadata. Structured data is more easily managed, analyzed, and monetized because it has rich and easily processed metadata. For example, a column titled "Name" will correspond to the name of the person linked to the rest of the data in the row.

It may be rather surprising, but unstructured information is stored rather haphazardly. Every day, knowledge workers create documents and send e-mails to communicate with other knowledge workers. Our personally unique and inconsistent preferences for what we name our everyday office e-documents and where we save them makes for a labyrinth of data. Even the nature of the information within them is rather chaotic. Free-flowing sentences do not make sense to computers like databases full of 1s and 0s. As a result, analysis is more difficult, at least until the proper metadata fields are created and leveraged—then the benefits can be quite significant.

Unstructured information is stored rather haphazardly; applying file analysis can help to insert metadata to organize unstructured information.

Structured data is very useful for determining what is happening in the market or within your organization. However, relying on it will leave you missing the most important piece of the puzzle: why. Clearly, it is advantageous to know what is happening, but without the why it is impossible to act on.

Historically, data analytics has been an imperfect science of observation. Systems produce massive amounts of data, and data scientists correlate data points to determine trends. Unfortunately, correlation does not imply or translate to

causation. Think about all the information that isn't included. Behind every one of these data points are e-mails and instant messages formulating ideas, as well as documents describing the thoughts and processes involved. There is a treasure trove of information to be found in the crucial unstructured data.

Take, for example, a typical enterprise of 10,000 employees. On average, this organization will generate over one million e-mail messages and nearly 250,000 IM chats every single day. During that same time, they will also create or update 100,000 e-documents. The problem is simply being able to corral and cull massive amounts of information into a useable data set. This is not an easy task, especially at scale. The challenge only increases with the production of more data. Not only are established technologies not designed to cope with this type of information, they're also unable to function at such high volumes.

IG Is Key

Without an active IG program, all this relevant information remains opaque in desktop computers, file shares, content management systems, and mailboxes. This underutilized or unknown information is referred to as **dark data**. Not only does understanding and controlling this information add value to your analytics program, it also reduces risk across the enterprise.

To control your information, you must own your information; when you own your information, you can utilize your information. This is much harder with unstructured information because most organizations have environments full of stand-alone, siloed solutions that each solve their own designated issue. This is fine from a business perspective, but a nightmare to manage for RIM and IG professionals.

> IG programs that deploy analytics help to unlock the value of unstructured information.

A single document sent as an e-mail attachment could be saved in a different location by every person included in the chain. Multiple copies of the same document make it difficult, if not impossible, to apply universal classification and retention policies. The same file may be stored and classified separately by legal, compliance, HR, and records––*and no one would know!* When this happens, organizations lose control and expose themselves to undue risks.

Large organizations have petabytes of dark data haphazardly stored throughout their file shares and collaboration software. Much of this information is ROT (redundant, obsolete, or trivial). ROT data hogs storage and can slow down systems and search capabilities—thus hindering business function. ROT may also be stored in retired legacy systems. These legacy systems can be a thorn in the side of IG professionals because of the amount of dark data and ROT intermingled with important business records. Mass deletion is not possible, but neither is the status

quo. Implementing modern, proactive IG strategies can be a daunting task that requires input from a number of sources.

So Where Do We Begin?

IG is not something to jump into all at once, but rather to ease into step by step. Often, the best place to start is with file analysis classification and remediation (FACR) software. In short, file analysis performs a series of system scans across file shares and other disk storage to index this dark data and bring it to light. A deep content inspection is conducted and standardized metadata tags are inserted (the classification and remediation part), based on a predefined metadata design plan. File analysis can be performed on the metadata or the content of files, depending on the intricacy and accuracy needed. Metadata is information about the topic, who created the file, as well as where and when, and how often it has been accessed. *Think about the information on the outside of an envelope being metadata, while the actual letter enclosed is the content*. Performing file analysis helps determine what information is ROT and can be deleted, what can be utilized, and what needs to be preserved.

Leveraging Newfound Knowledge

Analytics can help improve compliance functions by tracking and mapping information flows and communications. A communication heat map allows an administrator to view "who is communicating with whom about what" at a high level, while also having the granularity to drill down into any of the conversations that may set off compliance triggers. Beyond monitoring communications, tools are able to determine if there are sensitive or potentially illegal communications being shared or stored in documents and files. Doing so proactively is an additional safeguard to keep an organization safe.

Analytics can help improve compliance functions by tracking and mapping information flows and communications.

These communication maps are also valuable to Human Resources. Knowing who communicates with whom, and about what, helps determine who the big impact players are within an organization. Understanding who knows and owns important information and tracking communication trends can help assess leadership potential and award promotions based on merit. It can also alert management about potential negative sentiments and potential insider threats to an organization. It can also provide insights to a potential acquiring organization in an M&A scenario.

For legal teams, the data insight can drastically improve Early Case Assessment (ECA) abilities during litigation. Since the legal team knows what information the organization has and where it is stored, there's no mad scramble to find information when litigation is initiated. Being able to analyze what information the

organization holds saves time and effort in collection, while also providing a more accurate data set. It is not necessary to send massive amounts of information to outside counsel to be analyzed, which is very costly and time consuming. When litigation does arise, the legal team can quickly and accurately determine what, if any, liability the organization faces and can make informed decisions on how to proceed. A process that used to take weeks or months can now be completed in hours or days.

The benefits for records management teams are substantial as well. The insights gained from analysis provide important information about which documents are business records and which are unnecessary to retain. This goes beyond typical records, too. Items that are historically not considered records, such as private information discussed in an e-mail, now may be discoverable for litigation. This means record managers need to be able to identify this information and apply retention to it.

> Analytics can provide insights for records managers as to which files are records and which are duplicates or older versions.

File analysis also makes compliance with new regulations (like the European Union's privacy law, GDPR) much easier. Many vendors have promised one-stop GDPR solutions, but the truth is there really is no such thing. GDPR is not something you can solve with a single-point solution, but rather something that requires the implementation of proper IG tools, techniques, and policies. Having in-depth knowledge of the information within an organization makes GDPR dSAR (digital subject access requests) a breeze.

Summing It Up

It is crucial to focus on the cross-functional benefits of IG in order to spur executives into action. The knowledge gained from analytics within IG helps create new value and possibly revenue, while minimizing risk. It is a competitive advantage that will shape the next few decades in the corporate world.

Artificial Intelligence

Artificial intelligence (AI) is simulation of human intelligence by computers: the ability of software to “learn” and make decisions based on inputs and conditions. This creates intelligent computers that can reason (using pre-set rules) and make adjustments or corrections on a fundamental level like humans, only much more rapidly and efficiently. The use of AI has drastically increased and is used for applications like robotics, the Internet of Things (IoT), speech recognition, complex classification, expert systems like medical and maintenance diagnostics, advertising and pricing, and even compliance.

AI tools can significantly assist in business functions, but AI systems can be

expensive to develop and maintain, so they are increasingly being embedded into software applications as well as AI as a service (AIaaS), which allows experimentation and testing prior to making a major investment in AI.

Leading AI platforms include Amazon Machine Learning, IBM Watson, Microsoft Cognitive Services, and Google's Cloud Machine Learning.

There are deep ethical issues that arise with AI, worth considering, but more complex than the scope of this book.

Deep Learning

Deep learning is a type of AI, where the software continuously "learns" based on results it creates, without continual human input. David Naylor noted the "term *deep learning* was adopted to differentiate" between a "new generation of neural net technology" from previous machine-reading technology.[2] In 2012, the University of Toronto used deep learning to "significantly improve speech recognition on Android devices and the prediction of drug activity in a Merck competition." The key is prediction. AI thinks by using deep learning to predict things and make insights.

The Role of Artificial Intelligence in IG

AI solutions, if pundits are to be believed, will solve everything from data storage to transportation. The use of AI to assist in IG program tasks and activities is steadily rising.

Fostering GDPR Compliance

The European Union's new GDPR has left companies across the globe scrambling to gain control over the consumer data they have housed. Some software companies are offering AI tools to assist in this effort. One example is using **machine learning** technology (to automate analytical model building) to simplify compliance tasks. The software can give enterprises a holistic, comprehensive view of consumer data, regardless of where it is stored. It uses machine learning to identify relationships among data in different databases and data stores, including e-mail, instant messages, social media, transactional data, and other sources.

The goal is to ensure compliance with GDPR by gathering technical, business, operational and usage metadata, and providing more accurate compliance analytics and reporting.

AI in Health Care

Investment in healthcare-related AI is "expected to reach $6.6 billion by 2021," resulting "in annual savings of $150 billion by 2026."[3] Organizations that participate in this growth will conceptually utilize an IG program leveraged with state-of-the-art technology infrastructure. This approach allows AI to do what

humans cannot do, that is, structuring massive sets of big data. Facilitating these partnerships requires cooperation and close support for the end user to ensure that regulatory concerns about information sharing are adequately addressed.

While all business sectors will benefit from AI, the healthcare industry will see widespread adoption as administrators and CEOs realize its potential. This is an emerging technology, and, as such, businesses operating within the healthcare industry that begin using AI will gain a competitive advantage. The following five key steps help define how to leverage AI in healthcare:

1. *Users must understand what AI is and what it does.* AI applications use the same data other systems use. Although the common perception is that AI simply replaces human ability, the key point is that it does some things "better" or "faster" or "more accurately" and/or some things humans want to do but cannot. For instance, Finnish company Fimmic Oy developed a deep learning AI application that helps pathologists identify abnormalities the human eye cannot see.[4] The key is to remember AI's ability to leverage data and information at levels humans cannot. It is not a magic robot, but it does have powerful information-processing capabilities. With baseline training, existing healthcare workers should be able to manage and control new AI applications.
2. *AI can use sophisticated algorithms to "learn" features from a large volume of health-care data, and then use the obtained insights to assist clinical practice.*[5] Much of today's AI literature uses the term *deep learning* to describe what AI does behind the scenes.
3. *The long-identified issues of injury and death caused by medical errors (the third leading cause of death in the United States) can be addressed at a micro level.* Given AI's learning capacity, the resulting data then becomes a point of leverage for funding elements such as Medicare reimbursements. Deep learning promotes "self-correcting abilities to improve system accuracy based on feedback."[6]
4. *AI can assist with evidence-based practice (EBP) protocols.* By monitoring the hundreds of accessible information databases, AI can enable real-time EBP. This physician/AI partnership adds to the benefits of **meaningful use** and other U.S. federal healthcare requirements.
5. *AI will contribute to public health initiatives.* By linking preventive medicine routines with elements such as diabetes risk factors, healthcare organizations can begin to project "healthy communities" that ultimately contribute to public health initiatives that are equitable. It certainly costs less money to provide preventive healthcare services than to perform surgeries or to administer extreme treatments in an acute care setting.

AI and file analysis can make compliance tasks easier to execute.

Auto-Classification and File Remediation

AI is also being applied to large collections of unstructured information. Unstructured generally information lacks detailed and organized metadata and must be classified and mapped to an organization's information taxonomy so it can be managed. AI can be used to inspect the contents of e-documents and e-mails to make a "best guess" at how they should be categorized. Some of the more sophisticated FACR software can actually insert basic metadata tags to help organize (remediate) content. This is an essential task for executing defensible disposition, that is, following an established records retention schedule (RRS) and dispositioning (usually destroying) information that has met its life cycle retention requirements.

E-Discovery Collection and Review

AI is also used commonly to locate information that is responsive in a particular legal matter. Using predictive coding software (which uses AI and analytics), a human expert, usually an attorney working on a case, feeds the software examples of content (e.g. documents and e-mails) that are relevant. Then the software goes out into information stores and looks for similar content. It serves up the content and the expert reviewer goes through a sample and teaches the software "more like this" and "not like this" so the AI software gets better and better at narrowing its searches. After a few iterations, the software becomes quite efficient at finding the relevant information. But it doesn't have to be perfect. Courts in the United States have ruled that if the predictive coding software locates 70 percent or more of the responsive information, then that is acceptable, since that is about the accuracy rate of humans, due to fatigue and error.

> Predictive coding software "learns" to find information relevant in legal matters through initial guidance by a human expert.

AI has proven to be a good tool for IG programs to utilize to accomplish key tasks, and the use of AI in IG programs will continue to grow.

Blockchain: A New Approach with Clear Advantages

By Darra Hoffman

This article first appeared in Information Governance World*, Fall 2018. Used with permission.*

By now, we've all heard about **blockchain technology**—or at least its famous progenitor, Bitcoin. According to its evangelists, blockchain technology will secure our records, protect our privacy, democratize our technology, and probably fix us a cup of tea in the process. Blockchain's detractors tend to agree with John Oliver's takedown of Bitcoin and other cryptocurrencies as, "Everything you don't

understand about money combined with everything you don't understand about computers." So, what's the real deal? Is blockchain technology the miracle cure that will soothe the aches and pains of digital information governance? Or is it just so much snake oil?

What Is Blockchain?

That one guy who wears only T-shirts with memes told you that blockchain is the future. So why is it so hard to find out what blockchain actually is? In part, it's because there's no agreed-upon definition as to what constitutes a "blockchain," and in part because there are actually a number of different kinds of "blockchains." While academics can debate the nuances of exactly which technologies are and aren't "blockchain," a blockchain can be understood as:

- A distributed ledger with a decentralized architecture
- Where transactions are:
  - Immutable
  - Secured through cryptography

"There's no agreed-upon definition as to what constitutes a "blockchain," and in part because there are actually a number of different kinds of "blockchains."

> A blockchain is a decentralized, distributed ledger where transactions are secured through cryptography and immutable.

Breaking Down the Definition of Blockchain

A **distributed ledger**, or distributed ledger technology (DLT), is its own technology—of which blockchain is a form. A distributed ledger is a database of transactions. The "distributed" part comes in from the fact that every computer or server running the ledger (every "node") runs that ledger in its entirety; there is no master-slave or master-copy setup. With a **decentralized architecture,** there is no centralized control over who can participate in the ledger. Instead of a centralized authority—say, Janice in accounting—maintaining the ledger, each node can construct and record its own updates to the ledger. *The nodes then all vote on whether each update is valid and what order they occurred in through a* ***consensus mechanism***. While different consensus mechanisms operate differently, they all trust math (instead of Janice in accounting). This is why blockchain is considered a "trust-less" technology: there is no human or institutional intervention necessary to verify transactions. If the nodes reach consensus that a transaction is valid, it stays. If the nodes find a transaction invalid, it must sashay away.

With blockchain's decentralized architecture, there is no centralized control over who can participate in the ledger.

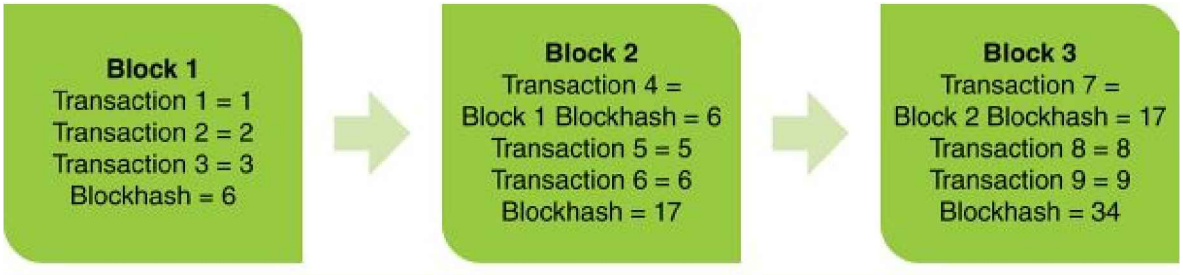
Transactions on the blockchain are made *immutable and secured to the blockchain* through a clever bit of math. With a blockchain, each transaction is cryptographically hashed—a cryptographic hashing algorithm makes an alphanumeric "fingerprint" of the transaction based on its exact content, down to the bit. A block of 10 transactions will have 10 hashes. Those hashes are then all hashed together to make the block hash. That block hash becomes the first hash of the next block, "chaining" all of the blocks together to make ... a chain of blocks (or a "blockchain").

See What We Did There?

In the above illustration (which uses simple addition, as opposed to the incredibly complex math of a real hashing algorithm), Block 2's hash value is dependent on Block 1's value; Block 3, in turn, depends on both Block 1 and 2. Changing the hash of any transaction—which, remember, happens when any bit of that transaction is changed—destroys the entire chain of hashes going forward. Because every block is unbreakably chained to the previous block, the blockchain is considered immutable. Furthermore, the cryptographic hash function works in such a way that it is virtually impossible to reconstruct the original transaction from its hash (much like you can't build a person from a fingerprint). This means that it's impossible to tamper and then go back and hide the tampering.

So What Can Blockchain Do for Your Organization?

Blockchain is a new technology that uses math to secure transactions on a ledger that anyone can read or write to without permission from a central authority. So why do you—a busy information professional—care? Blockchain is way up in the hype cycle; your team might well be asking whether a blockchain makes sense for your organization. *A few benefits of the blockchain get touted pretty often: a blockchain will make our records more secure; a blockchain is more private; or a blockchain is auditable.* To evaluate whether a blockchain makes sense for your organization, you need to know how true each of those claims is.



Claims that blockchains are secure (or at least, more secure than other databases)

rely on a few things. The first is the distributed nature of the blockchain ledger; being able to falsify records on the blockchain typically requires a "51% attack"—or gaining control of 51% of the nodes running the ledger. However, each user controls his/her/their own account through use of a private key; if that key is comprised, just like when a password is compromised, an attacker can then do anything the user could do. This is a real threat when considering the complexity of private keys and the elevated privileges in designs where a trusted body holds users' keys in escrow. People are always a security threat; blockchains are no exception to that rule.

The second element of the blockchain that leads people to claim it is secure is its usage of cryptography (such as the cryptographic hashing). People sometimes think this means data on the blockchain is natively encrypted. It's not. In a public blockchain, like Bitcoin, transaction data cannot be encrypted; if it were, nodes couldn't validate the transaction without decrypting the data. If every node in a private blockchain is going to decrypt in order to validate transactions, then you have to ask why you're spending the time and money to encrypt in the first place. So, even though blockchains use public key infrastructure (PKI) and cryptographic hashing, there's a whole lot of unencrypted data (which, remember, anyone running a node can read) running around on a blockchain. Since encryption is pointed to as a reason that the blockchain is both more secure and more private, it's difficult to overstate how important it is to understand exactly what data is, and isn't, encrypted when considering a blockchain solution.

People are always a security threat; blockchains are no exception to that rule.

Finally, claims that the blockchain will make records more secure often point to the immutability of transactions secured to the blockchain. It's true: this is an excellent tool for ensuring the integrity of records. It also makes auditability a native feature of the blockchain. However, for records to be trustworthy—for information assets to retain their strategic or, in the case of litigation, evidentiary value—they must be accurate, reliable, and authentic.

Integrity Is Only Half of Authenticity

Blockchain cannot ensure the accuracy of a record; it's entirely possible for a user to enter a false or incorrect record onto a blockchain. Reliability is a condition of how a record is created; if Bob enters, say, an employee record into the blockchain without complying with the company's record's procedures, then that will be an unreliable record. Nothing that happens after a record's creation can make it reliable.

Lastly, authenticity—of which integrity is part—requires that a record is what it purports to be. There is nothing in the blockchain that instantiates the archival bond, which means a blockchain doesn't ensure a record's authenticity. Creating, managing, and preserving trustworthy records in a blockchain solution requires a

lot of thought to build and integrate features that are not native to the blockchain.

When Is a Blockchain a Good Solution?

Are blockchains a complete write-off? A fad, doomed to the dustbin of history with Betamax and MySpace? *No!* Blockchains are still a technology in development, but they *offer an excellent solution when you need a database with shared read/write permissions, have low trust between parties, need disintermediation, and have relationships between the transactions in the database*.

The threshold question, then, is why do you need a blockchain (as opposed to simply a secure database)? The best answer is that you have parties who don't particularly trust one another, and you have some reason not to use a trusted third-party intermediary: cost, time, or simply the struggle finding someone all the parties can agree to trust. Like IG itself, blockchain technology integrates social considerations of trust with data and technical considerations.

> Blockchains fit best when a shared database is needed, there is low trust between parties, there is a need for disintermediation, and there are relationships between transactions.

As such, *blockchains are rarely a good solution for information assets within an organization;* the problems of trust and disintermediation (theoretically) shouldn't be an intra-organizational problem. However, they can be very useful for inter-organizational IG. Some of the problem spaces in which blockchain are being explored include land registries, supply chain management, food provenance, healthcare, and financial services. Examples include:

- The Linux Foundation's open-source collaborative blockchain, Hyperledger, being used by IBM to develop a banking application;
- Oracle developing preassembled, cloud-based, enterprise blockchain networks and applications;
- The National Association of Realtors developing a member-engagement data blockchain that allows permission-based access.

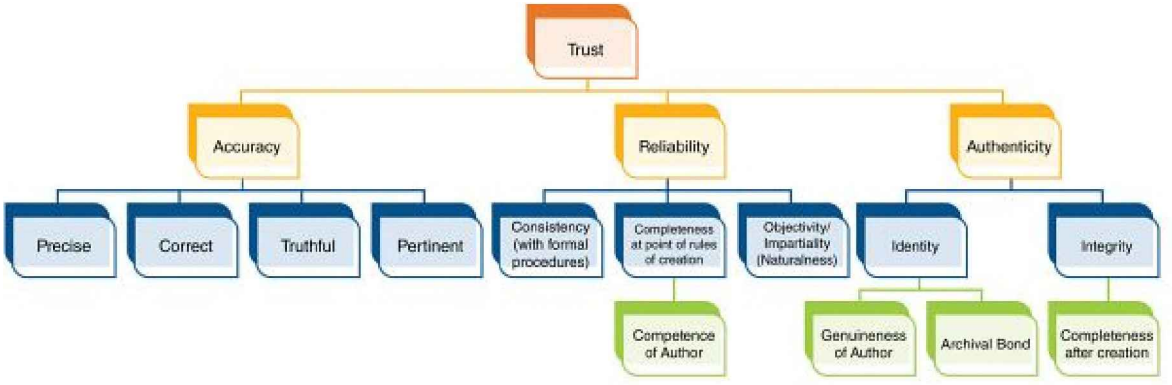


Figure 1: Taxonomy of Trust, by Dr. Victoria Lemieux

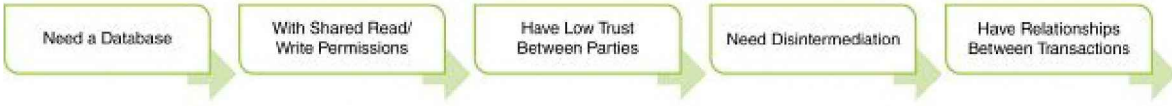
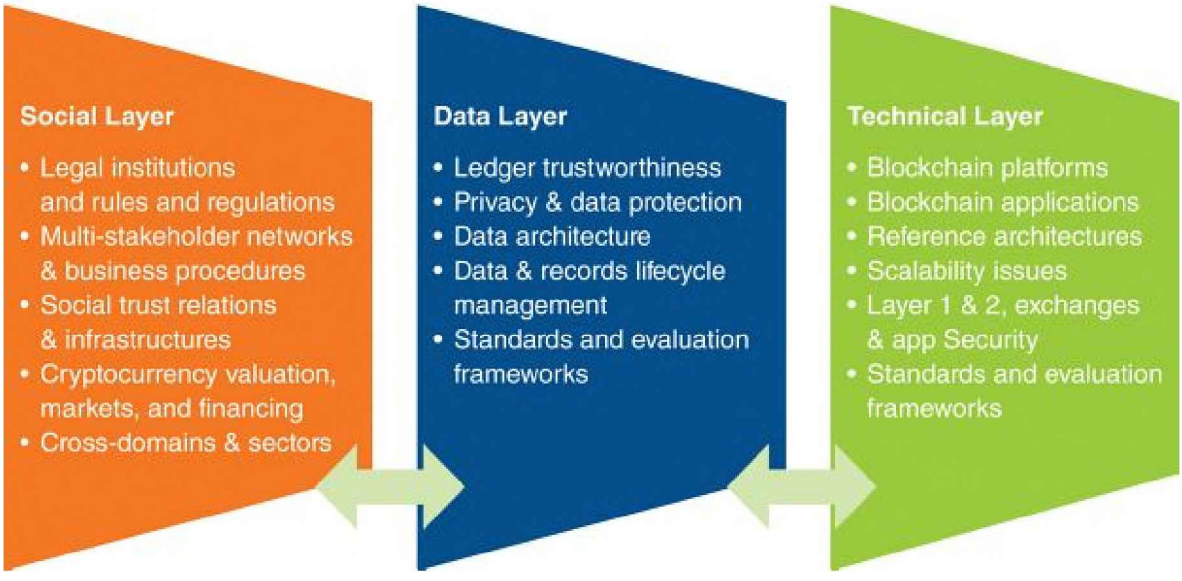


Figure 2: When Is a Blockchain a Good Solution?

Figure 3: The Interrelated Solution Layers of the Blockchain, by Dr. Victoria Lemieux



For those cases where a blockchain makes sense, *design matters*.

Implementing a successful blockchain requires asking in-depth technical questions:

- What consensus mechanisms?
- Permissioned or permission-less?
- What data will be encrypted?
- What kind of transaction speeds do we need?
- How scalable does this system need to be?

But it also requires asking a lot of people- and organization-oriented questions.

- Why do we need a trustless, disintermediated system?
- What are we trying to fix by implementing a blockchain?
- How do we make this accessible and useable to the end users, so that they trust the system where they didn't trust the previous processes?
- What regulatory challenges arise from using such a new technology?
- What makes the blockchain worth the extra investment, and how do we leverage that investment to maximize our return?

Implementing a blockchain should be a strategic choice.

> Blockchains are rarely a good solution for information assets within an organization, as trust between parties is not a major issue.

Conclusion

Blockchains are new and sexy. They combine distributed ledger technology and cryptography in a way that lets transactions be processed without human intervention—and thus no need to trust human fallibility. But new and sexy is often the wrong strategic choice, especially if old and dependable is sufficient to meet organizational needs. Before implementing a blockchain, an organization should ask itself: *Why?*

Blockchain is fundamentally a technology that addresses a social problem—trust. For those cases where low trust and intermediation are problems, blockchain can offer a real solution to serious data management problems, bringing efficiency and transparency to processes that have long challenged interorganizational IG. However, in cases where trust is not the fundamental problem, blockchain technology is not the best solution. The key is asking what organizational needs a blockchain can meet that can't be met by its plainer ancestor, the relational database. Blockchain probably won't get us a cup of tea (though who knows where the Internet of Things will go), but it is a very useful tool to have in the toolbox, as long as one remembers that a hammer does not make every problem into a nail.

The Internet of Things: IG Challenges

First came the Industrial Revolution. Then the Internet Revolution. And today we have made a firm step into the dawn of a third revolution called the Internet of Things or IoT.[7] Or at least this is how IoT's arrival was characterized by the head of the Industrial Internet Consortium—founded in March 2014 by household name companies like AT&T, Cisco, GE, Intel, and IBM to advance and coordinate the rapid rise of IoT. This is no hyperbole. Indeed, Cisco forecasts that IoT will have an economic impact of over $14 trillion by 2022, while per GE's prognosis, IoT could

add $15 trillion to the world economy over the next 20 years.[8]

Stated simply, the IoT is the concept of connecting any device with an "on and off switch" to the Internet (and/or to each other). This includes everything from smartphones, lights and light switches, security doors, refrigerators, cars, washing machines, headphones, wearable devices, and almost anything else you can think of. The IoT also includes components of devices and machines, such as the engine in a car or truck, the sprayers on crop duster planes, vital signs monitors in hospitals, and much more.

The IoT is a clear example of the Big Data trend. With the massive amounts of data that will be created—from such a wide variety of devices, at unheralded speeds—new strategies and methods will be required in order to capitalize on the upside of IoT opportunities while minimizing its inherent risks and planning for e-discovery collection and review.

What created this emerging revolution in the IoT? Faster, cheaper Internet connections make it possible to transmit large amounts of data. WiFi connectivity is being built into devices and sensors, and mobile device use is exploding. New developments in nanotechnology have created micro-electromechanical systems (MEMS), and new software engineering approaches have created microservices, which are a suite of modular software services, each supporting a specific goal.

The onslaught of the IoT is undeniable and the changes in technology and the business environment make it difficult for organizations to develop successful, practical IG policies and practices.

As the IoT becomes a reality, the deluge of data discoverable in legal actions will dwarf the data tsunami that is seemingly engulfing litigation teams today. With the imminent influx of connections in the future, the number of devices and objects that e-discovery professionals will be called upon to collect data from will be infinite. *The IoT will lead to the discovery of everything electronic.*[9]

The onslaught of the IoT make it difficult for organizations to develop successful, practical IG policies and practices.

In layman's terms, IoT represents the exciting and, for some, terrifying ecosystem of interconnected sensory devices performing coordinated, preprogrammed—or even learned—tasks without the need for continuous human input. Think thermostats, which "know" when to expect you to come home, so they automatically cool or warm your home just before you arrive. Now imagine your fitness activity tracker telling your thermostat that it needs to turn the A/C down a bit lower than usual before you come home because you have had an exhausting run. And maybe your auto insurance premiums will be determined in part by your driving habits as transmitted by embedded sensors."[10]

The IoT is a growing area of interest for business and technology professionals. It

"is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction."[11]

Some basic definitions of IoT:

1. TechTarget defines IoT[12] *as: a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.*
2. The IoT European Research Cluster (IERC) states that IoT[13] is —*A dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual ... things have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network.*

> As the IoT develops, the deluge of discoverable data in legal actions will dwarf what litigation teams must manage today.



Source: European Research Cluster on the Internet of Things.

Stated simply, the IoT is the concept of connecting any device with an "on and off switch" to the Internet (and/or to each other). This includes everything from smartphones, lights and light switches, security doors, refrigerators, cars, drones, washing machines, headphones, wearable devices, and "almost anything else you can think of." The IoT also includes components of devices and machines, such as the engine in a car or truck, the sprayers on crop duster planes, vital signs monitors in hospitals, and much more.

The IoT is a massive network of connected devices, which includes people. The interconnections will be "between people-people, people-things, and things-things."[14] The IoT has evolved from the convergence of wireless technologies, microelectromechanical systems (MEMS), microservices, and the Internet. The convergence has helped tear down the silo walls between operational technology (OT) and information technology (IT), allowing unstructured machine-generated data to be analyzed for insights that will drive improvements.[15]

The IoT, fully implemented and exploited, could radically change the way humans function in their everyday lives, including significant changes at work.[16] But is also presets tremendous challenges in cybersecurity, e-discovery during litigation, and IG.

Some may cast off the IoT as a somewhat frivolous development that reminds you to buy more milk or automatically adjusts the heat in your home. However, its implications are much more far-reachng than that, and its impact will affect nearly all industries.[17]

The IoT extends the end node far beyond the human-centric world to encompass specialized devices with human-accessible interfaces, such as smart home thermostats and blood pressure monitors. And even those without human interfaces, including industrial sensors, network-connected cameras, and traditional embedded systems.

As IoT grows, the need for real-time scalability to handle dynamic traffic bursts also increases. There also may be the need to handle very low bandwidth small data streams, such as a sensor identifier or a status bit on a door sensor or large high-bandwidth streams such as high-def video from a security camera.

Homes and Offices

Utility companies will receive constant updates from meters and sensors in the field to monitor systems in real time and to proactively detect and remediate problems such as blackouts, water leaks and circuit overloads. Efficiency is optimized through continual analysis of trends, demand, and outages. In one instance, the city of Oslo was able to reduce energy costs by over 60 percent using IoT technology."

"A *thing*, in the Internet of Things, can be a person with a heart monitor implant, a farm animal with a biochip transponder, an automobile that has built-in sensors to alert the driver when tire pressure is low—or any other natural or man-made object that can be assigned an IP address and provided with the ability to transfer data over a network."[18]

The IoT is a massive network of connected devices, which includes people. The interconnections will be "between people-people, people-things, and things-things."[19]

Gartner predicts that *by 2020 there will be over 26 billion connected devices.*

These trends have created the ideal environment for IoT to take off and radically change work and life through technology assistance and augmentation.[20] There will be big changes in healthcare, smart agriculture, precision manufacturing, building management, energy and transportation, and more.[21]

By 2025, the global Internet of Things market will be worth $3 trillion, up four times from $750 billion in 2015.[22] In 2025, smart city services will reach 1.2 billion, more than 10 times the 115.4 million devices in 2015 worldwide.[23]

IoT as a System of Contracts

Contracts are an everyday tool used in business. But with the IoT, a digital contract can contain the terms and conditions to be carried out based on various scenarios.[24]

A digital contract is an explicit and machine-executable process between several business-entities, primarily, things and people. (Italics added for emphasis).

For instance, a new IoT-enabled refrigerator could have four types of contracts simultaneously: (1) with household members, (2) with the original manufacturer and service provider, (3) with e-commerce connections to grocery and food delivery providers, and (4) with the smart TV, cell phones, thermostat, and other things within a particular household.

> A digital contract is an explicit and machine-executable process between several business entities, primarily, things and people.

Digital contracts are standardized, machine-executable processes that already know how to handle all these complexities. Each digital contract may be executed and records/audit trails would be kept by a blockchain.

IoT Basic Risks and IG Issues

Rachel Teisch succinctly stated the risks of IoT in a post on Lexology.com, grouping them into three main categories: cybersecurity, privacy, and Information Governance:[25]

> Last year [in 2015], there were an estimated 16 billion devices connected to the Internet, and predictions say that the number will rise as high as 30 billion devices by 2020. The Internet of Things (IoT) has gained publicity (or notoriety) as yet another data source that may be subject to litigation or investigations in an e-discovery context, and there have been a few cases already (e.g. a Fitbit device has already played a prominent role in a criminal case) involving IoT data. There's likely more to come, presenting yet another major headache for corporate counsel and their legal teams, given the wealth of data stored in IoT devices.
>
> Yet, so many devices collecting data on every aspect of our lives gives rise to a number of critical issues—so many so that to call it merely the Internet of Things trivializes its importance. Rather, it can also be coined the Cybersecurity, Privacy, Information Governance, and eDiscovery of Things.

Cybersecurity for IoT

If a cybercriminal hacks into a firewall and delves into a company's benefits database, there are risks of data loss and privacy breaches. But when an Internet of Things device is breached, the ramifications can be even more dire. If a hacker intercepts an unmanned vehicle's Internet connection and sends it malicious directions, it could lead to property damage as well as severe injuries or even death. Or, if a hospital is the subject of a denial of service attack, it could compromise medical records or *even allow outsiders to manipulate medicine dosage through Internet-controlled pumps*.

Privacy for IoT

Who owns the data that these devices create? Is it the exerciser tracking steps with a fitness device, the homeowner adjusting a thermostat, or the driver of a car? In general, no. The companies collecting this data often reserve the right to access, use, and even sell their data within their terms of service agreements. And sometimes, devices are collecting data without users consenting or being aware: for example, connected streetlights and beacons in retail outlets. Depending on where the data is collected, processed, and stored, it could also implicate cross-border data protection laws.

Information Governance for IoT

With the Internet of Things, *companies have access to a number of new data streams that may be relevant for regulatory matters or lawsuits*. But because most of these devices simply gather data and send it to the cloud, it creates complicated issues relating to identifying where the data is located and then negotiating with a third-party provider regarding its retention, control, and custody.

Each of these issues has the potential to spawn litigation and regulatory actions. Recognizing the risks, legal counsel at the forefront of these challenges are taking several steps:

1. They are embracing their ethical duty to remain technologically competent by understanding the implications of these devices.
2. They are considering what technologies their organizations are deploying and what data they are collecting and storing.
3. They are ensuring the impact of the Internet of Things is contemplated in policies and procedures that address privacy, security, records management, and litigation readiness.
4. They are working with eDiscovery specialists to devise ways to preserve and collect the relevant data for litigation and investigations.

Information governance is going to be an essential element to include in planning as the IoT develops. "The sheer volume of information from Internet enabled devices is enough to bog down even the most robust systems. *The 'keep everything' approach cannot work with the oceans of data generated by Internet enabled devices*. Everything from data maps to back-up procedures is impacted. Deciding what to keep and what to eliminate is critical. Data of sufficient importance must be protected while records of low value must be disposed of intelligently. Automated classification can help sift essential data from the chaff. Additionally, policy and procedures should regularly eliminate redundant, outdated, or trivial (ROT) records. Because the amount of data is so much greater from IoT sources, **defensible deletion** becomes a governance priority."[26]

IoT E-Discovery Issues

The amount of discoverable data will drastically escalate as the IoT matures. The "IoT will lead to the discovery of everything electronic."[27]

In today's litigation environment, data stewards and litigation attorneys search for electronically stored information (ESI) from traditional sources, like computers, phones, disk drives, and USB drives to find e-mail and data relevant to a pending legal matter. But new devices and data types are constantly being introduced so litigation professionals are seemingly always developing new ways to locate and preserve data. "For example, e-discovery professionals today are working to develop collection protocols for data contained in social media platforms, SnapChat applications, and text messages. When IoT devices become the norm, legal and IT professionals will need to quickly address some of the following concerns: who is in control of the IoT device, the format of the data being generated from relevant IoT devices, and how IoT data can be cost-effectively gathered for litigation processing and review."[28]

The amount of discoverable data will drastically escalate as the IoT matures.

Once all the sources of this new ESI are located, that data must be added to the pool of discoverable information, which may include e-mails, spreadsheets, word

processing documents, and presentation files. Today, this information is typically housed in some type of online litigation repository and document review tool. In the last decade, these online repositories and review technologies have advanced quickly to help with the increasingly daunting discovery process. For example, some of these platforms even possess cutting-edge "predictive coding" features that use artificial intelligence to weigh the responsiveness of a document in a legal matter. Although e-discovery innovations are keeping pace, document review databases and e-discovery technology platforms will have to kick it into ludicrous warp-speed to deal with the immense volume of new and diverse discoverable data coming from the IoT revolution.

"Determining whether IoT data is relevant to a suit and if any privilege or privacy concerns exist will present an additional challenge. The first step will be to efficiently remove superfluous data in the growing sea of information from IoT devices. An added complexity for businesses will be to draw the line between personal data and data that is relevant to the legal matter, while sufficiently protecting private data, such as personally identifiable information or financial and health information. This will be especially challenging as the line between personal use and corporate use is blurred by IoT devices."[29]

E-Discovery Dangers

"Beyond IoT issues like data privacy and information security, there are *eDiscovery dangers* lurking beneath the surface of companies' IG programs."[30] These dangers, which are particularly acute in the context of litigation holds and data preservation, are becoming better known through industry education efforts. For example, a highly publicized session from the 2014 Georgetown Law Advanced E-Discovery Institute brought much-needed attention to the issues. During that session, speakers representing various constituencies observed that the IoT could raise any number of preservation and production challenges in the discovery phase of civil litigation. Ignatius Grande from the law firm of Hughes, Hubbard & Reed explained that the IoT was not designed to accommodate eDiscovery demands:

"Many products in the IoT sphere are not created with litigation hold, preservation, and collection in mind," he said. "In terms of liability ... companies will most likely be responsible to preserve data produced by the capabilities of their products and services in the event of a litigation hold."

This is the case: unless *appropriate measures are adopted to ensure that IoT data is preserved for litigation or regulatory matters, relevant IoT materials could be lost, setting the stage for expensive and time-consuming satellite litigation*.[31]

Security

Security is also a major issue. Organizations will have to develop new policies and processes to manage and govern IoT data. And sometimes real-time analytics will need to take place. New data processing capacities will need to be developed for this massive amount of data flowing from the IoT.

In addition to data storage concerns, the IoT explosion is driving backup and retention difficulties. With limited storage space on current IoT devices, most IoT technologies integrate with existing technologies. As such, it is likely that any IoT data also resides on another device such as a smartphone, tablet, or server, meaning that the data is probably already backed up and governed by existing corporate policies and practices. For example, smart watches need to be connected to a smartphone to access e-mail, text messages, and social media accounts. However, as IoT devices become more stand-alone, with greater capacities to store data without the assistance of another device, corporate information governance and "bring your own device" policies will need to expand to include considerations for these hypermobile IoT devices.

Lastly, IoT innovations will create a multitude of security issues, including cyberattacks, data breaches, and hacking. A recent study from HP Security Research found that 70% of Internet-connected devices are vulnerable to some form of hacking. As IoT continues to expand, there are rising concerns about the increased number of entry points for hackers into the smart home or office. The challenge will be to make IoT devices tamper-proof to ensure their physical connections cannot be modified, their operating system or firmware is unalterable, and any data they contain is void of extraction in an unencrypted form. In addition, no matter how secure the IoT infrastructure, companies will still have to pay extra attention to the security of the data center that stores and processes data that comes in from IoT equipment."[32]

Challenges Ahead

It should be abundantly clear now that *organizations need to have an actionable plan to prepare for the data privacy, information security, and e-discovery implications of the IoT*. As an initial phase in this preparation, companies should determine the extent to which the IoT affects or will affect their consumers and employees. Understanding the range of potential IoT issues will provide clarity on the next steps that should be taken.

One of those steps likely will involve the development of an IG strategy that accounts for the massive data generated from the IoT. "Such a strategy should include a plan for identifying information that must be kept for business or legal purposes while isolating other data (particularly PII) for eventual deletion. It also should encompass steps to ensure compliance with the privacy expectations of domestic and international data protection authorities. Enterprises also will need to ensure that their litigation readiness programs are updated to include a process for preserving and producing relevant IoT data."[33]

> Organizations need to have an actionable plan to prepare for the data privacy, information security, and e-discovery implications of the IoT.

Taking a proactive approach that addresses these issues will help companies avoid

many of the treacherous problems associated with the IoT. While it may not lead to smooth sailing all of the time, it will establish a process that can enable the successful disposition of IoT issues.

As your organization deploys various devices connected to the Internet, bear in mind that any and *all of the data these devices generate and exchange is discoverable in litigation and must be considered in compliance planning*.

The Federal Rules of Civil Procedure (FRCP), since the landmark 2006 changes to accommodate electronically stored information (ESI), and its update in 2015, have made it clear that "all data within the enterprise is discoverable. As businesses expand their network of connected devices and collect more and more data, there will increasingly be an expectation that the data is discoverable."[34]

Due to the variety and velocity of data from multiple device types, the collection and preservation of IoT data will be challenging.[35]

That is why you must have an IG strategy and Information Governance Framework (IGF) in place to drive and guide your IG program. Your IGF will present the guardrails or guidelines within which your organization will manage the collection and preservation of IoT data.

Why IoT Trustworthiness Is a Journey and Not a Project

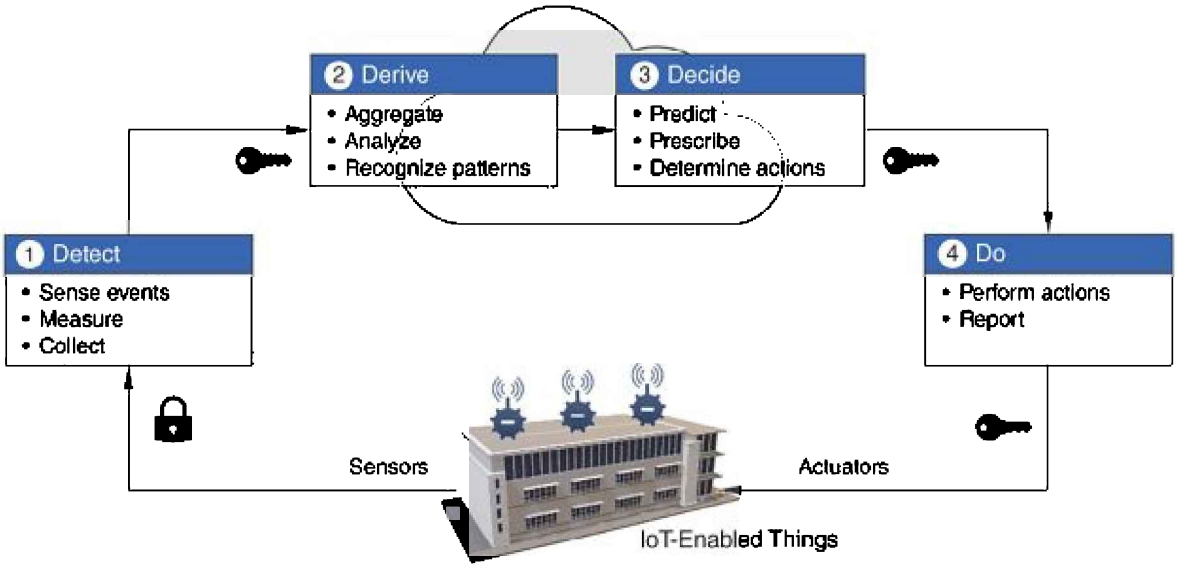
By Bassam Zarkout

Overview

We have seen several iterations of the "Internet of X" mantra over the years: The Internet of Content, the Internet of Commerce, and the Internet of People.

The most recent iteration and arguably the most significant one is the Internet of Things. Recognized as one of the key enablers for Digital Transformation, IoT is the ability to configure sensors on *things*[36] in order to capture operational data, exploit that data, gain insight about the operation of these *things*, control them, alter their behavior, and ultimately produce "better outcomes."[37]

Although IoT systems tend to be architecturally complex, the overall principle of their operation is fairly consistent: you Detect, you Derive, you Decide, and then you Do.



The 4Ds of IoT - source IGnPower

Most people associate the term IoT with consumer-oriented devices like home thermostats. But it is in various industries[38] that IoT applications have the most impact. In the last few years the number of IoT sensors[39] has grown exponentially. By 2020 that number is expected to exceed 20 billion. This means that IoT systems are destined to generate volumes of data that will dwarf the volumes of data and information generated by business systems.

As an information covernance professional, I consider that IoT data is corporate data that must be governed in accordance with legal and regulatory obligations and internal corporate policies.

As an IoT professional, I would say ... yes, but not so fast.

This article will introduce the term *IoT Trustworthiness*, an emerging domain that overlaps with IG in some areas, but is potentially much more significant to the organization.

This article is thus a *call to action* to both IoT practitioners and IG professionals:

- IoT practitioners should heed the growing governance debt that will inevitably result from the exponential growth of IoT data volumes.
- IG professionals should watch out for that incoming train called IoT and recognize the important role they are destined to play in *IoT Trustworthiness*.

Governing the IoT Data

As already stated, data produced and consumed by IoT systems should be considered as corporate data that is subject to governance controls mandated by laws, regulations, standards, and eDiscovery rules, as well as rules defined by internal policies.

Adopters of IoT solutions are facing a wide range of technical and organizational

challenges: how to cope with fast-evolving technology and architecture, how to deal with the challenges of integration, and above all how to reconcile IT with OT[40] issues and manage their convergence.

> Data produced and consumed by IoT systems should be considered as corporate data that is subject to governance controls mandated by laws, regulations, standards, and e-discovery rules, as well as internal rules.

As IoT solutions continue to expand and mature, the volume of IoT data generated by the sensors will witness exponential growth, and organizations will have to address several fundamental questions:

- What is the IoT data and who owns it?
- What are the rights of the IoT solution adopters?
- What are the obligations of the IoT solution providers[41] toward this data?
- What are the Data Protection best practices for this data?
- How long should this data be retained?
- How to deal with issues like data lineage and data residency?

Throughout my years in the IG space, I have always been struck by the years of inaction of organizations vis-à-vis their *mounting IG debt and the uphill battles IG practitioners continue to face in getting their initiatives off the ground*.

There is no question in my mind that the governance of IoT data will face similar challenges. But these challenges will be more complex here, however, due to the physical nature of these systems.

I will get into these challenges in the next section, but let me first get the "good news" out of the way:

- Governance debt for IoT data is still very low: Most IoT systems have been in production for a relatively short period of time. This means that the volume of IoT data is relatively low at the moment and the governance debt for the IoT data is still low. No time to waste here, however, since the volume of IoT data is about to explode.
- IoT data is structured and well organized: it should not be difficult to identify this data, classify it, and define governance rules for it. Adding governance frameworks to existing IoT systems to actually enforce the governance controls will require engineering efforts, but it is doable.

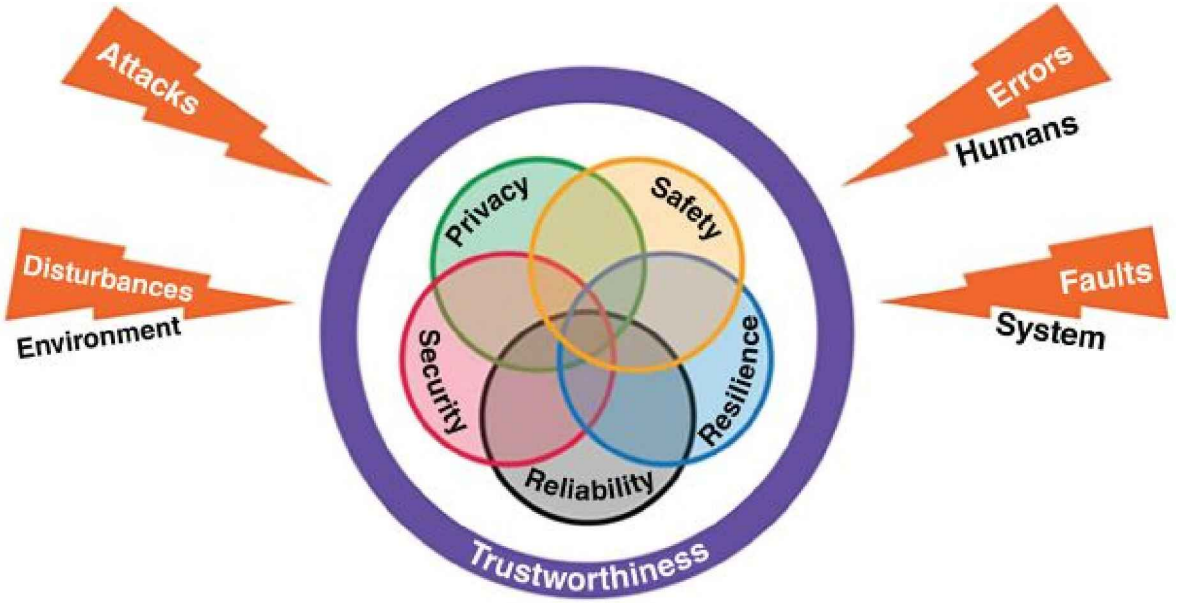
IoT Trustworthiness

The discussion about governing IoT data *cannot* be limited to data only. This is due to a very simple fact about IoT: *IoT is much more than IT for Things*.

By definition, IoT systems have a digital side and a physical side. The governance of the IT aspects of these IoT systems (security and privacy) cannot be separated from the governance of the OT aspects of these systems (safety, reliability, and resilience).

Enter the term: IoT Trustworthiness.

The Boston-based Industrial Internet Consortium or IIC[42,43] defines IoT Trustworthiness as follows:



IoT Trustworthiness - source Industrial Internet Consortium

Source: Industrial Internet Consortium.

It is the degree of confidence one has that the system performs as expected with characteristics including safety, security, privacy, reliability, and resilience in the face of environmental disturbances, human errors, system faults, and attacks.

Establishing and maintaining the trustworthiness objectives in an IoT system leads to better outcomes, such as a better alignment with the corporate business objectives, a better visibility of operational risks, and so on. On the other hand, failure to achieve and maintain the trustworthiness objectives can lead to significant negative consequences, such as serious accidents, equipment failures, data breaches, and operational interruptions to name a few.

Note: In so many IoT use cases, issues like safety and security far outweigh traditional IG concerns. For example, delaying a security patch in order not to affect production may introduce safety risks, which can lead to serious accidents where people may be physically harmed. The issues and choices that IG Professionals face in projects like shared drive cleanup of ROT[44] pale in comparison. Nobody was ever injured by duplicated documents in a shared drive.

In order to assess the overall trustworthiness state of an IoT system, one must look at the state of each of the IoT Trustworthiness characteristics: Security, Safety, Reliability, Resilience, and Privacy.

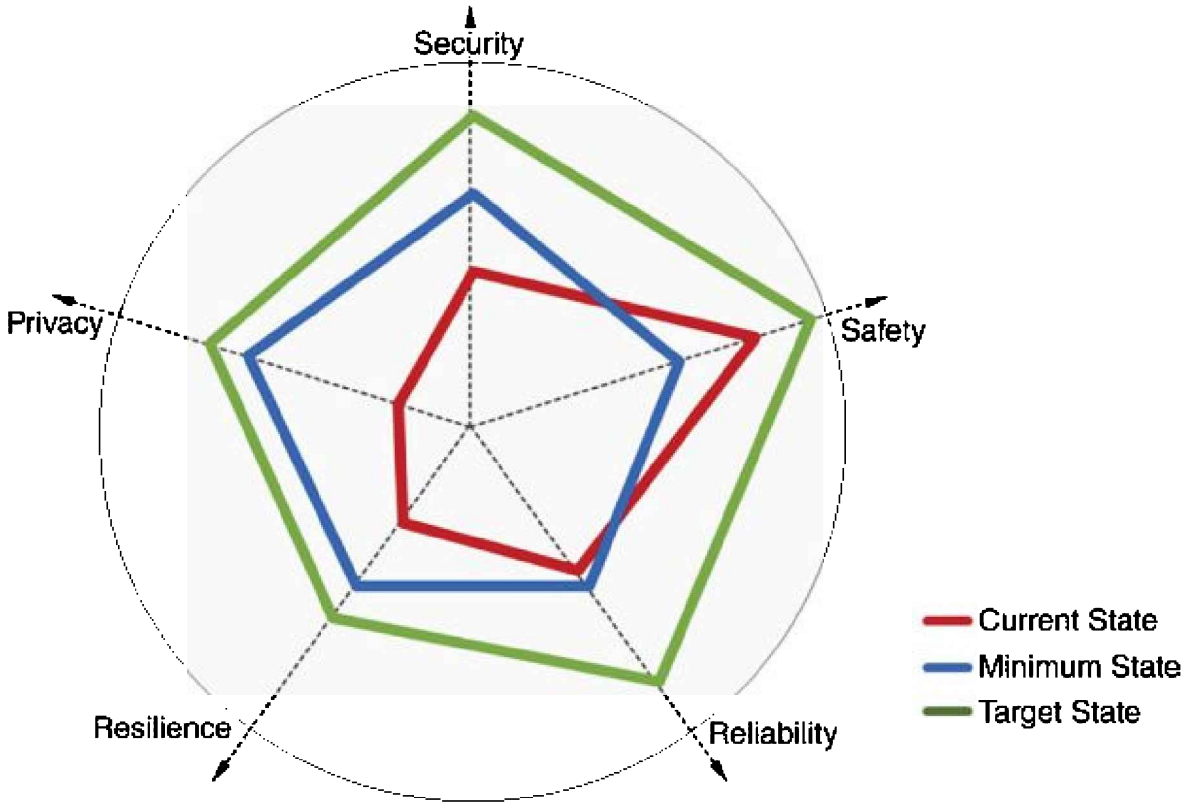
Establishing and maintaining the trustworthiness objectives in an IoT system leads to better outcomes.

For example, the *current* state of one characteristic may fall short of the *minimum* level mandated by laws and regulations for that characteristic. On the other hand, the *current* state of another characteristic may meet the *minimum* level but fall short of the *target* level set at a corporate level.

Below is a description of these *current*, *minimum,* and *target* states:

- *Current state (red)*: This is the "trustworthiness" status of the IoT system, based on how it is currently designed, implemented, and operating.
- *Minimum state (blue)*: This is a non-negotiable trustworthiness level mandated by external authorities and parties, for example, legal, regulatory, standards, and industry best practices.
- *Target state (green)*: This trustworthiness level exceeds the minimum state, and is based on internally defined and self-imposed drivers and objectives (business and technical).

The "radar map" in the diagram below provides an example of the IoT Trustworthiness states of a system. In this example, Safety exceeds the mandated minimum legal requirements while the other characteristics (Security, Reliability, Resilience, and Privacy) fall short of their respective mandated minimums and thus require efforts to become compliant.



IoT Trustworthiness Radar Diagram - source Industrial Internet Consortium

Source: Industrial Internet Consortium.

This visual view of IoT Trustworthiness will help the organization understand its current situation vis-à-vis the trustworthiness of IoT systems and prioritize the work needed to become compliant.

Information Governance Versus IoT Trustworthiness

Readers who have been trained in the art of information governance should have recognized by now that IG and IoT Trustworthiness share some similarities:

- IoT Trustworthiness may be complex as a topic, but at the end of the day IoT data is corporate data that must be governed. This data must be classified, its life cycle managed, and its e-discovery properly handled in case of litigation.
- Like IG, IoT Trustworthiness is a multifaceted discipline that requires a collaboration between multiple groups in the organization.
- Just like IG, IoT Trustworthiness needs a leader[45] who is empowered[46] to drive the trustworthiness efforts throughout the life cycle of the IoT system.

IoT Trustworthiness is also different from IG. Its scope is much wider, covering several well-established functions that have their own teams, long traditions, and mandates. Safety plays a very prominent role in IoT, and cybersecurity plays a central and enabling role in IoT and beyond (safety, privacy, etc.).

Like IG, IoT Trustworthiness is a multifaceted discipline that requires a collaboration between multiple groups in the organization.

IoT Trustworthiness Journey

IoT systems tend to have long life cycles. For example, the life cycle of a manufacturing plant and its systems may be decades long:

- During this long life cycle, some of the plant's internal systems and subsystems may be upgraded, IoT-enabled, or totally replaced.
- IoT data produced and consumed by the plant's systems may have long life cycles.
- Trustworthiness requirements for the system may change over time due to changes in laws and regulations or changes in the architecture of the system itself.

What all this means is that establishing and maintaining the system's trustworthiness is not a project. It is an effort that must be sustained throughout the life cycle journey of the system (diagram below):

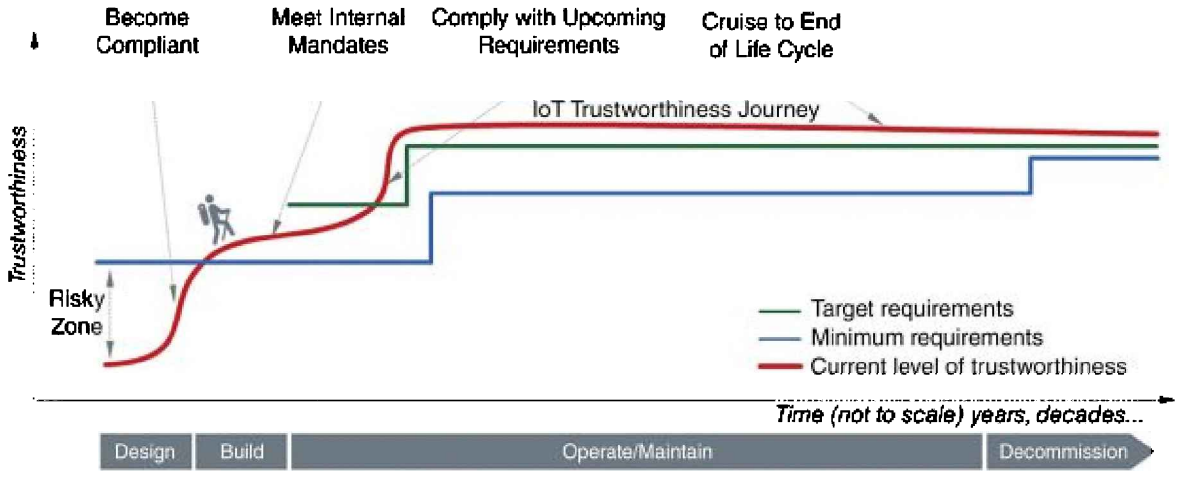
The IoT Trustworthiness journey must be piloted by a program that acts as a framework for organizing, directing, implementing, and maintaining trustworthiness of an IoT system throughout its life cycle, and in accordance with established Corporate Business Objectives.

Similar to the Information Governance program within the organization, the IoT Trustworthiness program must have a corporate sponsor to set the mandate and empower the organization to achieve that mandate, a program tsar to lead and manage the program, and a steering committee for the stakeholders who will coordinate the cross-functional implementation of the various facets of trustworthiness.

The program must also deliver real value to the organization in the form of “better outcomes.” This value must be communicated to the various groups and stakeholders in the organization in terms they relate to and understand.

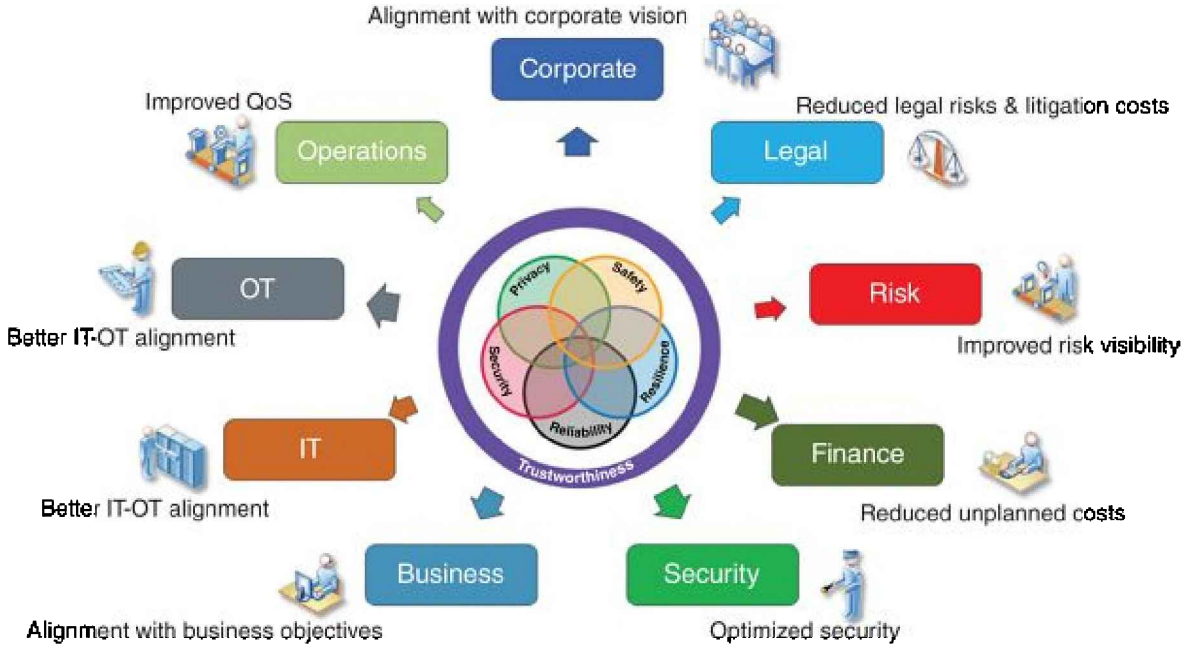
A lot to unpack here and perhaps I should dedicate a chapter in the future to the subject of the IoT Trustworthiness Program, its structure and its activities.

Suffice it to say that a core component of the initial stages of this program is an assessment of the *current* state of the IoT system and a determination of the *minimum* state based on external drivers like laws and regulations, and the desired *arget* state based on internal drivers like corporate strategy.



The IoT Trustworthiness Journey - source IGnPower

Source: IgN Power.



Value delivered by IoT Trustworthiness Program - source IGnPower

Source: IgN Power.

IG professionals have an important role to play in this regard.

Conclusion

The trustworthiness of IoT systems and the governance of their IoT data are key to ensuring that these systems can deliver on their intended objectives. Both efforts should be maintained throughout the full life cycle journey of the IoT systems and their IoT data.

There is little time to waste here as IoT technologies and architectures are evolving fast. AI and distributed ledger technologies like blockchain are starting to play

central roles within IoT systems. Issues like AI ethics (why did the AI make this versus that decision) and the seemingly irreconcilable conflicts between blockchain and privacy (for example, GDPR's right-to-forget) are coming to the forefront.

Terms like *safety-by-design*, *security-by-design,* and *privacy-by-design* are not mere catchy buzzwords. They have a significant impact on the success of IoT systems and ultimately on the Digital Transformation strategies of organizations. These terms must be understood and the principles behind them weaved into the fabric of the IoT systems.

To close, I think it is safe to say that the need to govern IoT data is real and looming ... it is also inescapable. But it is part of a wider conversation in which issues related to the trustworthiness of IoT systems will dominate the conversation.

Again, IG professionals will have an important role to play in all this.

CHAPTER SUMMARY: KEY POINTS

- The use of analytics is key to the success of IG programs.
- There are four main types of analytics: descriptive, diagnostic, predictive, and prescriptive. Descriptive and diagnostic approaches are reactive; predictive and prescriptive are proactive.
- Unstructured information is stored rather haphazardly; applying file analysis can help to insert metadata to organize unstructured information.
- IG programs that deploy analytics help to unlock the value of unstructured information
- Analytics can help improve compliance functions by tracking and mapping information flows and communications.
- Analytics can provide insights for records managers as to which files are records and which are duplicates or older versions.
- AI and file analysis can make compliance tasks easier to execute.
- Predictive coding software “learns” to find information relevant in legal matters through initial guidance by a human expert.
- A blockchain is a decentralized, distributed ledger where transactions are secured through cryptography and immutable.
- With blockchain's **decentralized architecture,** there is no centralized control over who can participate in the ledger.
- Blockchain is considered a “trustless” technology: there is no human or institutional intervention necessary to verify transactions.
- People are always a security threat; blockchains are no exception to that rule.

- Blockchains fit best when a shared database is needed, there is low trust between parties, there is a need for disintermediation, and there are relationships between transactions.
- Blockchain is fundamentally a technology that addresses a social problem—trust.
- Blockchains are rarely a good solution for information assets within an organization, as trust between parties is not a major issue.
- Stated simply, the IoT is the concept of connecting any device with an "on and off switch" to the Internet (and/or to each other).
- IoT is an ecosystem of interconnected sensory devices performing coordinated, preprogrammed—or even learned—tasks without the need for continuous human input.
- The onslaught of the IoT makes it difficult for organizations to develop successful, practical IG policies and practices.
- As the IoT develops, the deluge of discoverable data in legal actions will dwarf what litigation teams must manage today.
- A digital contract is an explicit and machine-executable process between several business entities, primarily things and people.
- The amount of discoverable data will drastically escalate as the IoT matures.
- Organizations need to have an actionable plan to prepare for the data privacy, information security, and e-discovery implications of the IoT.
- Data produced and consumed by IoT systems should be considered as corporate data that is subject to governance controls mandated by laws, regulations, standards, and e-discovery rules, as well as internal rules.
- Establishing and maintaining the trustworthiness objectives in an IoT system leads to better outcomes.
- Like IG, IoT trustworthiness is a multifaceted discipline that requires a collaboration between multiple groups in the organization.

Notes

1. Sam Fossett, "The Role of Analytics in IG Programs," *Information Governance World* (Fall 2018), 47.
2. C. David Naylor, "On the Prospects for a (Deep) Learning Health Care System," *JAMA* 320, no. 11 (2018): 1099–1100. doi:10.1001/jama.2018.11103.
3. "AI And Healthcare: A Giant Opportunity," *Forbes*, February 11, 2019, https://www.forbes.com/sites/insights-intelai/2019/02/11/ai-and-healthcare-

a-giant-opportunity/#599177fd4c68.

4. J. H. Tibbetts, "The Frontiers of Artificial Intelligence," *BioScience* 68, no. 1 (2018): 5–10, https://doi-org.wgu.idm.oclc.org/10.1093/biosci/bix136.

5. Fei Jiang et al., 2017. "Artificial Intelligence in Healthcare: Past, Present and Future." *Stroke and Vascular Neurology* 2, no. 4 (2017): 230–243.

6. Ibid.

7. Reproduced with permission from Electronic Commerce & Law Report, 20 ECLR 562, April 15, 2015. Copyright © 2015 by The Bureau of National Affairs, Inc. (800-372-1033), www.bna.com.

8. Ibid.

9. Michele C. S. Lange, "E-discovery and the Security Implications of the Internet of Things," April 13, 2015, https://sm.asisonline.org/Pages/Ediscovery-and-the-Security-Implications-of-the-Internet-of-Things.aspx.

10. Reproduced with permission from Electronic Commerce & Law Report, 20 ECLR 562, April 15, 2015. Copyright © 2015 by The Bureau of National Affairs, Inc. (800-372-1033) www.bna.com.

11. http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT (accessed September 16, 2016).

12. Ibid.

13. European Research Cluster and the Internet of Things, www.internet-of-things-research.eu/about_iot.htm (accessed December 24, 2018).

14. Jacob Morgan, "A Simple Explanation of 'The Internet of Things,'" *Forbes*, May 13, 2014, www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#2440d0e86828.

15. http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT (accessed August 23, 2016).

16. Morgan, "A Simple Explanation of 'The Internet of Things.'"

17. Ben Rossi, "Why the Internet of Things Is More than Just a Smart Fridge," *InformationAge*, September 22, 2014, http://www.information-age.com/technology/mobile-and-networking/123458485/why-internet-things-more-just-smart-fridge.

18. http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT (accessed August 23, 2016).

19. Morgan, "A Simple Explanation of 'The Internet of Things.'"

20. Ibid.

21. http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT (accessed August 23, 2016).

22. Machina Research, IoT Global Forecast & Analysis 2015–25, August 5, 2016, https://machinaresearch.com/report/iot-global-forecast-analysis-2015-25/.

23. IHS Markit, "Smart City Devices to Top 1 Billion Units in 2025, IHS Says," May 24, 2016, http://press.ihs.com/press-release/technology/smart-city-devices-top-1-billion-units-2025-ihs-says.

24. Alexander Samarin, "The IoT as a System of Contracts," August 6, 2015, https://www.linkedin.com/pulse/iot-system-digital-contracts-thanks-blockchain-bpm-ecm-samarin.

25. http://www.lexology.com/library/detail.aspx?g=594bb08c-e31d-4570-a4e7-bc2517d93e83 (accessed April 15, 2016).

26. http://www.sherpasoftware.com/blog/information-governance-and-the-internet-of-things/ (accessed August 23, 2016).

27. Michele C. S. Lange, "E-discovery and the Security Implications of the Internet of Things," April 13, 2015. https://sm.asisonline.org/Pages/Ediscovery-and-the-Security-Implications-of-the-Internet-of-Things.aspx.

28. Ibid.

29. Ibid.

30. Philip Favro, "IoT Data Collection Raises Legal Ediscovery Questions, Data Informed.com, May 21, 2015, http://data-informed.com/iot-data-collection-raises-legal-ediscovery-questions/.

31. Ibid.

32. Ibid.

33. Ibid.

34. www.insidecounsel.com/2016/04/28/will-the-iot-become-the-ediscovery-of-things.

35. Ibid.

36. Automotive, aerospace, machines in plants, agricultural equipment, city lights, elevators, and so on.

37. New business models, enhanced productivity, and so on.

38. Manufacturing, cities, transportation, retail, agriculture, healthcare, and so on.

39. IR sensors, image sensors, motion sensors, accelerometer sensors, temperature sensors, and so on.

40. Operational Technology such as SCADA systems and ICS.

41. The Data Controllers and Processors in the GDPR terminology.

42. https://www.iiconsortium.org/.

43. https://www.iiconsortium.org/news/joi-articles/2018-Sept-IoT-Trustworthiness-is-a-Journey_IGnPower.pdf.

44. Removal of redundant, obsolete, and trivial content from corporate shared drives.

45. Gartner recommends the appointment of a Chief Data Officer to own the Information Governance function.

46. Empowered with authority and budgets.

PART FIVE

Long-Term Program Issues

CHAPTER 17
Long-Term Digital Preservation*

By Charles M. Dollar and Lori J. Ashley

Every organization—public, private, or not-for-profit—now has electronic records and digital content that it wants to access and retain for periods in excess of 10 years. This may be due to regulatory or legal reasons, a desire to preserve organizational memory and history, or entirely for operational reasons. But *long-term continuity of digital information does not happen by accident*—it takes information governance (IG), planning, sustainable resources, and a keen awareness of the information technology (IT) and file formats in use by the organization, as well as evolving standards and computing trends.

Defining Long-Term Digital Preservation

Information is universally recognized as a key asset that is essential to organizational success. Digital information, which relies on complex computing platforms and networks, is created, received, and used daily to deliver services to citizens, consumers and customers, businesses, and government agencies. Organizations face tremendous challenges in the twenty-first century to manage, preserve, and provide access to electronic records for as long as they are needed.

Digital preservation is defined as long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required to be retained. *Digital preservation applies to content that is born digital as well as content that is converted to digital form.*

Some digital information assets must be preserved permanently as part of an organization's documentary heritage. Dedicated repositories for historical and cultural memory, such as libraries, archives, and museums, need to move forward to put in place trustworthy digital repositories that can match the security, environmental controls, and wealth of descriptive metadata that these institutions have created for analog assets (such as books and paper records). Digital challenges associated with records management affect all sectors of society—academic, government, private, and not-for-profit enterprises—and ultimately all citizens of all developed nations.

Digital preservation is defined as long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span that the information is required to be retained.

The term “preservation” implies permanence, but it has been found that electronic

records, data, and information that is retained for only 5 to 10 years is likely to face challenges related to storage media failure and computer hardware/software obsolescence. A useful point of reference for the definition of "long term" comes from the International Organization for Standardization (ISO) standard 14721, which defines long-term as "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely."[1]

> Organizational capabilities for properly ensuring access to authentic electronic records over time (in addition to the challenges of technological obsolescence) is a sophisticated combination of policies, strategies, processes, specialized resources, and adoption of standards.

Long-term records are common in many different sectors, including government, health care, energy, utilities, engineering and architecture, construction, and manufacturing. During the course of routine business, thousands or millions of electronic records are generated in a wide variety of information systems. Most records are useful for only a short period of time (up to seven years), *but some may need to be retained for long periods or permanently*. For those records, organizations must plan for and allocate resources for preservation efforts to ensure that the data remains accessible, usable, understandable, and trustworthy over time.

In addition, *there may be the requirement to retain the metadata associated with records even longer than the records themselves.*[2] A record may have been destroyed according to its scheduled disposition at the end of its life cycle, but the organization still may need its metadata to identify the record, its life cycle dates, and the authority or person who authorized its destruction.

Key Factors in Long-Term Digital Preservation

Some electronic records must be preserved, protected, and monitored over long periods of time to ensure they remain authentic, complete, and unaltered and available into the future. Planning for the proper care of these records is a component of an overall records management program and should be integrated into the organization's **information governance** (IG) policies and technology portfolio as well as its privacy and security protocols.

Enterprise strategies for sustainable and trustworthy digital preservation repositories have to take into account several prevailing and compound conditions: the complexity of electronic records, decentralization of the computing environment, obsolescence and aging of storage media, massive volumes of electronic records, and software and hardware dependencies.

Most records are useful for only a short period of time, but some may need to be retained for long periods or permanently.

The challenges of managing electronic records significantly increased with the trend of decentralization of the computing environment. In the centralized environment of a mainframe computer, prevalent from the 1960s to 1980s but also in use today, it is relatively easy to identify, assess, and manage electronic records. This is not the case in the decentralized environment of specialized business applications and office automation systems, where each user creates electronic objects that may constitute a formal record and thus will have to be preserved under IG policies that address record retention and disposition rules, processes, and accountability.

Electronic records have evolved from simple text-based word processing files or reports to include complex mixed media digital objects that may contain embedded images (still and animated), drawings, sounds, hyperlinks, or spreadsheets with computational formulas. Some portions of electronic records, such as the content of dynamic Web pages, are created on demand from databases and exist only for the duration of the viewing session. Other digital objects, such as electronic mail, may contain multiple attachments, and they may be threaded (i.e. related e-mail messages linked in send-reply chains). These records cannot be converted to paper or text formats for preservation without the loss of context, functionality, and metadata.

Electronic records are being created at rates that pose significant threats to our ability to organize, control, and make them accessible for as long as they are needed. This accumulating volume of digital content includes documents that are digitally scanned or imaged from a variety of formats to be stored as electronic records.

Electronic records are stored as representations of bits—1s and 0s—and therefore depend on software applications and hardware networks for the entire period of retention, whether it is 3 days, 3 years, or 30 years or longer. As information technologies become obsolete and are replaced by new generations, the capability of a specific software application to read the representations of 1s and 0s and render them into human-understandable form will degrade to the point that the records are neither readable nor understandable. As a practical matter, this means that the readability and understandability of the records can never be recovered, and there can be serious legal consequences.

Electronic records are being created at rates that pose significant threats to our ability to organize, control, and make them accessible for as long as they are needed.

Storage media are affected by the dual problems of obsolescence and decay. They are fragile, have limited shelf life, and become obsolete in a matter of a few years. *Mitigating media obsolescence is critical to long-term digital preservation* (LTDP) because the bitstreams of 1s and 0s that comprise electronic records must be kept "alive" through periodic transfer to new storage media.

In addition to these current conditions associated with technology and records management, organizations face tremendous internal **change management** challenges with regard to reallocation of resources, business process improvements, collaboration and coordination between business areas, accountability, and the dynamic integration of evolving recordkeeping requirements. Building and sustaining the capability to manage digital information over long periods of time is a shared responsibility of all stakeholders.

Threats to Preserving Records

A number of known threats may degrade or destroy electronic records and data:

- *Failure of storage media*. Storage media is inherently vulnerable to errors and malfunction, including disk crashes. Solid-state drives (SSD) largely address these concerns as there are no moving parts, and data can be stored without needing electrical power.
- *Failure of computer systems*. Computer hardware has moving parts and circuits that deteriorate and fail over time, at an average rate called mean time between failures. Some failures are complete and irrecoverable, and some are minor and can be fixed with no loss of data. Computer software is prone to bugs and malware that can compromise the safekeeping of data.
- *Systems and network communications failures*. A small number of network communications is likely to contain errors or misreads, especially undetected checksum errors, which may impact the authenticity of a record. Network errors can occur from changes or redirection of URLs, and any communication over a network is subject to intrusions, errors, and hackers.
- *Component obsolescence*. As hardware, software, and media age, they become obsolete over time, due to the continued innovation and advances by the computer industry. Sometimes obsolescence is due to outdated component parts, changes in software routines, or changes in the hardware to read removable media.
- *Human error*. People make mistakes, and they can make mistakes in selecting, classifying, storing, or handling archived records. Some of these errors may be detected and can be remedied; some go unnoticed or cannot be fixed.
- *Natural disaster*. Hurricane Katrina is the clearest US example of how a natural disaster can interrupt business operations and destroy business records, although in some instances, damaged records were able to be recovered. Floods, fires, earthquakes, and other natural disasters can

completely destroy or cause media or computer hardware/software failures.

- *Attacks*. Archived electronic records are subject to external attacks from malware, such as viruses and worms, so preserved records must be scanned for malware and kept separate from external threats. Preserved records also can be subject to theft or damage from insiders, such as the theft of historical radio recordings by a National Archives and Records Administration employee, which was reported in 2012. Proper monitoring and auditing procedures must be in place to detect and avoid these types of attacks.
- *Financial shortfall*. It is expensive to preserve and maintain digital records. Power, cooling and heating systems, personnel costs, and other preservation-associated costs must be budgeted and funded.
- *Business viability*. If an organization has financial or legal difficulties or suffers a catastrophic disaster, it may not survive, placing the preserved records at risk. Part of the planning process is to include consideration of successor organization alternatives, should the originating organization go out of business.[3]

Threats to LTDP of records can be internal or external, from natural disasters, computer or storage failures, and even from the financial viability of an organization.

The impact on the preserved records can be gauged by determining what percentage of the data has been lost and cannot be recovered or, for the data that can be recovered, what the impact or delay to users may be.

It should be noted that threats can be interrelated and more than one type of threat may impact records at a time. For instance, in the event of a natural disaster, operators are more likely to make mistakes, and computer hardware failures can create new software failures.

Digital Preservation Standards

The digital preservation community recognizes that open standard technology-neutral standards play a key role in ensuring that digital records are usable, understandable, and reliable for as far into the future as may be required.

There are two broad categories of digital preservation standards. The first category involves systems infrastructure capabilities and services that support a trustworthy repository. The second category relates to open standard technology-neutral file formats.

Digital preservation infrastructure capabilities and services that support trustworthy digital repositories include the international standard ISO 14721:2003, 2012 Space Data and Information Transfer Systems—Open Archival Information System (**OAIS**)—Reference Model, which is a key standard applicable

to LTDP.[4]

The fragility of digital storage media in concert with ongoing and sometimes rapid changes in computer software and hardware poses a fundamental challenge to ensuring access to trustworthy and reliable digital content over time. Eventually, every digital repository committed to LTDP must have a strategy to mitigate computer technology obsolescence. Toward this end, the Consultative Committee for Space Data Systems developed an Open Archival Information System (OAIS) reference model to support formal standards for the long-term preservation of space science data and information assets. OAIS was not designed as an implementation model.

The OAIS Reference Model defines an archival information system as an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available and understandable for a designated community (i.e. potential users or consumers), who should be able to understand the information. Thus, the context of an OAIS-compliant digital repository includes producers who originate the information to be preserved in the repository, consumers who retrieve the information, and a management/organization that hosts and administers the digital assets being preserved.

OAIS encapsulates digital objects into information packages. Each information package includes the digital object content (a sequence of bits) and representation information that enables rendering of an object into human usable information along with **preservation description information** (PDI) such as provenance, context, and fixity.

The OAIS Information Model employs three types of information packages: a **submission information package** (SIP), an **archival information package** (AIP), and a **dissemination information package** (DIP). An OAIS-compliant digital repository preserves AIPs and any PDI associated with them. A SIP encompasses digital content that a producer has organized for submission to the OAIS. After the completion of quality assurance and transformation procedures, an AIP is created, which is the focus of preservation activity. Subsequently, a DIP is created that consists of an AIP or information extracted from an AIP customized to the requirements of the designated community of users and consumers.

The core of OAIS is a functional model that consists of six entities:

1. *Ingest* processes the formal incorporation (in archival terms, *accession*) of submitted information (i.e., a SIP) into the digital repository. It acknowledges the transfer, conducts quality assurance, extracts metadata from the SIP, generates the appropriate AIP, and populates PDI and extracted metadata into the AIP.
2. *Archival storage* encompasses all of the activities associated with storage of AIPs. They include receipt of AIPs, transferring AIPs to the appropriate storage location, replacing media as necessary, transforming AIPs to new file formats as necessary, conducting quality assurance tests, supporting backups

and business continuity procedures, and providing copies of AIPs to the access entity.

3. *Data management* manages the storage of description and system information, generates reports, and tracks use of storage media.
4. *Administration* encompasses a host of technical and human processes that include audit, policy making, strategy, and provider and customer service, among other management and business functions. OAIS administration connects with all of the other OAIS functions.
5. *Preservation planning* does not execute any preservation activities. Rather, it supports a technology watch program for sustainable standards, file formats, and software for digital preservation, monitoring changes in the access needs of the designated community, and recommending updated digital preservation strategies and activities.
6. *Access* receives queries from the designated community, passes them to archival storage, and makes them available as DIPs to the designated community.

Figure 17.1 displays the relationships between these six functional entities.[5]

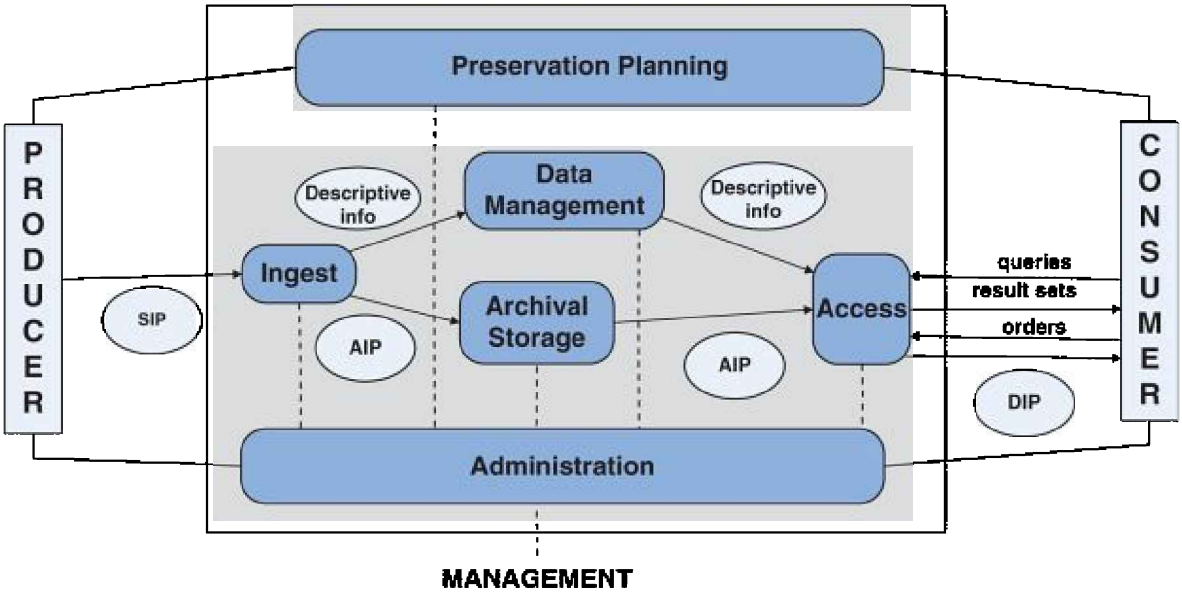


Figure 17.1 Open Archival Information System Reference Model

In archival storage, the OAIS reference model articulates a migration strategy based on four primary types of AIP migration that are ordered by an increasing risk of potential information loss: refreshment, replication, repackage, and transformation.[6]

1. *Migration refreshment* occurs when one or more AIPs are copied exactly to the same type of storage media with no alterations occurring in the packaging information, the content information, the preservation description information (PDI), or the AIP location and access archival storage mapping

infrastructure.

2. *Migration replication* occurs when one or more AIPs are copied exactly to the same or new storage media with no alterations occurring in the packaging information, the content information, and the PDI. However, there is a change in the AIP location and access archival storage mapping infrastructure.
3. *Migration repackage* occurs when one or more AIPs are copied exactly to new storage media with no alterations in the content information and the PDI. However, there are changes in the packaging information and the AIP location and to the access to the archival storage mapping infrastructure.
4. *Migration transformation* occurs when changes in bitstreams result when a new content encoding procedure replaces the current encoding procedure (e.g. Unicode representation of A through Z replaces the ASCII representation of A through Z), a new file format replaces an existing one, or a new software application is required to access and render the AIP content.

OAIS is the lingua franca of digital preservation. The international digital preservation community has embraced it as the framework for viable and technologically sustainable digital preservation repositories. *An LTDP strategy that is OAIS-conforming offers the best means available today for preserving the digital heritage of all organizations, private and public.*

> An OAIS-conforming LTDP strategy is the best way to preserve an organization's digital heritage.

ISO TR 18492 (2005), Long-Term Preservation of Electronic Document-Based Information

ISO 18492 provides practical methodological guidance for the long-term preservation and retrieval of authentic electronic document-based information, when the retention period exceeds the expected life of the technology (hardware and software) used to create and maintain the information assets. It emphasizes both the role of open standard technology–neutral standards in supporting long-term access and the engagement of IT specialists, document managers, records managers, and archivists in a collaborative environment to promote and sustain a viable digital preservation program.

ISO 18492 takes note of the role of ISO 15489 but does not cover processes for the capture, classification, and disposition of authentic electronic document-based information. Ensuring the usability and trustworthiness of electronic document-based information for as long as necessary in the face of limited media durability and technology obsolescence requires a robust and comprehensive digital preservation strategy. ISO 18492 describes such a strategy, which includes media renewal, software dependence, migration, open standard technology-neutral formats, authenticity protection, and security:

- *Media renewal*. ISO 18492 defines media renewal as a baseline requirement for digital preservation because it is the only known way to keep bitstreams of information based on electronic documents alive. It specifies the conditions under which copying and reformatting of storage media and storage devices should occur.
- *Open standard technology-neutral formats*. The fundamental premise of ISO 18492 is that open standard technology-neutral formats are at the core of a viable and technologically sustainable digital preservation strategy because they help mitigate software obsolescence. ISO 18492 recommends the use of several standard formats, including: eXtensible Markup Language (XML), Portable Document Format/Archival (PDF/A), tagged image file format (TIFF), and Joint Photographic Experts Group (JPEG).
- *Migrating electronic content*. ISO 18492 recommends two ways of migrating electronic content to new technologies. The first relies on backwardly compatible new open standard technology-neutral formats that are displacing existing ones. Generally, this is a straightforward process that typically can be executed with minimal human intervention. The second involves writing computer code that exports the electronic content to a new target application or open standard technology-neutral format. This can be a very labor-intensive activity and requires rigorous quality control.
- *Authenticity*. ISO 18492 recommends the use of hash digest algorithms to validate the integrity of electronic content after execution of media renewal activities that do not alter underlying bit streams of electronic content. In instances where bitstreams are a result of format conversion, comprehensive preservation metadata should be captured that documents the process.
- *Security*. ISO 18492 recommends protecting the security of electronic records by creating a firewall between electronic content in a repository and external users. In addition, procedures should be in place to maintain backup/disaster recovery capability, including at least one off-site storage location.

ISO 18492 provides practical methodological guidance for the long-term preservation of e-documents when the retention period exceeds the expected life of the technology that created it.

ISO 16363 (2012)—Space Data and Information Transfer Systems—Audit and Certification of Trustworthy Digital Repositories

ISO 14721 (OAIS) acknowledged that an audit and certification standard was needed that incorporated the functional specifications for records producers, records users, ingest of digital content into a trusted repository, archival storage of this content, and digital preserving planning and administration. *ISO 16363 is this audit and certification standard.* Its use enables independent audits and certification of trustworthy digital repositories and thereby promotes public trust

in digital repositories that claim they are trustworthy. To date only a handful of ISO 16363 test audits have been undertaken; additional time is required to determine how widely adopted the standard becomes.

ISO 16363 is organized into three broad categories: organization infrastructure, digital object management, and technical infrastructure and security risk management. Each category is decomposed into a series of primary elements or components, some of which may be more appropriate for digital libraries than for public records digital repositories. In some instances there are secondary elements or components. An explanatory discussion of each element accompanies "empirical metrics" relevant to that element. The empirical metrics typically include high-level examples of how conformance can be demonstrated. Hence, they are subjective high-level conformance metrics rather than explicit performance metrics.

ISO 16363 is an audit and certification standard organized into three broad categories: organization infrastructure, digital object management, and technical infrastructure and security risk management.

Organizational infrastructure[7] consists of these primary elements:

- *Mission statement* that reflects a commitment to the preservation of, long-term retention of, management of, and access to digital information
- *Preservation strategic plan* that defines the approach the repository will take in the long-term support of its mission
- *Collection policy* or other document that specifies the types of information it will preserve, retain, manage, and provide access to
- *Identification and establishment of the duties identified* and establishment of the duties and roles that are required to perform along with a staff with adequate skills and experience to fulfill these duties
- *Dissemination of the definitions* of its designated community and associated knowledge base(s)
- *Preservation policies* that ensure that the preservation strategic plan will be met
- *Documentation* of the history of changes to operations, procedures, software, and hardware
- *Commitment to transparency and accountability* in all actions supporting the operation and management of the repository that affect the preservation of digital content over time
- *Dissemination* as appropriate of the definition, collection, and tracking of information integrity measurements

- *Commitment to a regular schedule of self-assessment* and external certification
- *Short- and long-term business planning* processes in place to sustain the repository over time
- *Deposit agreements* for digital materials transferred to the custody of the organization
- *Written policies* that specify when the preservation responsibility for contents of each set of submitted data objects occurs
- *Intellectual property ownership rights* policies and procedures

Digital object management,[8] *which is the core of the standard*, comprises these primary elements:

- Methods and factors used to determine the different types of information for which an organization accepts preservation responsibility
- An understanding of digital collections sufficient to carry out the preservation necessary for as long as required
- Specifications that enable recognition and parsing of SIPs
- An ingest procedure that verifies each SIP for completion and correctness
- An ingest procedure that validates successful ingest of each SIP
- Definitions for each AIP or class of AIPs used that are adequate for parsing and suitable for long-term preservation requirements
- Descriptions of how AIPs are constructed from SIPs, including extraction of metadata
- Documentation of the final disposition of SIPs, including those not ingested
- A convention that generates unique, persistent identifiers of all AIPs
- Reliable linking services that support the location of each uniquely identified object, regardless of its physical location
- Tools and resources that support authoritative representation information for all of the digital objects in the repository, including file type
- Documented processes for acquiring and creating PDI
- Understandable content information for the designated community at the time of creation of the AIPs
- Verification of the completeness and correctness of AIPs at the point of their creation
- Contemporaneous capture of documentation of actions and administration processes that are relevant to AIP creation
- Documented digital preservation strategies

- Mechanisms for monitoring the digital preservation environment
- Documented evidence of the effectiveness of digital preservation activities
- Specifications for storage of AIPs down to the bit level
- Preservation of the content information of AIPs
- Monitoring the integrity of AIPs
- Documentation that preservation actions associated with AIPs complied with the specifications for those actions
- Specification of minimum information requirements that enable the designated community to discover and identify material of interest
- Bidirectional linkage between each AIP and its associated descriptive information
- Compliance with access policies
- Policies and procedures that enable the dissemination of digital objects that are traceable to the "originals," with evidence supporting their authenticity
- Procedures that require documentation of actions taken in response to reports about errors in data or responses from users

Technical infrastructure and security risk management primary elements[9] include these elements:

- Technology watches or other monitoring systems that track when hardware and software is expected to become obsolete
- Procedures, commitment, and funding when it is necessary to replace hardware
- Procedures, commitment, and funding when it is necessary to replace software
- Adequate hardware and software support for backup functionality sufficient for preserving the repository content and tracking repository functions
- Effective mechanisms that identify bit corruption or loss
- Documentation captures of all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data
- Defined processes for storage media and/or hardware change (e.g. refreshing, migration)
- Management of the number and location of copies of all digital objects
- Systematic analysis of security risk factors associated with data, systems, personnel, and physical plant
- Suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s)

ISO 16363 represents the gold standard of audit and certification for trustworthy digital repositories. In some instances the resources available to a trusted repository may not support full implementation of the audit and certification specifications. Decisions about where full and partial implementation is appropriate should be based on a risk assessment analysis.

ISO 16363 represents the gold standard of audit and certification for trustworthy digital repositories.

PREMIS Preservation Metadata Standard

ISO 14721 specifies that preservation metadata associated with all archival storage activities (e.g. generation of hash digests, transformation, and media renewal) should be captured and stored in PDI. *This high-level guidance requirement demands greater specificity in an operational environment*.

Toward this end, the US Library of Congress and the Research Library Group supported a new international working group called PREservation Metadata Information Strategies (PREMIS)[10] to define a core set of preservation metadata elements with a supporting data dictionary that would be applicable to a broad range of digital preservation activities and to identify and evaluate alternative strategies for encoding, managing, and exchanging preservation metadata. Version 2.2 was released in June 2012.[11]

PREMIS enables designers and managers of digital repositories to have a clear understanding of the information required to support the "functions of viability, renderability, understandability, authenticity, and identity in a preservation context." PREMIS accomplishes this through a data model that consists of five "semantic units" (think of them as high-level metadata elements, each of which is decomposed into subelements) and a data dictionary that decomposes these "semantic units" into a structure hierarchy. The five semantic units and their relationships are displayed in Figure 17.2.

Note the arrows that define relationships between these entities:

- *Intellectual entities* are considered a single intellectual unit such as a book, map, photograph, database, or records (e.g. an AIP).
- *Objects* are discrete units of information in digital form that may exist as a bitstream, a file, or a representation.
- *Events* denote actions that involve at least one digital object and/or agent known to the repository. Events may include the type of event (e.g. media renewal), a description of the event, and the agents involved in the event. Events support the chain of custody of digital objects.

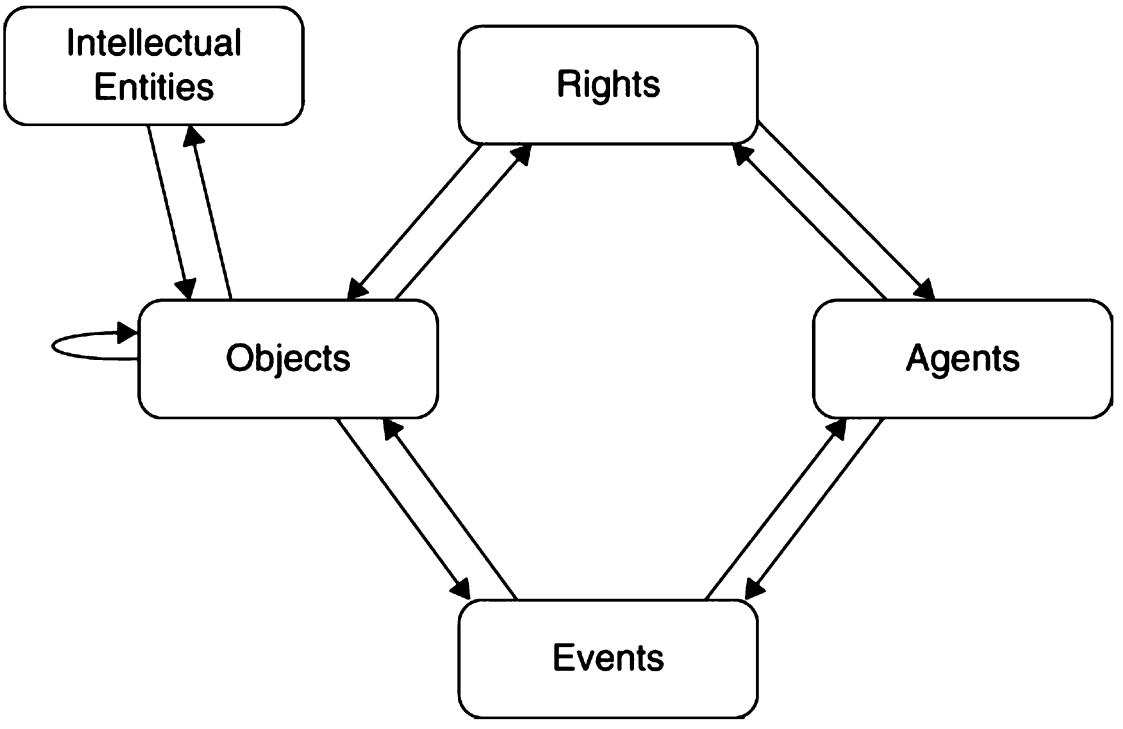


Figure 17.2 PREMIS Data Model

Source: Library of Congress, *PREMIS Data Dictionary Version 2.2: Hierarchical Listing of Semantic Units*, September 13, 2012, www.loc.gov/standards/premis/v2/premis-dd-Hierarchical-Listing-2-2.html.

- *Agents* are actors in digital preservation that have roles. An agent can be an individual, organization, or a software application.
- *Rights* involve the assertion of access rights and access privileges that relate to intellectual property, privacy, or other related rights

The PREMIS Data Dictionary decomposes objects, events, agents, and rights into a structured hierarchical schema. In addition, it contains semantic units that support documentation of relationships between Objects. An important feature of the PREMIS is an XML schema for the PREMIS Data Dictionary. The primary rationale for the XML schema is to support the exchange of metadata information, which is crucial in ingest and archival storage. The XML schema enables automated extraction of preservation related metadata in SIPs and population of this preservation metadata into AIPs. In addition, the XML schema can enable automatic capture of preservation events that are foundational for maintaining a chain of custody in archival storage.

The PREMIS standard defines a core set of preservation metadata elements with a supporting data dictionary applicable to a broad range of digital preservation activities.

Recommended Open Standard Technology–Neutral Formats

A digital file format specifies the internal logical structure of digital objects (i.e. binary bits of 1s and 0s) and signal encoding (e.g. text, image, sound, etc.). File formats are crucial to long-term preservation because a computer can open, process, and render file formats that it recognizes. *Many file formats are proprietary* (also known as native), *meaning that digital content can be opened and rendered only by the software application used to create, use, and store it*. However, as IT changed, some software vendors introduced new products that no longer support earlier versions of a file format. In such instances these formats become "legacy" format, and digital content embedded in them can be opened only with computer code written expressly for this purpose. Other vendors, such as Microsoft, support backward compatibility across multiple generations of technology so Microsoft Word 2010 can open and render documents in Microsoft Word 95. Nonetheless, it is unrealistic to expect any software vendor to support backward compatibility for its proprietary file formats for digital content that will be preserved for multiple decades.

> Many digital file formats are proprietary, meaning that content can be viewed and controlled only by the software application used to create, use, and store it.

In the late 1980s, an alternative to vendor-supported backward compatibility emerged to mitigate dependence on proprietary file formats through open system interoperable file formats. Essentially, this meant that digital content could be exported from one proprietary file format and imported to one or more other proprietary file formats. Over time, interoperable file formats evolved into open standard technology-neutral formats that today have these characteristics:

- *Open* means that the process is transparent and that participants in the process reach a consensus on the properties of the standard.
- *Standard* means that a recognized regional or international organization (e.g. the ISO) published the standard.
- *Technology neutral* means that the standard is interoperable on almost any technology platform that asserts conformance to the standard.

Because even open standard technology-neutral formats are not immune to technology obsolescence, their selection must take into account their technical sustainability and implementation in digital repositories. The PRONON program of the National Archives of the United Kingdom and long-term sustainability of file formats of the US Library of Congress assess the sustainability of open standard technology-neutral formats.

The recommended open standard technology-neutral formats for nine content

types listed in Table 17.1 are based on this ongoing work along with preferred file formats supported by Library and Archives Canada and other national archives. Unlike PDF/A, several of these file formats (e.g. XML, JPEG 2000, and Scalable Vector Graphics [SVG]) were not explicitly designed for digital preservation. *It cannot be emphasized too strongly that this list of recommended open standard technology–neutral formats (or any other comparable list) is not static and will change over time as technology changes*.

> The PDF/A file format was designed specifically for digital preservation.

Table 17.1 Recommended Open Standard Technology-Neutral Formats

	PDF/A	XML	TIFF	PNG	JPEG 2000	SVG	MPEG-2	BWF	WARC
Text	√	√							
Spreadsheets	√								
Images (raster)	√		√	√					
Photographs (digital)					√				
Vector graphics						√			
Moving images							√		
Audio								√	
Web									√
Databases		√							

ISO 19005 (PDF/A)—Document Management—Electronic Document File Format for Long-Term Preservation (2005, 2011, and 2012)

PDF/A is an open standard technology-neutral format that enables the accurate representation of the visual appearance of digital content without regard for the proprietary format or application in which it was created or used. PDF/A is widely used in digital repositories as a preservation format for static textual and image content. Note that PDF/A is agnostic with regard to digital imaging processes or storage media. PDFA/A supports conversion of TIFF and PNG images to PDF/A. There are two levels of conformance to PDF/A specifications. PDF/A-1a references the use of a “well-formed” hierarchical structure with XML tags that enable searching for a specific tag in a very large digital document. PDF/A-1b does not require this conformance, and as a practical matter, it does not affect the accurate

representation of visual appearance.

Since its publication in 2005, there have been two revisions of PDF/A. The first revision, PDF/A-2, was aligned with the Adobe Portable Document Format 1.7 published specifications, which Adobe released to the public domain in 2011. The second revision, PDF/A-3, supports embedding documents in other formats, such as the original source document, in a PDF document.

Extensible Markup Language (XML)—World Wide Web Consortium (W3C) Internet Engineering Group (1998)

XML is a markup language that is a derivative of Standard General Markup Language (SGML) that logically separates the rendering of a digital document from its content to enable interoperability across multiple technology platforms. Essentially XML defines rules for marking up the structure of content and its content in ASCII text. Any conforming interoperable XML parser can render the original structure and content. XML-encoded text is human-readable because any text editor can display the marked-up text and content. XML is ubiquitous in IT environments because many communities of users have developed document type definitions unique to their purposes, including genealogy, math, and relational databases. Structure data elements work with relational databases, so this enables relational database portability.

Tagged Image File Format: 1992

Tagged image file format (TIFF) was initially developed by the Aldus Corporation in 1982 for storing black-and-white images created by scanners and desktop publishing application. Over the next six years, several new features were added, including a wide range of color images and compression techniques, including lossless compression. The most recent version of TIFF 6.0 was released by Aldus in 1992. Subsequently, Adobe purchased Aldus and chose not to support any further significant revisions and updates. Nonetheless, TIFF is widely used in desktop scanners for creating digital images for preservation. With such a large base of users, it is likely to persist for some time, but Adobe's decision to discontinue further development of TIFF means that it will lack features of other current and future image file formats. Fortunately, there are tools available to convert TIFF images to PDF and PNG images.

ISO/IEC 15498: 2004—Information Technology-Computer Graphics and Image Processing-Portable Network Graphics (PNG): Functional Specifications

The W3C Internet Engineering Task Force supported the development of PNG as a replacement for graphics image format (GIF) because the GIF compression algorithm was protected by patent rights rather than being in the public domain, as many believed. In 2004, PNG became an international standard that supports lossless compression, grayscale, and true-color images with bit depths that range from 1 to 16 bits per pixel, file integrity checking, and streaming capability.

PNG replaced GIF as an international standard for grayscale and color images in 2004.

Scalable Vector Graphics (SVG): 2003—W3C Internet Engineering Task Force

Vector graphics images consist of two-dimensional lines, colors, curves, or other geometrical shapes and attributes that are stored as mathematical expressions, such as where a line begins, its shape, where it ends, and its color. Changes in these mathematical expressions will result in changes in the image. Unlike raster images, there is no loss of clarity of a vector graphics image when it is made larger. SVG images and their behavior properties are defined in XML text files, which means any named element in a SVG image can be indexed and searched. SVG images also can be accessed by any text editor, which minimizes on a specific software application to render and edit the images.

ISO/IEC 15444:2000—Joint Photographic Engineers Group (JPEG 2000)

JPEG 2000 *is an international standard for compressing full-color and grayscale digital images* and rendering them as full-size images and thumbnail images. Unlike JPEG, its predecessor, which supported only lossy compression, JPEG 2000 supports both lossy and lossless compression. Lossy compression means that during compression, bits that are considered technically redundant are permanently deleted. Lossless compression means no bits are lost or deleted. *The latter is very important for LTDP because lossy compression is irreversible.* JPEG 2000 is widely used in producing digital images in digital cameras and is an optional format in many digital scanners.

JPEG 2000 is an international standard for compressing and rendering full-color and grayscale digital images in full size or as thumbnails.

ISO/IEC 13818–3: 2000—Motion Picture Expert Group (MPEG-2)

MPEG-2 is an international broadcast standard for lossy compression of moving images and associated audio. The major competitor for MPEG-2 appears to be Motion JPEG 2000, which is used in small devices, such as cell phones.

European Broadcasting Tech 3285: 2011—Broadcast Wave Format (BWF)

First issued by the European Broadcasting Union in 1997 and revised in 2001 (v1) and 2011 (v2), BWF is a file format for audio data that is an extension of the

Microsoft Wave audio format. Its support of metadata ensures that it can be used for the seamless exchange of audio material between different broadcast environments and between equipment based on different computer platforms.

ISO 28500: 2009—WebARChive (WARC)

WebARChive (WARC) is an extension of the Internet Archive's ARC format to store digital content harvested through "Web crawls." WARC was developed to support the storage, management, and exchange of large volumes of "constituent data objects" in a single file. Currently, WARC is used to store and manage digital content collected through Web crawls and data collected by environmental sensing equipment, among others.

Digital Preservation Requirements

Implementing a sustainable LTDP program is not an effort that should be undertaken lightly. Digital preservation is complex and costly and requires collaboration with all of the stakeholders who are accountable for or have an interest in ensuring access to usable, understandable, and trustworthy electronic records for as far into the future as may be required.

As noted earlier, ISO 14721 and ISO 16363 establish the baseline functions and specifications for ensuring access to usable, understandable, and trustworthy electronic records, whether this involves regulatory and legal compliance for a business entity, vital records, accountability for a government unit, or cultural memory for a public or private institution. Most first-time readers who review the functions and specifications of ISO 14721 and ISO 16363 are likely to be overwhelmed by the detail and complexity of almost 150 specifications.

Long-Term Digital Preservation Capability Maturity Model®

A useful approach that both simplifies these specifications and provides explicit criteria regarding conformance to ISO 14721 and ISO 16363 is the Long-Term Digital Preservation Capability Maturity Model® (DPCMM).[12] The DPCMM, which is described in some detail in this section, draws on functions and preservation services identified in ISO 14721(OAIS) as well as attributes specified in ISO 16363, Audit and Certification of Trustworthy Repositories. It is important to note that the DPCMM is not a one-size-fits-all approach to ensuring long-term access to authentic electronic records. Rather, it is a flexible approach that can be adapted to an organization's specific requirements and resources.

> The Long-Term Digital Preservation Capability Maturity Model (DPCMM) systematically organizes high-level conformance to ISO 14721 and ISO 16363.

DPCMM can be used to identify the *current state* capabilities of digital

preservation that form the basis for debate and dialogue regarding the *desired future state* of digital preservation capabilities, and the level of risk that the organization is willing to assume. In many instances, this is likely to come down to the question of what constitutes digital preservation that is good enough to fulfill the organization's mission and meet the expectations of its stakeholders. The DPCMM has five incremental stages, which are depicted in Figure 17.3. In Stage 1, a systematic digital preservation program has not been undertaken or the digital preservation program exists only on paper, whereas Stage 5 represents the highest level of sustainable digital preservation capability and repository trustworthiness that an organization can achieve.

The DPCMM is based on the functional specifications of ISO 14721and ISO 16363 and accepted best practices in operational digital repositories. It is a systems-based tool for charting an evolutionary path from disorganized and undisciplined management of electronic records, or the lack of a systematic electronic records management program, into increasingly mature stages of digital preservation capability.

The goal of the DPCMM is to identify at a high level where an electronic records management program is in relation to optimal digital preservation capabilities, report gaps, capability levels, and preservation performance metrics to resource allocators and other stakeholders to establish priorities for achieving enhanced capabilities to preserve and ensure access to long-term electronic records.

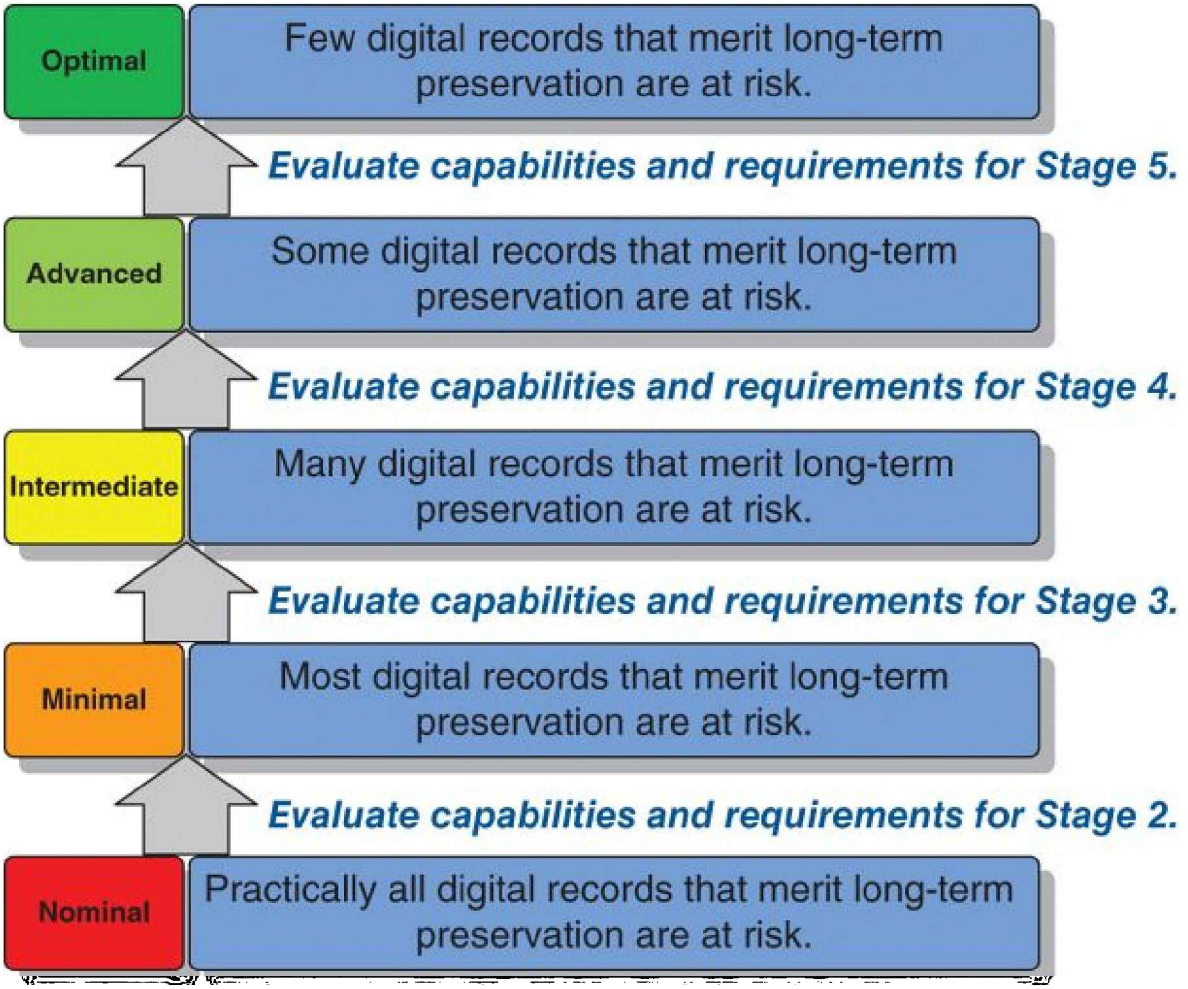


Figure 17.3 Five Levels of Digital Preservation Capabilities

Stage 5: Optimal Digital Preservation Capability

Stage 5 is the highest level of digital preservation readiness capability that an organization can achieve. It includes a strategic focus on digital preservation outcomes by continuously improving the manner in which electronic records life cycle management is executed. Stage 5 digital preservation capability also involves benchmarking the digital preservation infrastructure and processes relative to other best-in-class digital preservation programs and conducting proactive monitoring for breakthrough technologies that can enable the program to significantly change and improve its digital preservation performance. *In Stage 5, few if any electronic records that merit long-term preservation are at risk.*

Stage 4: Advanced Digital Preservation Capability

Stage 4 capability is characterized by an organization with a robust infrastructure and digital preservation processes that are based on ISO 14721 specifications and ISO 16363 audit and certification criteria. At this stage, the preservation of electronic records is framed entirely within a collaborative environment in which

there are multiple participating stakeholders. Lessons learned from this collaborative framework serve as the basis for adapting and improving capabilities to identify and proactively bring long-term electronic records under life cycle control and management. *Some electronic records that merit long-term preservation still may be at risk.*

Stage 3: Intermediate Digital Preservation Capability

Stage 3 describes an environment that embraces the ISO 14721 specifications and other best practice standards and schemas and thereby establishes the foundation for sustaining an enhanced digital preservation capability over time. This foundation includes successfully completing repeatable projects and outcomes that support the enterprise digital preservation capability and enables collaboration, including shared resources, between record-producing units and entities responsible for managing and maintaining trustworthy digital repositories. *In this environment, many electronic records that merit long-term preservation are likely to remain at risk.*

Stage 2: Minimal Digital Preservation Capability

Stage 2 describes an environment where an ISO 14721–based digital repository is not yet in place. Instead, a surrogate repository for electronic records is available to some records producers that satisfies some but not all of the ISO 14721 specifications. Typically, the digital preservation infrastructure and processes of the surrogate repository are not systematically integrated into business processes or universally available so the state of digital preservation is somewhat rudimentary and life cycle management of the organization's electronic records is incomplete. There is some understanding of digital preservation issues, but it is limited to a relatively few individuals. There may be virtually no relationship between the success or failure of one digital preservation initiative and the success or failure of another one. Success is largely the result of exceptional (perhaps even heroic) actions of an individual or a project team. Knowledge about such success is not widely shared or institutionalized. *Most electronic records that merit long-term preservation are at risk.*

Stage 1: Nominal Digital Preservation Capability

Stage 1 describes an environment in which the specifications of ISO 14721 and other standards may be known, accepted in principle, or under consideration, but they have not been formally adopted or implemented by the record-producing organization. Generally, there may be some understanding of digital preservation issues and concerns, but this understanding is likely to consist of ad hoc electronic records management and digital preservation infrastructure, processes, and initiatives. Although there may be some isolated instances of individuals attempting to preserve electronic records on a workstation or removable storage media (e.g. DVD or hard drive), *practically all electronic records that merit long-term preservation are at risk.*

Scope of the Capability Maturity Model

This capability maturity model consists of 15 components, or key process areas, that are necessary and required for the long-term preservation of usable, understandable, accessible, and trustworthy electronic records. Each component is identified and is accompanied by explicit performance metrics for each of the five levels of digital preservation capability.

The objective of the model is to provide a process and performance framework (or benchmark) against best practice standards and foundational principles of digital preservation, records management, information governance, and archival science. Figure 17.4 displays the components of the DPCMM.

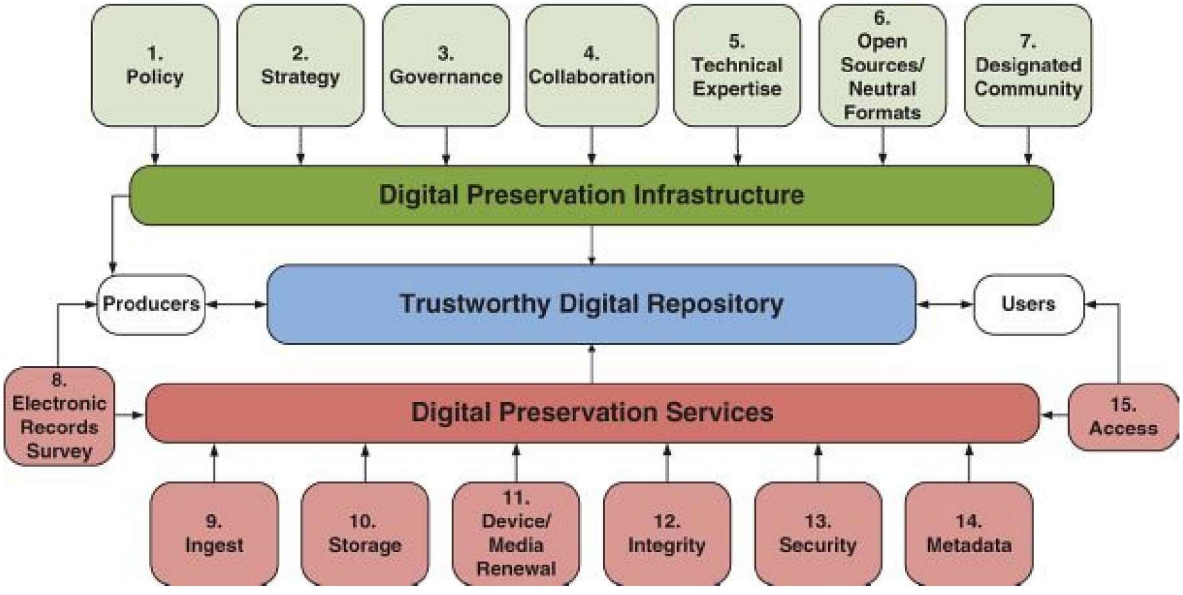


Figure 17.4 Digital Preservation Capability Maturity Model

Scope notes for each of the graphic elements in Figure 17.4 diagram are provided next for additional clarity. Numbered components in the model are associated with performance metrics and capability levels described in the next section.

Producers and Users

- *Records creators and owners* are stakeholders who have either the obligation or the option to transfer permanent and long-term (10-plus-years' retention) electronic records to one or more specified digital repositories for safekeeping and access.
- *Users*. Individuals or groups that have an interest in and/or right to access records held in the digital repository. These stakeholders represent a variety of interests and access requirements that may change over time.
- *Digital preservation infrastructure*. Seven key organizational process areas required to ensure sustained commitment and adequate resources for the long-term preservation of electronic records are:

1. *Digital preservation policy*. The organization charged with ensuring preservation and access to long-term and permanent legal, fiscal, operational, and historical records should issue its digital preservation policy in writing, including the purpose, scope, accountability, and approach to the operational management and sustainability of trustworthy repositories.
2. *Digital preservation strategy*. The organization charged with the preservation of long-term and permanent business, government, or historical electronic records must proactively address the risks associated with technology obsolescence, including plans related to periodic renewal of storage devices, storage media, and adoption of preferred preservation file formats.
3. *Governance*. The organization has a formal decision-making framework that assigns accountability and authority for the preservation of electronic records with long-term and permanent historical, fiscal, operational, or legal value, and articulates approaches and practices for trustworthy digital repositories sufficient to meet stakeholder needs. Governance is exercised in conjunction with information management and technology functions and with other custodians and digital preservation stakeholders, such as records-producing units and records consumers, and enables compliance with applicable laws, regulations, record retention schedules, and disposition authorities.
4. *Collaboration*. Digital preservation is a shared responsibility. The organization with a mandate to preserve long-term and permanent electronic business, government, or historical records in accordance with accepted digital preservation standards and best practices is well served by maintaining and promoting collaboration among its internal and external stakeholders. Interdependencies between and among the operations of records producing units, legal and statutory requirements, IT policies and governance, and historical accountability should be addressed systematically.
5. *Technical expertise*. A critical component in a sustainable digital preservation program is access to professional technical expertise that can proactively address business requirements and respond to impacts of evolving technologies. The technical infrastructure and key processes of an ISO 14721/ISO 16363–conforming archival repository requires professional expertise in archival storage, digital preservation solutions, and life cycle electronic records management processes and controls. This technical expertise may exist within the organization or be provided by a centralized function or service bureau or by external service providers, and should include an in-depth understanding of critical digital preservation actions and their associated recommended practices.
6. *Open standard technology-neutral formats*. A fundamental requisite for a sustainable digital preservation program that ensures long-term access

to usable and understandable electronic records is mitigation of obsolescence of file formats. Open standard platform-neutral file formats are developed in an open public setting, issued by a certified standards organization, and have few or no technology dependencies. Current preferred open standard technology file format examples include:

- XML and PDF/A for text
- PDF/A for spreadsheets
- JPEG 2000 for photographs
- PDF/A, PNG, and TIFF for scanned images
- SVG for vector graphics
- BWF for audio
- MPEG-4 for video
- WARC for Web pages

Over time, new digital preservation tools and solutions will emerge that will require new open standard technology-neutral standard file formats. Open standard technology-neutral formats are backwardly compatible so they can support interoperability across technology platforms over an extended period of time.

7. *Designated community*. The organization that has responsibility for preservation and access to long-term and permanent legal, operational, fiscal, or historical government records is well served through proactive outreach and engagement with its designated community. There are written procedures and formal agreements with records-producing units that document the content, rights, and conditions under which the digital repository will ingest, preserve, and provide access to electronic records. Written procedures are in place regarding the ingest of electronic records and access to its digital collections. Records producers will submit fully conforming ISO 14721/ISO 16363 SIPs while DIPs are developed and updated in conjunction with its user communities.

- *Trustworthy digital repository*. This includes the integrated people, processes, and technologies committed to ensuring the continuous and reliable design, operation, and management of digital repositories entrusted with long-term and permanent electronic records. A trustworthy digital repository may range from a simple system that involves a low-cost file server and software that provide nonintegrated preservation services, to complex systems comprising data centers and server farms, computer hardware and software, and communication networks that interoperate.

  The most complete trustworthy digital repository is based on models and standards that include ISO 14721, ISO 16363, and generally accepted best digital preservation practices. The repository may be managed by the organization that owns the electronic records or may be provided as a service

by an external third party. It is likely that many organizations initially will rely on surrogate digital preservation capabilities and services that approximate some but not all of the capabilities and services of a conforming ISO14721/ISO 16363 trustworthy digital repository.

- *Digital preservation processes and services*. Eight key business process areas needed for continuous monitoring of the external and internal environments in order to plan and take actions to sustain the integrity, security, usability and accessibility of electronic records stored in trustworthy digital repositories.

    1. *Electronic records survey*. A trustworthy repository cannot fully execute its mission or engage in realistic digital preservation planning without a projected volume and scope of electronic records that will come into its custody. It is likely that some information already exists in approved retention schedules, but it may require further elaboration as well as periodic updates, especially with regard to preservation ready, near preservation ready, and legacy electronic records held by records-producing units.
    2. *Ingest*. A digital repository that conforms to ISO 14721/ISO 16363 has the capability to systematically ingest (receive and accept) electronic records from records-producing units in the form of SIPs, move them to a staging area where virus checks and content and format validations are performed, transform electronic records into designated preservation formats as appropriate, extract metadata from SIPs and write it to PDIPDI, create AIPs, and transfer the AIPs to the repository's storage function. This process is considered the minimal work flow for transferring records into a digital repository for long-term preservation and access.
    3. *Archival storage*. ISO 14721 delineates systematic automated storage services that support receipt and validation of successful transfer of AIPs from ingest, creation of PDI for each AIP that confirms its "fixity"[13] during any preservation actions through the generation of hash digests, capture and maintenance of error logs, updates to PDI including transformation of electronic records to new formats, production of DIPs from access, and collection of operational statistics.
    4. *Device and media renewal*. No known digital device or storage medium is invulnerable to decay and obsolescence. A foundational digital preservation capability is ensuring the readability of the bitstreams underlying the electronic records. ISO 14721/ ISO 16363 specify that a trustworthy digital repository's storage devices and storage media should be monitored and renewed ("refreshed") periodically to ensure that the bitstreams remain readable over time. A projected life expectancy of removable storage media does not necessarily apply in a specific instance of storage media. Hence, it is important that a trustworthy digital

repository have a protocol for continuously monitoring removable storage media (e.g. magnetic tape, external tape drive, or other media) to identify any that face imminent catastrophic loss. Ideally, this renewal protocol would execute renewal automatically after review by the repository.

5. *Integrity*. A key capability in conforming ISO 14721/ISO 16363 digital repositories is ensuring the integrity of the records in its custody, which involves two related preservation actions. The first action generates a hash digest algorithm (also known as a cyclical redundancy code) to address a vulnerability to accidental or intentional alterations to electronic records that can occur during device/media renewal and internal data transfers. The second action involves integrity documentation that supports an unbroken electronic chain of custody captured in the PDI in AIPs.
6. *Security*. Contemporary enterprise information systems typically execute a number of shared or common services that may include communication, name services, temporary storage allocation, exception handling, role-based access rights, security, backup and business continuity, and directory services, among others. A conforming ISO 14721/ISO 16363 digital repository is likely to be part of an information system that may routinely provide some or perhaps all of the core security, backup, and business continuity services, including firewalls, role-based access rights, data-transfer-integrity validations, and logs for all preservation activities, including failures and anomalies, to demonstrate an unbroken chain of custody.
7. *Preservation metadata*. A digital repository collects and maintains metadata that describes actions associated with custody of long-term and permanent records, including an audit trail that documents preservation actions carried out, why and when they were performed, how they were carried out, and with what results. *A current best practice is the use of a PREMIS-based data dictionary to support an electronic chain of custody that documents authenticity over time as preservation actions are executed.* Capture of all related metadata, transfer of the metadata to any new formats/systems, and secure storage of metadata are critical. All metadata is stored in the PDI component of conforming AIPs.
8. *Access*. Organizations with a mandate to support access to permanent business, government, or historical records are subject to authorized restrictions. A conforming ISO 14721/ISO 16363 digital repository will provide consumers with trustworthy records in "disclosure-free" DIPs redacted to protect, privacy, confidentiality, and other rights, where appropriate, and searchable metadata that users can query to identify and retrieve records of interest to them. Production of DIPs is tracked, especially when they involve extractions, to verify their trustworthiness and to identify query trends that are used to update electronic accessibility tools to support these trends.

Digital Preservation Capability Performance Metrics

Digital preservation performance metrics for each level of the five levels of the model have been mapped to each of the 15 numbered components described in the previous section. The performance metrics are explicit empirical indicators that reflect an incremental level of digital preservation capability. The digital preservation capability performance metrics for digital preservation strategy listed in Table 17.2 illustrate the results of this mapping exercise.[14]

Conducting a gap analysis of its digital preservation capabilities using these performance metrics enables the organization to identify both its current state and desired future state of digital preservation capabilities. In all likelihood, this desired future state will depend on available resources, the organization's mission, and stakeholder expectations. "Good-enough" digital preservation capabilities will vary by organization; what is good enough for one organization is unlikely to coincide with what is good enough for another.

Table 17.2 Digital Preservation Performance Metrics

Level	Capability Description
0	A formal strategy to address technology obsolescence does not exist.
1	A strategy to mitigate technology obsolescence consists of accepting electronic records in their native format with the expectation that new software will become available to support these formats. During this interim period, viewer technologies will be relied on to render usable and understandable electronic records.
2	Electronic records in interoperable "preservation-ready"* file formats and transformation of one native file format to an open standard technology-neutral file format are supported. Changes in information technologies that may impact electronic records collections and the digital repository are monitored proactively and systematically.
3	The organization supports transformation of selected native file formats to preferred/supported preservation file formats in the trustworthy digital repository. Records-producing units are advised to use preservation-ready file formats for permanent or indefinite long-term (e.g. case files, infrastructure files) electronic records in their custody.
4	Electronic records in all native formats are transformed to available open standard technology-neutral file formats.

* The term "preservation-ready file formats" refers to open standard technology-neutral formats that the organization has identified as preferred for long-term digital preservation.

Digital Preservation Strategies and Techniques

Any organization with long-term or permanent electronic records in its custody

must ensure that the electronic records can be read and correctly interpreted by a computer application, rendered in an understandable form to humans, and trusted as accurate representations of their logical and physical structure, substantive content, and context. To achieve these goals, a digital repository should operate under the mandate of a digital preservation strategy that addresses 10 digital preservation processes and activities:

1. *Adopt preferred open standard technology-neutral formats*. Earlier, nine open standard technology-neutral file formats that covered text, images, photographs, vector graphics, moving images, audio, and Web pages were discussed. Adoption of these file formats means that the digital repository will support their use in its internal digital preservation activities and notify the producers of records of the preferred formats for preservation-ready electronic records to be transferred to the repository's custody.
2. *Acquire electronic records in preservation-ready formats*. Likely many born-digital electronic records along with scanned images will be created or captured in a preservation-ready format. Acquisition or ingest of electronic records already in preservation-ready formats can significantly reduce the workload of the repository because it will not be necessary to transform records to open standard technology-neutral formats.
3. *Acquire and transform electronic records in near-preservation-ready formats*. Near-preservation-ready format are native proprietary file formats that can be easily transformed to preservation-ready file formats through widely available software plug-ins. Ideally, over time, the volume of near-preservation-ready records will diminish as records producers increasingly convert records scheduled for long-term retention into preservation-ready formats before they are transferred to the repository.
4. *Acquire legacy electronic records*. Legacy electronic records initially were created in a proprietary file format that is obsolete and no longer supported by a vendor. In most instances, electronic records embedded in legacy file formats can be recovered and saved in a preservation-ready format only if special computer code is written to extract the records from their legacy format. Once extracted from the legacy format, they can be written to a contemporary format. Niche vendors provide this kind of service, but it is relatively expensive and perhaps beyond the resources of many repositories.

   An alternative is to forgo this costly process in the hope that a future technology, such as **emulation**, will be widely available and relatively inexpensive. Meanwhile, the repository would rely on a file viewer technology, such as Inside Out, to render legacy electronic records into format understandable to humans with the exact logical and physical structure and representation at the time they were created and used.
5. *Maintain bitstream readability through device/media removal*. No known digital storage device or media is exempt from degradation and technology obsolescence. Consequently, the bitstreams of 1s and 0s that underlie

electronic records are stored on media that are vulnerable to degradation and technology obsolescence. Technology obsolescence may occur when a vendor introduces a new form factor for storage device/media, such as the transition from 5.25-inch disk drives and disks to 3.5-inch disk drives and media to thumb drives. With today's technology, periodic device/medial renewal is the only known way to keep bitstreams available. *A rule of thumb is to renew storage device/media at least every 10 years*. Failure to maintain the readability of bitstreams over time is an absolute guarantee the electronic records cannot be recovered and that the records will be permanently lost for all practical purposes.

6. *Migrate to new open standard technology-neutral formats*. These formats are not immune to technology obsolescence. The inevitable changes in IT mean that new open standard technology formats will be created that displace current ones. The solution to this issue is migration from an older or current open standard technology-neutral format to newer ones. Seamless migration from old to new open standard technology–neutral formats is made possible through backward compatibility. "Backward compatibility" means that a new standard can interpret digital content in an old standard and then save it in the new format standard. **Migration** is the most widely used tool to mitigate file format obsolescence.
7. *Protect the integrity and security of electronic records*. Imperfect information technologies inevitably have glitches that, along with accidental human error and intentional human actions, can corrupt or otherwise compromise the trustworthiness of electronic records though some alteration in the underlying bitstream. Accidental alteration occurs when preservation actions are initiated for electronic records. These actions may occur during transformation, migration, media renewal, accessions to digital records, and relocation of electronic records from one part of the repository to another. The most effective tool for validating that no unauthorized changes to electronic records occur is to compute a hash digest before a preservation action occurs and after the action is completed. If there is change of only one bit, a comparison of the two will identify it. Capturing these pre- and posthash digests and saving them as preservation description information can contribute to an electronic chain of custody.

   A robust firewall that blocks unauthorized access with tightly controlled role-based permission rights will help protect the security of records in the custody of the repository.

   A further enhancement to protect against a cataclysmic natural or man-made disaster is maintaining a backup copy of the repository's holdings at an off-site facility.
8. *Capture and save preservation metadata*. Preservation metadata, which consists of tracking, capturing, and maintaining documentation of all preservation actions associated with electronic records, involves identifying these events, the agents that executed the actions, and the results of the

actions, including any corrective action taken. Saving this metadata along with the hash digest integrity validations just discussed enables robust electronic chain of custody and establishes a strong basis for the trustworthiness of electronic records in the custody of the digital repository.

9. *Provide access. Access to usable and trustworthy records is the ultimate justification for digital preservation*. In some respects, this may be the most challenging aspect of digital preservation because user expectations for customized retrieval tools, access speed, and delivery formats of electronic records may exceed the current resources of a trusted digital repository. Nonetheless, some form of user access, through replication of records in a single open standard technology format such as PDF/A for text and scanned images and JPEG 2000 for digital photographs, would be a major accomplishment and form the basis for a more aggressive access program over time.
10. *Engage proactively with records producers and other stakeholders*. The traditional notion of an archive being in a reactive mode with regard to records producers and other stakeholders in LTDP simply will not work in today's world. Proactive engagement with records producers about how capturing electronic records in open standard technology-neutral can support both current business operation requirements and long-term requirements for usable, understandable, and trustworthy can be a win-win for the digital repository and the records producers. Equally important is the notion of proactive engagement with all of the stakeholders in ensuring long-term access to usable, understandable, and trustworthy electronic records. Support of other stakeholders can be leveraged to gain broad organizational support for the digital repository.

Evolving Marketplace

The design and implementation of a digital repository that operates under this digital preservation strategy can be carried out in several different ways. One way is to use internal expertise to build a stand-alone repository that conforms to these digital preservation strategy requirements. Typically, an internally built repository is costly, takes considerable time to implement, and may not meet all expectations because of technical inexperience. An alternative is to use the services and/or solutions offered by an external institution or supplier. A third-party solution is offered by Archivematica, a Vancouver, British Columbia, company that specializes in the use of open-source software and conformance to the specifications of ISO SO 14721. “Archivematica is a free and open-source digital preservation system that is designed to maintain standards-based, long-term access to collections of digital objects.”[15] Another company, Preservica,[16] has an ISO 14721–conforming digital preservation SaaS and on-premise solution that has been implemented in national and pan-national archives as well as 19 US state archives. *It is likely that other repository solutions and preservation services will emerge over the next few years as demand for digital archiving increases*. The digital preservation strategy

discussed earlier can be used to assess the capabilities of these solutions. Spain-based Libnova also offers a cloud-based digital preservation solution, especially for handling large collections at national libraries or archives.

In November 2017 The National Cultural AudioVisual Archives (NCAA), hosted by the Indira Gandhi National Centre for the Arts Audio/Visual Repository, became the first digital repository in the world to be awarded ISO 16363 certification. The audit was conducted by PTAB, the Primary Trustworthy Digital Repository Authorisation Body. Its members are the group of global digital preservation who developed the ISO standards 14721, 16363, and 16919. PTAB was the first organization to be accredited to perform repository audits. A number of public sector repositories have announced plans to undergo audits and seek certification. It remains to be seen whether commercial enterprises and/or those who fund repositories for valuable digital encoded information will seek certification to ensure repositories are worthy of trust and sustained funding.

Looking Forward

It wasn't long ago—5 to 10 years—that LTDP required a relatively expensive and complicated set of internal processes to store digital records needed for 10 years or more. Migrating digital images from older, proprietary file formats and maintaining records in industry-standard, technology-neutral file formats while ensuring readability presented major challenges.

But today, there are new outsourced options that make digital preservation much easier and more cost effective for organizations needing to preserve digital documents. The approach that digital preservation suppliers take is to manage the entire digital conversion process (from paper or microfilm to digital) and to store five to six copies of each image with a major cloud supplier like Microsoft Azure or Amazon AWS on servers dispersed geographically around the world. Some approaches use more than one cloud supplier to reduce the risk of loss even further.

Error-detecting software uses checksum algorithms to scan digital records periodically for any degradation or loss of bits due to hardware failures, hacking attacks, or other anomalies. Then the damaged copy is either restored or replaced, ensuring that five to six viable copies are still available in various parts of the world.

This newer cloud-based approach has made digital preservation more accessible and viable for major organizations with the need to preserve digital information far into the future, especially movie studios, national libraries, universities, research organizations, and government entities.

New outsourced cloud-based options make digital preservation much easier and more cost effective for organizations needing to preserve digital documents.

Conclusion

Organizations, especially those whose primary mission is to preserve and provide access to permanent records, face significant challenges in meeting their LTDP needs. They must collaborate with internal and external stakeholders, develop governance policies and strategies to govern and control information assets over long periods of time, inventory records in the custody of records producers, monitor technology changes and evolving standards, and sustain trustworthy digital repositories. The most important consideration is to determine what level of LTDP maturity is appropriate, achievable, and affordable for the organization and to begin working methodically toward that goal for the good of the organization and its stakeholders over the long term. In addition, organizations should focus on what is doable over the next 10 to 20 years rather than the next 50 or 100 years.

CHAPTER SUMMARY: KEY POINTS

- Digital preservation is defined as long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required to be retained.
- Digital preservation applies to content that is born digital as well as content that is converted to digital form.
- Capability for properly ensuring access to authentic electronic records over time (regardless of the challenges of technological obsolescence), is a sophisticated combination of policies, strategies, processes, specialized resources, and adoption of standards.
- Most records are useful for only a short period of time, but some may need to be retained for long periods or permanently. For those records, organizations will need to plan for their preservation to ensure that they remain accessible, trustworthy, and useful.
- Electronic records are being created at rates that pose significant threats to our ability to organize, control, and make them accessible for as long as they are needed.
- Threats to LTDP of records can be internal or external, from natural disasters, computer or storage failures, and even from the financial viability of an organization, which can limit needed funding.
- Building and sustaining the capability to manage digital information over

long periods of time is a shared responsibility among all stakeholders.

- ISO 14721 is the lingua franca of digital preservation. The international digital preservation community has embraced it as the framework for viable and technologically sustainable digital preservation repositories.
- An ISO 14721 (OAIS)–compliant repository is the best way to preserve an organization's long-term digital assets.
- ISO 18492 provides practical methodological guidance for the long-term preservation of e-documents, when the retention period exceeds the expected life of the technology that created it.
- ISO 16363 is an audit and certification standard organized into three broad categories: organization infrastructure, digital object management, and technical infrastructure and security risk management.
- ISO 16363 represents the gold standard of audit and certification for trustworthy digital repositories.
- The PREMIS standard defines a core set of preservation metadata elements with a supporting data dictionary applicable to a broad range of digital preservation activities.
- Many digital file formats are proprietary, meaning that content can viewed and controlled only by the software application used to create, use, and store it.
- The digital preservation community recognizes that open standard technology-neutral standards play a key role in ensuring that digital records are usable, understandable, and reliable for as far into the future as may be required.
- The PDF/A file format was specifically designed for digital preservation.
- PNG replaced GIF as an international standard for grayscale and color images in 2004.
- JPEG 2000 is an international standard for compressing and rendering full-color and grayscale digital images in full size or as thumbnails.
- The Long-Term Digital Preservation Capability Maturity Model simplifies conformance to ISO 14721 and ISO 16363.
- Migration, refreshment, and replication are examples of specific preservation techniques.
- New outsourced cloud-based options make digital preservation much easier and more cost effective for organizations needing to preserve digital documents.

Notes

1. Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)* (Washington, DC: CCSDS Secretariat, 2002), p. 1.

2. Kate Cumming, "Metadata Matters," in Julie McLeod and Catherine Hare, eds., *Managing Electronic Records* (London: Facet, 2005), 48.

3. David Rosenthal et al., "Requirements for Digital Preservation Systems," *D-Lib Magazine* 11, no. 11 (November 2005), www.dlib.org/dlib/november05/rosenthal/11rosenthal.html.

4. "ISO 14721:2003, 2012Space Data and Information Transfer Systems—Open Archival Information System—Reference Model," www.iso.org/iso/catalogue_detail.htm?csnumber=24683 (accessed May 21, 2012).

5. ISO 14721:2003(E), section 4.1.

6. ISO 14721:2003(E), section 5.4.

7. See ISO 16363:2012 (E), sections 3.1–3.5.2.

8. See ibid., sections 4.1–4/6/2/1.

9. See ibid., sections 5.1–5.2.3.

10. For a useful overview of PREMIS, see Priscilla Caplan, "Understanding PREMIS," Library of Congress, February 1, 2009, www.loc.gov/standards/premis/understanding-premis.pdf.

11. Library of Congress, "PREMIS Data Dictionary Version 2.2: Hierarchical Listing of Semantic Units," September 13, 2012, www.loc.gov/standards/premis/v2/premis-dd-Hierarchical-Listing-2-2.html.

12. Charles Dollar and Lori Ashley are codevelopers of this model. Since 2007 they have used it successfully in both the public and private sectors. The most recent instance is a digital preservation capability assessment for the U.S. Council of State Archivists (CoSA). For more information about the model, see "Digital Preservation Capability Maturity Model" at www.savingthedigitalworld.com (accessed December 12, 2013).

13. ISO 1472 uses "fixity" to express the notion that there have been no unauthorized changes to electronic records and associated Preservation Description Information in the custody of the repository. See ISO 14721: 2003 (E): 1.6.

14. For information about digital preservation capability performance metrics, visit "Digital Preservation Capability Maturity Model," https://www.statearchivists.org/resource-center/resource-library/digital-preservation-capability-maturity-model-dpcmm/

15. Archivematica, "What Is Archivematica?" October 15, 2012, www.archivematica.org/wiki/Main_Page.

16. Preservica, www.Preservica.com (accessed June 10, 2019).

* Portions of this chapter are adapted from Chapter 17, Robert F. Smallwood, *Managing Electronic Records: Methods, Best Practices, and Technologies*, © John Wiley & Sons, Inc., 2013. Reproduced with permission of John Wiley & Sons, Inc.

CHAPTER 18
Maintaining an Information Governance Program and Culture of Compliance

Maintaining your information governance (IG) program beyond an initial project effort is key to realizing the continued and long-term benefits of IG. This means that the IG program must become an everyday part of an organization's operations. This requires vigilant and consistent monitoring and auditing to ensure that IG policies and processes are effective and consistently followed and enforced. Using audits and the proper controls should become a regular part of the enterprise's operations.

Monitoring and Accountability

This requires a continuous tightening down and expansion of protections and the implementation of newer strategic technologies. Information technology (IT) developments and innovations that can foster the effort must be steadily monitored and evaluated, and those technology subsets that can assist in providing security need to be incorporated into the mix.

The IG policies themselves must be reviewed and updated periodically to accommodate changes in the business environment, laws, regulations, and technology. Program gaps and failures must be addressed, and the effort should continue to improve and adapt to new types of security threats.

That means accountability—some individual must remain responsible for an IG policy's administration and results[1]—perhaps the executive sponsor for the initial project becomes the chief information governance officer (CIGO) or IG “Czar” of sorts; or the chief executive officer (CEO) continues ownership of the program and drives its active improvement. The organization may also decide to form a standing IG board, steering committee, or team with specific responsibilities for monitoring, maintaining, and advancing the program.

However it takes shape, an IG program must be ongoing, dynamic, and aggressive in its execution in order to remain effective.

> Maintaining an IG program requires that someone is accountable for continual monitoring and refinement of policies and tools.

Change Management—Required

By Monica Crocker

IG programs inherently involved **change management** (CM), to coach and train employees on the value of information and the risks it carries. They also must become fully aware of how the IG program objectives will help the organization meet its overall business objectives.

Here are some change management (CM) guidelines specific to IG. There are several steps to an IG change Management Plan:

1. Defining Objectives
2. Defining Scope
3. Developing the Plan
4. Executing the Plan

During the initial phase, defining objectives, there is one key question you will need to answer: What is the change that you need to manage? This must be more specific than "Practice good IG." It should be concrete and measurable. An example might be: "Task every team member to inventory the records they own by the end of this fiscal year. And then take the appropriate action to secure or destroy those records as appropriate." The CM objective can then be further refined for each of the various stakeholders in the organization. The CM initiative may be asking some stakeholders to lead the change, a few to champion it, others to simply comply with or accept it, and the rest to refrain from working against the change.

In conjunction with defining the objectives of the change management effort, it is necessary to define the scope of that effort. The IG team must determine the desired results of the CM effort. Examples of questions to consider include the following:

- Do stakeholders need to have a basic understanding of IG principles?
- Do stakeholders need to have a basic understanding of the organizational strategy for IG?
- Do stakeholders need to have an awareness of the specific IG activities that will take place?
- Does each stakeholder need to understand what they need to do to support the IG strategy?
- Do stakeholders need an awareness of future communications on the topic (what additional communications they will receive and when)?

Key to this is, of course, identification of the IG Change Management stakeholders. This identification should include a description of their role in the IG Change Management effort. Each of these roles may require a separate communication plan when those are developed.

Steps to an IG Change Management Plan include: Defining Objectives, Defining Scope, Developing the Plan, and Executing the Plan.

Change management is driven by individuals, but is manifested at various levels of the organization. For instance, when the CEO publishes a letter in the annual report stating "Here is how we do business now," that is an enterprise level organizational change. The CEO may have signed the letter, but there were likely many individual stakeholders behind it. At an organizational level, CM requires planning. Project management methodology reinforces the idea that *the CM effort should be built into the initial planning for the IG program and included as part of any resulting project plans*. This planning should identify timelines for specific CM actions and the resources necessary to accomplish those actions.

Change management efforts should be integrated into the IG Communications Plan. Those communications must take the abstract concept of IG and make it tangible for the various stakeholders. For instance, instead of communicating "we need to govern our information better," a communication could say: "We retain and store a lot of e-mail, those e-mails put us at unnecessary risk, and therefore, we are going to take actions that reduce unnecessary e-mail storage." Within the CM communication plan, additional roles may be identified, such as change advocates (cheerleaders) and change management experts. CM experts (such as a representative from Human Resources) could attend the kickoff meeting for each department affected by the Information Governance initiative to talk frankly about change management challenges the department could expect and techniques for dealing with those.

Change Management efforts should be integrated into the IG Communications Plan.

At its core, CM happens at an individual level. The CM strategy directed at individuals should incorporate the following assumptions:

- Most people feel a strong sense of ownership of their work, and, often, the information associated with that work.
- People have been developing their own strategies for managing information since they first started collecting it.
- Therefore, asking individuals to change the way they manage the information associated with their work is a significant change.

As a result, communications to affected individuals will comprise the critical component of an IG Change Management plan. At a minimum, individuals will want to know why the change is happening, the logistics of the change (what and when) and how it impacts them. In addition to those factual components of a CM

communication effort, it is important that the effort allows affected individuals to feel like their input is valued and their concerns have been acknowledged. For instance, a possible response to an objection to placing e-mails into a shared repository is "Those e-mails are too valuable to be unavailable to your successor should you win the lottery and not come into work tomorrow." This acknowledges that the e-mails and the work they support are valuable to the organization.

A change management plan should incorporate the following factors:

- CM communications are typically another manifestation of the 80/20 rule: a given communication may satisfy 80% of the recipients, but the remaining 20% may need follow up or a communication that is tailored to their situation. Therefore, communications might be customized in advance. Alternatively, the communication plan can incorporate the resources to respond to questions after a communication is delivered.
- Multiple options for message delivery exist in every organization. To improve the effectiveness of the communication, the same message can be delivered through multiple channels: a broadcast e-mail, posters in the breakroom, and announcements during company meetings, and so on.
- Repetition will improve message retention, but it is necessary to vary those messages in order to keep people's attention. This could be accomplished by altering the tone of the message or the specific topic of the message. Consider the example of an airline safety briefing given by attendants before takeoff. The basic content of these messages has not changed much in years, but a given airline will refresh the message on a regular basis to keep passengers' attention.
- As with organizational communications, communications aimed at individuals should strive to take the abstract concept of Information Governance and make it more concrete.
- Finally, remember that most of the recipients of the change management communications have full-time jobs and those jobs are *not* information governance. Even asking them to take time away from those jobs to be aware of information governance initiatives is, on some level, imposing an additional burden on them.

  The CM strategy should include mechanisms that reinforce positive responses for each type of stakeholder. That could range from a positive "your record has been successfully uploaded and is now secure" message in a system to a gift card giveaway. This is one area of the IG strategy where creativity is especially important. Implement whatever works for the given organization and stakeholders.

A word about "naysayers." It is human nature to avoid negative feedback. However, critics may have excellent suggestions disguised as concerns. By incorporating extra time to collect input from those individuals into the change management plan, the IG program might gain valuable insights and be improved.

A good measure of a successful IG Change Management initiative is when the biggest critic switches from describing inevitable failure to asking when the next phase of the program will be rolled out.

> The CM strategy should include mechanisms that reinforce positive responses for each type of stakeholder.

A few other points to keep in mind:

- IG is a complex subject. A challenge for any IG communication is making the subject comprehensible to a casual participant.
- Assume positive intent; most people want to do the right thing, even if their words or actions may not reflect that.
- Change management is a specific area of expertise. The organization may need to consider bringing in outside resources.
- Stay flexible. Measure the success of the CM strategy as it proceeds and restrategize when necessary. Also, the IG program may be modified midstream, requiring the CM strategy to adjust accordingly.
- Do not be afraid to admit mistakes. Be prepared to change the CM strategy based on feedback and results after it is initiated. In summary, some might consider CM to be a science, but applying it to a particular organization with a unique IG strategy is an art. The only wrong way to do it is not to do it at all.

Continuous Process Improvement

This requires implementing principles of continuous process improvement (CPI). CPI is a "never-ending effort to discover and eliminate the main causes of problems." It accomplishes this by using small-steps improvements, rather than implementing one huge improvement. In Japan, the word *kaizen* reflects this gradual and constant process, as it is enacted throughout the organization, regardless of department, position, or level.[2] To remain effective, the program must continue using CPI methods and techniques.

Maintaining and improving the program will require monitoring tools, periodic audits, and regular meetings for discussion and approval of changes to improve the program. It will require a cross-section of representatives from IT, legal, records management, compliance, risk management, and functional business units participating actively and citing possible threats and sources of information leakage.

Why Continuous Improvement Is Needed

While the specific drivers of change are always evolving, the reasons that organizations need to continuously improve their program for securing

information assets are relatively constant, and include:

- *Changing technology*. New technology capabilities need to be monitored and considered with an eye to improving, streamlining, or reducing the cost of IG. The IG program needs to anticipate new types of threats and also evaluate adding or replacing technologies to continue to improve it.
- *Changing laws and regulations*. Compliance with new or updated laws and regulations must be maintained.
- *Internal information governance requirements*. As an organization updates and improves its overall IG, the program elements that concern critical information assets must be kept aligned and synchronized.
- *Changing business plans*. As the enterprise develops new business strategies and enters new markets, it must reconsider and update its IG program. If, for instance, a firm moves from being a domestic entity to a regional or global one, new laws and regulations will apply, and perhaps new threats will exist, and new security strategies must be formed.
- *Evolving industry best practices*. Best practices change, and new best practices arise with the introduction of each successive wave of technology, and with changes in the business environment. The program should consider and leverage new best practices.
- *Fixing program shortcomings*. Addressing flaws in the IG program that are discovered through testing, monitoring, and auditing; or addressing an actual breach of confidential information; or a legal sanction imposed due to noncompliance are all reasons why a program must be revisited periodically and kept updated.[3]

Maintaining the IG program requires that a senior level officer of the enterprise continues to push for enforcement, improvement, and expansion. This requires leadership, and a consistent and loud message to employees. IG and the security of information assets must be on the minds of all members of the enterprise; it must be something they are aware of and think about daily. They must be on the lookout for ways to improve it, and they should be rewarded for those contributions.

Gaining this level of mindshare in employees' heads will require follow-up messages in the form of personal speeches and presentations, newsletters, corporate announcements, e-mail messages, and even posters placed at strategic points (e.g., near the shared printing station). Everyone must be reminded that keeping e-documents and information assets secure is everyone's job, and that to lose or leak confidential information harms the organization over the long term and erodes its value.

CHAPTER SUMMARY: KEY POINTS

- Keeping an enterprise's IG program effective requires vigilant and consistent monitoring and auditing to ensure that IG are followed and enforced.
- Steps to an IG Change Management Plan include: Defining Objectives; Defining Scope; Developing the Plan; and Executing the Plan.
- Change management efforts should be integrated into the IG Communications Plan.
- The CM strategy should include mechanisms that reinforce positive responses for each type of stakeholder.
- Information technologies that can assist in advancing the program must be steadily monitored, evaluated, and implemented.
- To maintain and improve the IG program will require monitoring tools, regular audits, and regular meetings for discussion and approval of changes to the program to continually improve it.
- Organizations need to continuously improve their program for securing information assets due to: (1) changing technology; (2) changing laws and regulations; (3) internal information governance requirements; (4) changing business plans; (5) evolving industry best practices; and (6) fixing program shortcomings.
- Maintaining the program to secure information assets requires that a senior-level officer of the enterprise continues to push for enforcement, improvement, and expansion of the program to secure confidential information assets.

Notes

1. Mark Woeppel, "Is Your Continuous Improvement Organization a Profit Center?" June 15, 2009, www.processexcellencenetwork.com/process-management/articles/is-your-continuous-improvement-organization-a-prof/.
2. Donald Clark, Big Dog and Little Dog's Performance Juxtaposition, "Continuous Process Improvement," March 11, 2010, www.nwlink.com/~donclark/perform/process.html.
3. Randolph Kahn and Barclay T. Blair, *Information Nation: Seven Keys to Information Management Compliance* (Silver Spring, MD: AIIM International, 2004), 242–243.

APPENDIX A
Information Organization and Classification: Taxonomies and Metadata

Barb Blackburn, CRM, with Robert Smallwood; edited by Seth Earley

The creation of electronic documents and records is exploding exponentially, multiplying at an increasing rate, and sifting through all this information results in a lot of wasted, unproductive (and expensive) knowledge-worker time. This has real costs to the enterprise. According to the study, "The High Cost of Not Finding Information," an *IDC* report, "knowledge workers spend at least 15 to 25 percent of the workday searching for information. Only half the searches are successful." *Experts point to poor* ***taxonomy*** *design as being at the root of these failed searches and lost productivity*.

Taxonomies are at the heart of the solution to harnessing and governing information. *Taxonomies are hierarchical classification structures* used to standardize the naming and organization of information, and their role and use in managing electronic records cannot be overestimated.

Although the topic of taxonomies can get complex, in **electronic records management** (ERM), they are a sort of online card catalog that is cross-referenced with hyperlinks and is used to organize and manage records and documents.[2]

According to Forrester Research, taxonomies "represent agreed-upon terms and relationships between ideas or things and serve as a glossary or knowledge map helping to define how the business thinks about itself and represents itself, its products and services to the outside world."[3]

Gartner Group researchers warn that "to get value from the vast quantities of information and knowledge, enterprises must establish discipline and a system of governance over the creation, capture, organization, access, and utilization of information."[4]

Over time, organizations have implemented taxonomies to attempt to gain control over their mounting masses of information, creating an orderly structure to harness unstructured information (such as e-documents, e-mail messages, scanned records, and other digital assets), and to improve searchability and access.[5]

> Knowledge workers spend at least 15 to 25 percent of the workday searching for information, with only half the searches being successful.

Taxonomies for electronic records management (ERM) standardize the vocabulary used to describe records, making it easier and faster for searches and retrievals to be made.

Search engines can deliver faster and more accurate results from good taxonomy design by limiting and standardizing terms. A robust and efficient taxonomy design is the underpinning that indexes collections of documents uniformly and helps knowledge workers find the proper files to complete their work. The way a taxonomy is organized and implemented is critical to the long-term success of any enterprise, as it directly impacts the quality and productivity of knowledge workers who need organized, trusted information to make business decisions.

It does not sound so complicated, simply categorizing and cataloguing information, yet most enterprises have had disappointing or inconsistent results from the taxonomies they use to organize information. *Designing taxonomies is hard work.* Developing an efficient and consistent taxonomy is a detailed, tedious, labor-intensive team effort on the front end, and its maintenance must be consistent and regular and follow established **information governance** (IG) guidelines, to maintain its effectiveness.

Once a taxonomy is in place, it requires systematic updates and reviews, to ensure that guidelines are being followed and new document and record types are included in the taxonomy structure. Technology tools like **text mining, social tagging,** and **autoclassification** can help uncover trends and suggest candidate terms (more on these technologies later in this chapter).

When done correctly, the business benefits of good taxonomy design go much further than speeding search and retrieval; an efficient, operational taxonomy also is a part of IG efforts that help the organization to manage and control information so that it may efficiently respond to litigation requests, comply with governmental regulations, and meet customer needs (both external and internal).

Taxonomies are crucial to finding information and optimizing knowledge worker productivity, yet some surveys estimate that nearly half of organizations do not have a standardized taxonomy in place.[6]

> To maximize efficient and effective retrieval of records for legal, business, and regulatory purposes, organizations must develop and implement taxonomies.

According to the Montague Institute, "The way your company organizes information (i.e. its taxonomy) is critical to its future. A taxonomy not only frames the way people make decisions, but also helps them find the information to weigh all the alternatives. *A good taxonomy helps decision makers see all the perspectives, and 'drill down' to get details from each* and *explore lateral relationships among them*" (italics added).[7] Without it, your company will find it difficult to leverage intellectual capital, engage in electronic commerce, keep up with employee training, and get the most out of strategic partnerships.

With the explosion in growth of electronic documents and records, a standardized classification structure that a taxonomy imposes optimizes records retrievals for daily business operations and legal and regulatory demands.[8]

Since end-users can choose from topic areas, subject categories, or groups of documents, rather than blindly typing word searches, *taxonomies narrow searches and speed search time and retrieval.*[9]

> Taxonomies speed up the process of retrieving records because end-users can select from subject categories or topics.

"The link between taxonomies and usability is a strong one. The best taxonomies efficiently guide users to exactly the content they need. Usability is judged in part by how easily content can be found," according to the Montague Institute.[10]

Importance of Navigation and Classification

Taxonomies need to be considered from two main perspectives: navigation and classification. *Most people consider the former, but not the latter*. The navigational construct that is represented by a taxonomy is evident in most file structures and file shares—the nesting of folders within folders—and in many Web applications where users are navigating hierarchical arrangements of pages or links. However, classification is frequently behind the scenes. A document can "live" in a folder that the user can navigate to. *But within that folder, the document can be classified in different ways through the application of metadata*. In these cases, the records indicate what business function created them. Metadata are descriptive fields that delineate a (document or) record's characteristics, such as author, title, department of origin, date created, length, number of pages or file size, and so forth. The metadata is also part of the taxonomy or related to the taxonomy. In this way, usability can be impacted by giving the user *multiple ways* to retrieve their information,[11] while still maintaining the authenticity and evidence trail of the business function.

> Taxonomies need to be considered from two main perspectives: navigation and classification.

When Is a New Taxonomy Needed?

In some cases, organizations have existing taxonomy structures, but they have gone out of date or have not been maintained. They may not have been developed with best practices in mind or with correct representation of user groups, tasks, or applications. There are many reasons why taxonomies no longer provide the full value that they can provide. There are certain situations that clearly indicate that the

organization needs a refactored or new taxonomy.[12]

> Poor search results, inconsistent or conflicting file plans, and the inability to locate information on a timely basis are indications taxonomy work is needed.

If knowledge workers in your organization regularly conduct searches and receive hundreds of pages of results, then you need a new taxonomy. If you have developed a vast knowledge base of documents and records, and designated **subject matter experts** (SMEs), yet employees struggle to find answers, you need a new taxonomy. If there is no standardization of the way content is classified and catalogued, or there is conflict between how different groups or business units classify content, you need a new taxonomy. And if your organization has experienced delays, fines, or undue costs in producing documentation to meet compliance requests or legal demands, your organization needs to work on a new taxonomy.[13]

Taxonomies Improve Search Results

Taxonomies can improve a search engine's ability to deliver results to user queries in finding documents and records in an enterprise. The way the digital content is indexed (e.g. spidering, crawling, rule sets, algorithms) is a separate issue, and a good taxonomy improves search results regardless of the indexing method.[14]

Search engines struggle to deliver accurate and refined results since the wording in queries may vary and words can have multiple meanings. A taxonomy addresses these problems since the terms are set and defined in a **controlled vocabulary**.

> Taxonomies improve search results by increasing search accuracy and improving the user experience.

Metadata, which, as stated earlier, are data fields that describe content, such as document type, creator, date of creation, and so forth, *must be leveraged in the taxonomy design effort*.

A formal definition of metadata is “standardized administrative or descriptive data about a document [or record] that is common for all documents [or records] in a given repository.” Standardized metadata elements of e-documents should be utilized and supported by including them in controlled vocabularies when possible.[15]

The goal of a taxonomy development effort is to help users find the information they need, in a logical and familiar way, even if they are not sure what the correct search terminology is. *Good taxonomy design makes it easier and more comfortable for users to browse topics and drill down into more narrow searches to find the documents and records they need*. Where it really becomes useful and helps contribute to productivity is when complex or compound searches are conducted.

> Metadata, which are the characteristics of a document expressed in data fields, must be leveraged in taxonomy design.

Metadata and Taxonomy

One potential limitation of a purely hierarchical taxonomy is the lack of association between tiers (or nodes). There are often one-to-many or many-to-many associations between records. For example, an employee travels to a certification course. The resultant “expense report” is classified in the Finance/Accounts Payable/Travel Expense node of the taxonomy. The “course completion certificate” that is generated from the same travel (and is included as backup documentation for the expense report) is appropriately classified in the Human Resources/Training and Certification/Continuing Education node. *For ERM systems that do not provide the functionality for a multifaceted taxonomy, metadata is used to provide the link between the nodes in the taxonomy* (see Figure A.1).

Metadata schema must be structured to provide the appropriate associations as well as meet the users’

keyword search needs. *It is important to limit the number of metadata fields that a user must manually apply to records.* Most recordkeeping systems provide the functionality to automatically assign certain metadata to records based on rules that are established in advance and set up by a system administrator (referred in this book as **inherited metadata**). The record's classification or location in the taxonomy is appropriate for inherited metadata.

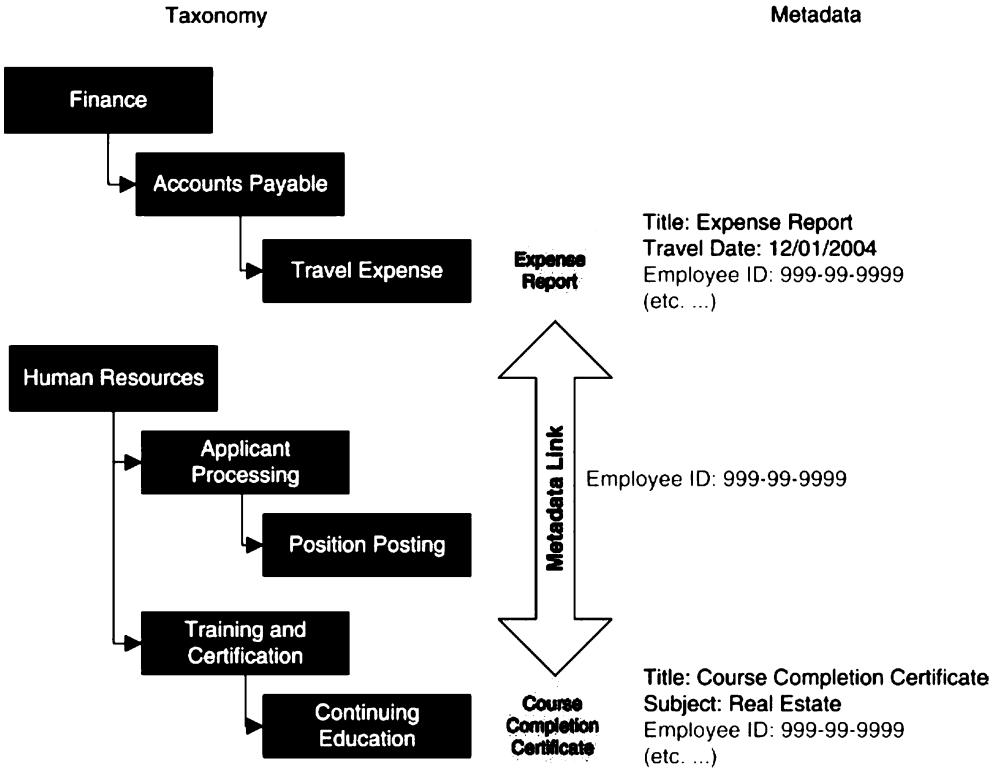


Figure A.1 Metadata Link to Taxonomy Example

Source: Blackburn Consulting.

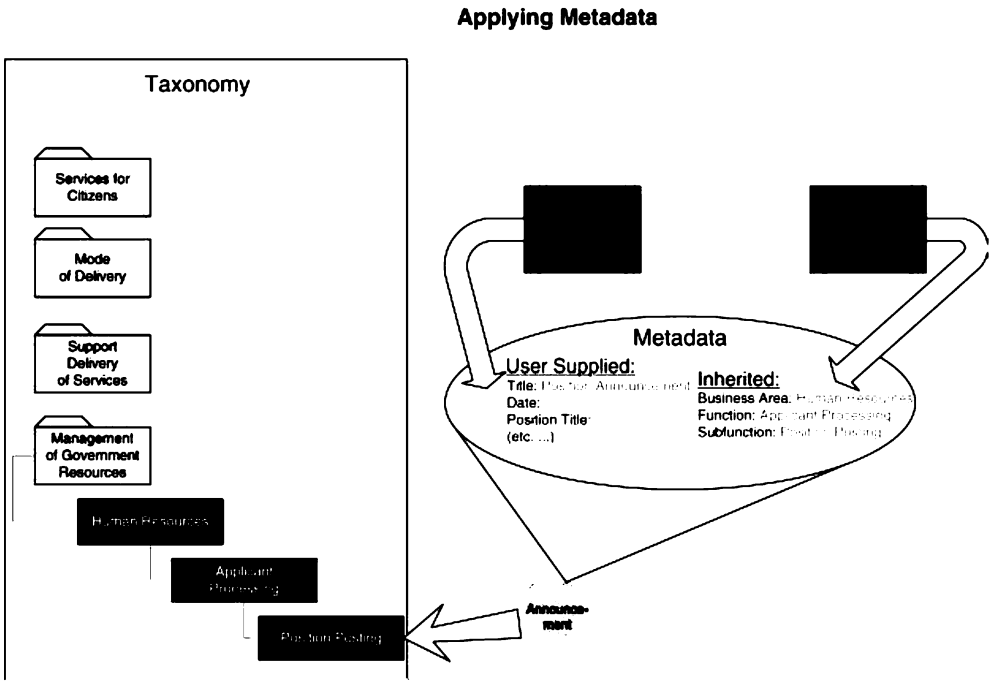


Figure A.2 Application of Metadata to Taxonomy Structure

Metadata can also be applied by autocategorization software. This can reduce the level of burden placed on the user and increase the quality and consistency of metadata. These approaches need to be tested and fine-tuned to ensure that they meet the needs of the organization.[16]

The File Plan will provide the necessary data to link the taxonomy to the document via inherited metadata. In most systems, this metadata is applied by the system and is transparent to the users. Additional metadata will need to be applied by the user. To maintain consistency, a **thesaurus**, which contains all synonyms and definitions, is used to enforce naming conventions (see Figure A.2).

Metadata Governance, Standards, and Strategies

Metadata can be a scary term to a lot of people. It just *sounds* complicated. And it can get complicated. It is often defined as "data about data," which is true but somewhat confusing, and this does not provide enough information for most people to understand.

"Meta" derives from the Greek word that means "alongside, with, after, next." Metadata can be defined as "structured data about other data."[17]

In **electronic records management** (ERM), metadata identifies a record and its contents. *ERM metadata describes a record's characteristics so that it may be classified more easily and completely*. Metadata fields, or *terms,* for e-records can be as basic as identifying the name of the document, the creator or originating department, the subject, the date it was created, the document type, the length of the document, its security classification, and its file type.

Creating standardized metadata terms is part of an information governance (IG) effort that enables faster, more complete, and more accurate searches and retrieval of records. This is important not only in everyday business operations, but also, for example, when searching through potentially millions of records during the discovery phase of litigation.

Good metadata management also assists in the maintenance of corporate memory and improving accountability in business operations.[18]

Using a standardized format and controlled vocabulary provides a "precise and comprehensible description of content, location, and value."[19] *Using a controlled vocabulary means your organization has standardized a set of terms used for metadata elements describing records*. This "ensures consistency across a collection" and helps with optimizing search and retrieval functions and records research, as well as

meeting e-discovery requests, compliance demands, and other legal and regulatory requirements. Your organization may, for instance, decide to use the standardized Library of Congress Subject Headings as standard terms for the "subject" metadata field.[20]

Metadata also describes a record's relationships with other documents and records, and what actions may have been taken on the record over time. This helps to track its history and development, and aid in any future e-discovery requests.

The role of metadata in managing records is multifaceted; it helps to:

- Identify the records, record creators and users, and the areas within which they are utilized.
- Determine the relationships between records and the knowledge workers who use them, and the relationships between the records and the business processes they are supporting.
- Assist in managing and preserving the content and structure of the record.
- Support IG efforts that outline who has access to records, and the context (when and where) in which access to the records is granted.
- Provide an audit trail to document changes to or actions upon the record and its metadata.
- Support the finding and understanding of records and their relationships.[21]

> Metadata terms or fields describe a record's characteristics so that it may be classified, managed, and found more easily.

In addition, good metadata management provides additional business benefits including increased management control over records, improved records authenticity and security, and reusability of metadata.[22]

Often, organizations will establish mandatory metadata terms that must accompany a record, and some optional ones that may help in identifying and finding it. *A record is more complete with more metadata terms included, which also facilitates search and retrieval of records.*[23] This is particularly the case when knowledge workers are not quite sure which records they are searching for, and therefore enter some vague or conceptual search terms. So, the more detail that is in the metadata fields, the more likely the end user is to find the records they need to complete their work. This provides a measurable productivity benefit to the organization, although it is difficult to quantify. Certainly, search times will decrease upon implementation of a standardized metadata program, and improved work output and decisions will also follow.

> Metadata terms can be as basic as the name of the document, the creator, the subject, the date it was created, the document type, the length of the document, its security classification, and its file type.

Standardizing the metadata terms, definitions, and classifications for documents and records is done by developing and enforcing IG policy. This standardization effort gives users confidence that the records they are looking for are, in fact, the complete and current set they need to work with. And it provides the basis for a *legally defensible* records management program that will hold up in court.

A metadata governance program must be an ongoing effort that keeps metadata up-to-date and accurate. Often, once a metadata project is complete, attention to it wanes and maintenance tasks are not executed and soon the accuracy and completeness of searches for documents and records deteriorates. So metadata maintenance is an ongoing process and it must be formalized into a program that is periodically checked, tested, and audited.

> A metadata governance and management program must be ongoing.

Types of Metadata

There are several types or categories of metadata, including:

Descriptive metadata. Metadata that describes the intellectual content of a resource and is used for the indexing, discovery, and identification of a digital resource.

Administrative metadata. Metadata that includes management information about the digital resource, such as ownership and rights management.

Structural metadata. Metadata that is used to display and navigate digital resources and describes relationships between multiple digital files, such as page order in a digitized book.

Technical metadata. Metadata that describes the features of the digital file, such as resolution, pixel dimension, and hardware. The information is critical for migration and long-term sustainability of the digital resource.

Preservation metadata. Metadata that specifically captures information that helps facilitate management and access to digital files over time. This inherently includes descriptive, administrative, structural, and technical metadata elements that focus on the provenance, authenticity, preservation activity, technical environment, and rights management of an object.[24]

The main types of metadata are: descriptive, administrative, structural, technical, and preservation metadata.

Core Metadata Issues

Some key considerations and questions that need to be answered for effective implementation of a metadata governance program are:

- *Who is the audience?* Which users will be using the metadata in their daily operations? What is their skill level? Which metadata terms/fields are most important to them? What has been their approach to working with documents and records in the past and how can it be streamlined or improved? What terms are important to management? How can the metadata schema be designed to accommodate the primary audience and other secondary audiences? Answers to these questions will come only with close consultation with these key stakeholders.[25]
- *Who else can help?* That is, which other stakeholders can help build a consensus on the best metadata strategy and approach? What other records creators, users, custodians, auditors, and legal counsel personnel can be added to the team to design a metadata approach that maximizes its value to the organization? Are there subject matter experts (SMEs)? What standards and best practices can be applied across functional boundaries to improve the ability of various groups to collaborate and leverage the metadata?
- *How can metadata governance be implemented and maintained?* Creating IG guidelines and rules for metadata assignment, input, and upkeep are a critical step—but how will the program continue to be updated to maintain its value to the organization? What business processes and audit checks should be in place? How will the quality of the metadata be monitored and controlled? Who is accountable?
- *What will the user training program look like?* How will users be trained initially, and how will continued education and reinforcement be communicated? Will there be periodic meetings of the IG or metadata team to discuss issues and concerns? What is the process for adding or amending metadata terms as the business progresses and changes? These questions must be answered, and a documented plan must be in place.
- *What will the communications plan be?* Management time and resources are also needed to continue the practice of informing and updating users and encouraging compliance with internal metadata standards and policies. Users need to know on a consistent basis why metadata is important and the value that good metadata management can bring to the organization.[26]

International Metadata Standards and Guidance

Metadata is what gives an e-record its record status, or, in other words, electronic records metadata is what makes an electronic file a record. There are several established international standards for metadata

structure, and additional guidance on strategy and implementation has been provided by standards groups such as ISO and ANSI/NISO, and other bodies, such as the Dublin Core Metadata Initiative (DCMI).

ISO 15489 Records Management Definitions and Relevance

The international records management standard ISO 15489 states that "a record should correctly reflect what was communicated or decided or what action was taken. It should be able to support the needs of the business to which it relates and be used for accountability purposes" and its metadata definition is "data describing context, content, and structure of records and their management through time."[27]

A key difference between a document and a record is that a record is fixed, whereas a document can continue to be edited. (This line has been blurred with the advent of blockchain technology, which keeps records in a sequence, and creates a new record each time a change or update is made.) Preventing records from being edited can be partly accomplished by indicating their formal record status in a metadata field, among other controls.

Proving that a record is, in fact, authentic and reliable, necessarily includes proving that its metadata has remained intact and unaltered through the entire chain of custody of the record.

> Proving that a record is authentic and reliable includes proving that its metadata has remained intact and unaltered through the record's entire chain of custody.

ISO Technical Specification 23081–1: 2006 Information and Documentation—Records Management Processes—Metadata for Records—Part 1: Principles

ISO 23081–1 "covers the principles that underpin and govern records management metadata. These principles apply through time to:

- Records and their metadata;
- All processes that affect them;
- Any system in which they reside;
- Any organization that is responsible for their management.[28]

The ISO 23081–1 standard provides guidance for metadata management within the "framework" of ISO 15489, and addresses the relevance and roles that metadata plays in records management intensive business processes. There are *no mandatory* metadata terms set, as these will differ by organization and by location and governing national and state/provincial laws.[29] The standard lists 10 purposes or benefits of using metadata in records management, which can help build the argument for convincing users and managers of the importance of good metadata governance and its resultant benefits.

> ISO 23081 defines needed metadata for records, and provides guidance for metadata management within the "framework" of ISO 15489.

Dublin Core Metadata Initiative

The DCMI produced a basic or core set of metadata terms that have served as the basis for many public and private sector metadata governance initiatives. Initial work in workshops filled with experts from around the world took place in 1995 in Dublin, Ohio (*not* Ireland). From these working groups the idea of a set of "core metadata" or essential metadata elements with generic descriptions arose.[30] "The fifteen-element 'Dublin Core' achieved wide dissemination as part of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and has been ratified as IETF RFC 5013, ANSI/NISO Standard Z39.85–2007, and ISO Standard 15836:2009."

Goals of the Dublin Core Metadata Initiative are simplicity, commonly understood semantics, international scope, and extensibility.

"Dublin Core has as its goals:[31]

Simplicity of creation and maintenance

The Dublin Core element set has been kept as small and simple as possible to allow a non-specialist to create simple descriptive records for information resources easily and inexpensively, while providing for effective retrieval of those resources in the networked environment.

Commonly understood semantics

Discovery of information across the vast commons of the Internet is hindered by differences in terminology and descriptive practices from one field of knowledge to the next. The Dublin Core can help the 'digital tourist'—a non-specialist searcher—find his or her way by supporting a common set of elements, the semantics of which are universally understood and supported. For example, scientists concerned with locating articles by an author, and art scholars interested in works by a particular artist, can agree on the importance of a 'creator' element. Such convergence on a common, if slightly more generic, element set increases the visibility and accessibility of all resources, both within a given discipline and beyond.

International scope

The Dublin Core Element Set was originally developed in English, but versions are being created in many other languages, including Finnish, Norwegian, Thai, Japanese, French, Portuguese, German, Greek, Indonesian, and Spanish. The DCMI Localization and Internationalization Special Interest Group is coordinating efforts to link these versions in a distributed registry.

Although the technical challenges of internationalization on the World Wide Web have not been directly addressed by the Dublin Core development community, the involvement of representatives from virtually every continent has ensured that the development of the standard considers the multilingual and multicultural nature of the electronic information universe.

Extensibility

While balancing the needs for simplicity in describing digital resources with the need for precise retrieval, Dublin Core developers have recognized the importance of providing a mechanism for extending the DC element set for additional resource discovery needs. It is expected that other communities of metadata experts will create and administer additional metadata sets, specialized to the needs of their communities. Metadata elements from these sets could be used in conjunction with Dublin Core metadata to meet the need for interoperability. The DCMI Usage Board is presently working on a model for accomplishing this in the context of 'application profiles.'

"The fifteen element 'Dublin Core' described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The full set of vocabularies, DCMI Metadata Terms [DCMI-TERMS], also includes sets of resource classes (including the DCMI Type Vocabulary [DCMI-TYPE]), vocabulary encoding schemes, and syntax encoding schemes. The terms in DCMI vocabularies are intended to be used in combination with terms from other, compatible vocabularies in the context of application profiles and on the basis of the DCMI Abstract Model [DCAM]."[32]

Global Information Locator Service

Global Information Locator Service (GILS) is ISO 23950, the international standard for information searching over networked (client/server) computers, which is a simplified version of structured query language (SQL). ISO 23950 is a federated search protocol that equates to the US standard ANSI/NISO Z39.50. The US Library of Congress is the official maintenance agency for both standards, "which are technically identical (though with minor editorial differences)."[33]

ISO 23950 (GILS) is the international standard for information searching over networked computers.

ISO 23950 (also known as ANSI/NISO standard Z39.50) grew out of the library science community, although it is widely used, particularly in the public sector.[34] The use of GILS has tapered off as other metadata standards, at the international, national, industry level, and agency level have been established.[35]

"It [GILS] specifies procedures and formats for a client to search a database provided by a server, retrieve database records, and perform related information retrieval functions." While it does not specify a format, information retrieval can be accomplished through full-text search, although it "also supports large, complex information collections." The standard specifies how searches are made and how results are returned.

GILS helps people find information, especially in large, complex environments, such as across multiple government agencies. It is used in over 40 US states and several countries, including Argentina, Australia, Brazil, Canada, France, Germany, Hong Kong, India, Spain, Sweden, Switzerland, United Kingdom, and many others.[36]

Text Mining

On a continuing basis, text mining can be conducted on documents to learn of emerging potential taxonomy terms. Text mining is simply performing detailed full-text searches on the content of document. And with more sophisticated tools like neural computing and artificial intelligence (AI), *concepts,* not just key words, can be discovered and leveraged for improving search quality for users.

Text mining is simply performing detailed full-text searches on the content of document.

Another tool is the use of **faceted search** (sometimes referred to as faceted navigation or faceted browsing) where, for instance, document collections are classified in multiple ways, rather than in a single, rigid taxonomy. Knowledge workers may apply multiple filters to search across documents and records and find better and more complete results. And when they are not quite sure what they are looking for, or if it exists, then a good taxonomy can help suggest terms, related terms, and associated content, truly contributing to enterprise **knowledge management** (KM) efforts, adding to corporate memory and increasing the organizational knowledge base.[37] Good KM helps to provide valuable training content for new employees, and helps to reduce the impact of turnover and retiring employees.

Search is ultimately about metadata—whether your content has explicit metadata or not. The search engine creates a forward index and determines what words are contained in the documents being searched. It then inverts that index to provide the documents that words are contained in. This is effectively metadata about the content. A taxonomy can be used to enrich that search index in various ways. This does require configuration and integration with search engines, but the result is the ability to increase both precision and recall of search results. Search results can also be grouped and clustered using a taxonomy. This allows large numbers of results to be more easily scanned and understood by the user. Many of these functions are determined by the capabilities of search tools and document and records management systems. As search functionality is developed, don't miss this opportunity to leverage the taxonomy.

Records Grouping Rationale

The primary reasons that records are grouped together are:

- They tie together documents with like content, purpose, or theme.
- To improve search and retrieval capabilities.
- To identify content creators, owners, and managers.
- To provide an understandable context.
- For retention and disposition scheduling purposes.[38]

Taxonomies group records with common attributes. The groupings are constructed not only for records management classification and functions, but also to support end users in their search and retrieval activities. Associating documents of a similar theme enables users to find documents when they do not know the exact document name. Choosing the theme or topic enables the users to narrow their search to find the relevant information.

The theme or grouping also places the document name into context. Words have many meanings and adding a theme to them further defines them. For example, the word "article" could pertain to a newspaper article, an item or object, or a section of a legal document. If it were grouped with publications, periodicals, and so on the meaning would be clear. The challenge here is when to choose to have a separate category for "article" or to group "article" with other similar publications. Some people tend to develop finer levels of granularity in classification structures. These people can be called the "splitters." Those who group things together are "lumpers." *But there can be clear rules for when to lump versus split.* Experts recommend splitting into another category when business needs demand that we treat the content differently or users need to segment the content for some purpose. This rule can be applied to many situations when trying to determine whether a new category is needed.[39]

Management, security, and access requirements are usually based on a user's role in a process. Grouping documents based on processes makes the job of assigning the responsibilities and access easier. For example, documents used in financial processes can be sensitive and there is a need to restrict access to only those users that have the role in the business with a need to know.

Records retention periods are developed to be applied to a series (or group) of documents. When similar documents are grouped, it is easier to apply retention rules. However, when the grouping for retention is not the same as the grouping for other user views, a cross-mapping (**file plan**) scheme must be developed and incorporated into the taxonomy effort.

Business Classification Scheme, File Plans, and Taxonomy

In its simplest definition a **business classification scheme** (BCS) is a hierarchical conceptual representation of the business activity performed by an organization.[40] The highest level of a BCS is called an *Information Series,* which signifies "high-level business functions" of a business or governmental agency, and the next level is *Themes,* which represent the specific activities that feed into the high-level functions at the information series level. These two top levels are rarely changed in an organization.[41]

A BCS is often viewed as synonymous with the term ***file plan,*** which is the shared file structure in an Electronic Records Management (ERM) System, but it is *not a direct file plan.*

Yet, a file plan can be developed and mapped back to the BCS and automated through an electronic document and records management system (EDRMS) or electronic records management (ERM) system.[42]

A BCS is required by ISO 15489, the international records management standard, and, together with the folders and records it contains, comprises what in the paper environment was called simply a "File plan." A BCS is therefore a full representation of the business of an organization.

Classification and Taxonomy

Classification of records extends beyond the categorization of records in the taxonomy. It also must include the application of retention requirements. These are legal and business requirements that specify the length of time a record must be maintained. A **Records Retention Schedule** (RRS) is a document that identifies regulatory relevant records and specifies the periods for which an organization should retain these records to meet its operational needs. The RRS also is a guide that indicates legal and other statutory requirements. *The Records Retention Schedule groups documents into records series that relate to specific business functions.* This grouping is performed because laws and regulations are mainly based on the *business functions* that creates the documents. These business functions are not necessarily the same as the activities described in the hierarchy of the taxonomy. Therefore, there must be a method to map the RRS to the Taxonomy. This is accomplished with a File Plan. The File Plan facilitates the application of retention rules during document categorization without requiring a user to know or understand the Records Retention Schedule (see Figure A.3).

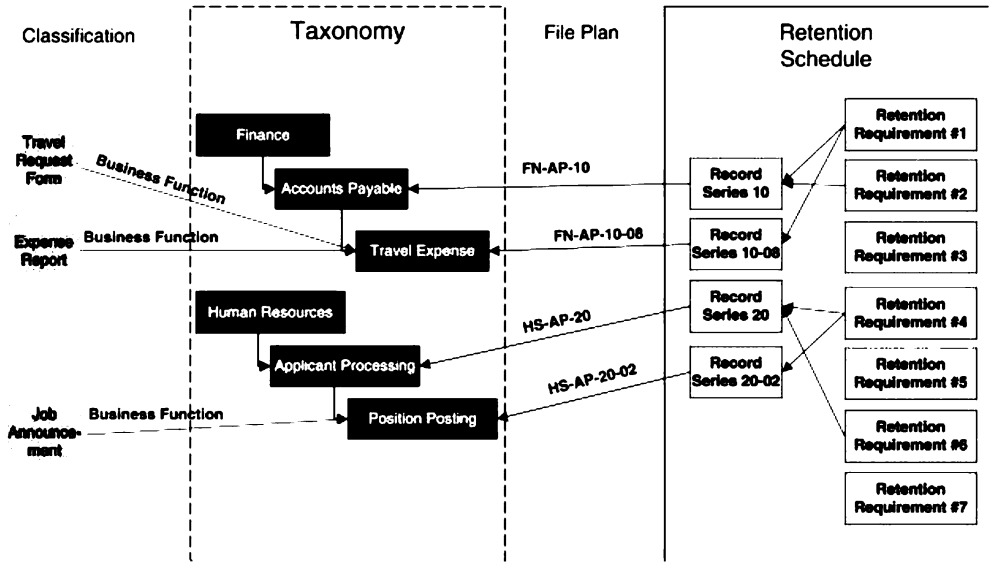


Figure A.3 Mapping the Records Retention Schedule to the Taxonomy

Source: Blackburn Consulting.

Prebuilt Versus Custom Taxonomies

Taxonomy templates for specific vertical industries (e.g. law, pharmaceuticals, aerospace) are provided by content and knowledge management software, enterprise search vendors, and trade associations. These prebuilt taxonomies use consistent terminology, have been tried and tested, and incorporate industry best practices, where possible. They can provide a jump-start and faster implementation at a lower cost than developing a custom taxonomy in-house or with external consulting assistance.

There are advantages and disadvantages to each approach. A prebuilt taxonomy will typically have some parameters that are able to be configured to better meet the business needs of an organization, yet compromises and trade-offs will have to be made. It may also introduce unfamiliar terminology that knowledge workers will be forced to adapt to, increasing training time and costs, and reducing its overall effectiveness. These considerations must be factored into the "build or buy" decision. Using the custom-developed approach, a taxonomy can be tailored to meet the precise business needs of an organization or business unit and can include nuances such as company-specific nomenclature and terminology.[43]

Frequently, the longer and more costly customized approach must be used, since there are no prebuilt taxonomies that fit well. This is especially the case with niche enterprises or those operating in developing or esoteric markets. For mature industries, more prebuilt taxonomies and template choices exist. *Attempting to tailor a prebuilt taxonomy can end up taking longer than building one from scratch if it is not a good fit in the first place,* so best practices dictate that organizations use prebuilt taxonomies where practical, and custom design taxonomies where needed.

There really is no "one size fits all" when it comes to taxonomy. And even when two organizations do the exact same thing in the exact same industry, there will be differences in their culture, process, and content that will require customization and tuning of the taxonomy. Standards are useful for improving efficiency of a process, and taxonomy projects really are internal standards projects. However, competitive advantage is attained through differentiation. A taxonomy specifically tuned to meet the needs of a particular enterprise is actually a competitive advantage.[44]

There is one other alternative, which is to "autogenerate" a taxonomy from the metadata in a collection of e-documents and records by using sophisticated statistical techniques like term frequency and entity extraction to attempt to create a taxonomy. It seems to be perhaps the "best of both worlds" in that it offers instant customization at a low cost, but, although these types of tools can help provide useful insights into the data on the front end of a taxonomy project, providing valuable statistical renderings, the only way to focus on user needs is to interview and work with users to gain insights into their business process needs and requirements, while considering the business objectives of the taxonomy project. This cannot be done with mathematical computations—the human factor is key.

Best practices dictate that taxonomy development includes designing the taxonomy structure and heuristic principles to align with user needs.

In essence, these autogenerated taxonomy tools can determine which terms and documents are used frequently, but they cannot assess the *real value* of information being used by knowledge workers and *how* they use the information. That takes consultation with stakeholders, studied observation, and business analysis. *Machine-generated taxonomies look like they were generated by machines*—which is to say that they are not very usable by humans.[45]

Thesaurus Use in Taxonomies

A **thesaurus** in the use of taxonomies contains the agreed-on synonyms and similar names for terms used in a controlled vocabulary. So, "invoice" may be listed as the equivalent term for "bill" when categorizing records. The thesaurus goes further and lists "information about each term and their relationships to other terms within the same thesaurus."

A thesaurus is like a hierarchical taxonomy but also includes "associative relationships."[46] An associative relationship is a conceptual relationship. It is the *"see also"* that we may come across in the back-of-the-book index. But the question is: Why do we want to see it? Associative relationships can provide a linkage to specific classes of information of interest to users and for processes. Use of associative relationships can provide a great deal of functionality in content and document management systems and needs to be considered in records management applications.[47]

There are international standards for thesauri creation from International Organization for Standardization (ISO), American National Standards Institute (ANSI), and the British Standards Institution (BSI).[48]

ISO 25964-1:2011, "Information and Documentation—Thesauri and Interoperability with Other Vocabularies," draws "on [the British standard, BS 8723] but reorganize[d] the content to fit into two parts." Part 1, "Thesauri for Information Retrieval," of the standard ISO 25964 was published in August 2011. Part 2, "Interoperability with Other Vocabularies," was published in 2013.[49]

Taxonomy Types

Taxonomies used in ERM systems are usually hierarchical where categories (nodes) in the hierarchy progress from general to specific. Each subsequent node is a subset of the higher-level function. There are three basic types of hierarchical taxonomies: subject, business-unit, and functional.[50]

A *subject* taxonomy uses controlled terms for subjects. The subject headings are arranged in alphabetical order by the broadest subjects, with more precise subjects listed under them. An example is the Library of Congress subject headings (LCSH) used to categorize holdings in a library collection (see Figure A.4). Even the Yellow Pages could be considered a subject taxonomy.

There are three basic types of hierarchical taxonomies: subject, business-unit, and functional.

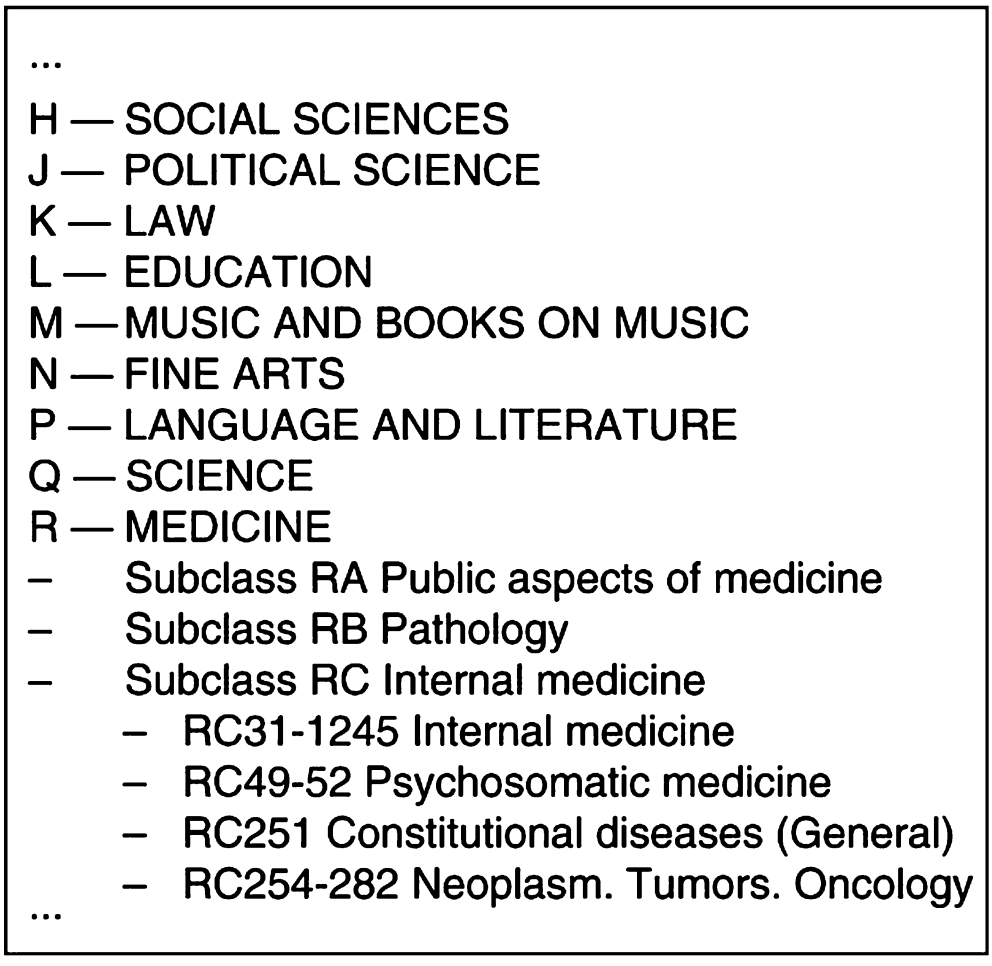


Figure A.4 Library of Congress Subject Headings

It is difficult to establish a universally recognized set of terms in a subject taxonomy. If users are unfamiliar with the topic, they may not know the appropriate term heading with which to begin their search. For example, say a person is searching through the Yellow Pages for a place to purchase eyeglasses. They begin their search alphabetically by turning to the E's and scanning for the term *eyeglasses*. Since there are no topics titled "eyeglasses," the person consults the Yellow Pages index, finds the term *eyeglasses*, and this provides a list of preferred terms or "see alsos" that direct the person to "Optical—Retail" for a list of eyeglass businesses (see Figure A.5).[51]

In both examples (LCSH and Yellow Pages), the subject taxonomy is supported by a thesaurus. Again, a thesaurus is a controlled vocabulary that includes synonyms, related terms, and preferred terms. In the case of the Yellow Pages, the index functions as a basic thesaurus.

In a *business-unit*-based taxonomy, the hierarchy reflects the organizational charts (e.g. Department/Division/Unit). Records are categorized based on the business unit that manages them. Figure A.6 shows the partial detail of one node of a business-unit based taxonomy that was developed for a county government.[52]

One advantage of a business-unit-based taxonomy is that it mimics most existing paper-filing system schemas. Therefore, users are not required to learn a "new" system. However, conflicts arise when documents are managed or shared among multiple business units. As an example, for the county government referenced earlier, a property transfer document called the "TD1000" is submitted to the Recording Office for recording and then forwarded to the Assessor for property tax evaluation processing. This poses a dilemma as to where to categorize the TD1000 in the taxonomy.[53]

Another issue arises with organizational changes. When the organizational structure changes, so must the business-unit based taxonomy.

In a *functional* taxonomy records are categorized based on the functions and activities that produce them (function/activity/transaction). The organization's business processes are used to establish the taxonomy. The highest or broadest level represents the business functions. The next level down the hierarchy constitutes the activities performed for the function. The lowest level in the hierarchy consists of the records that are created as a result of the activity (a.k.a., the *transactions*).

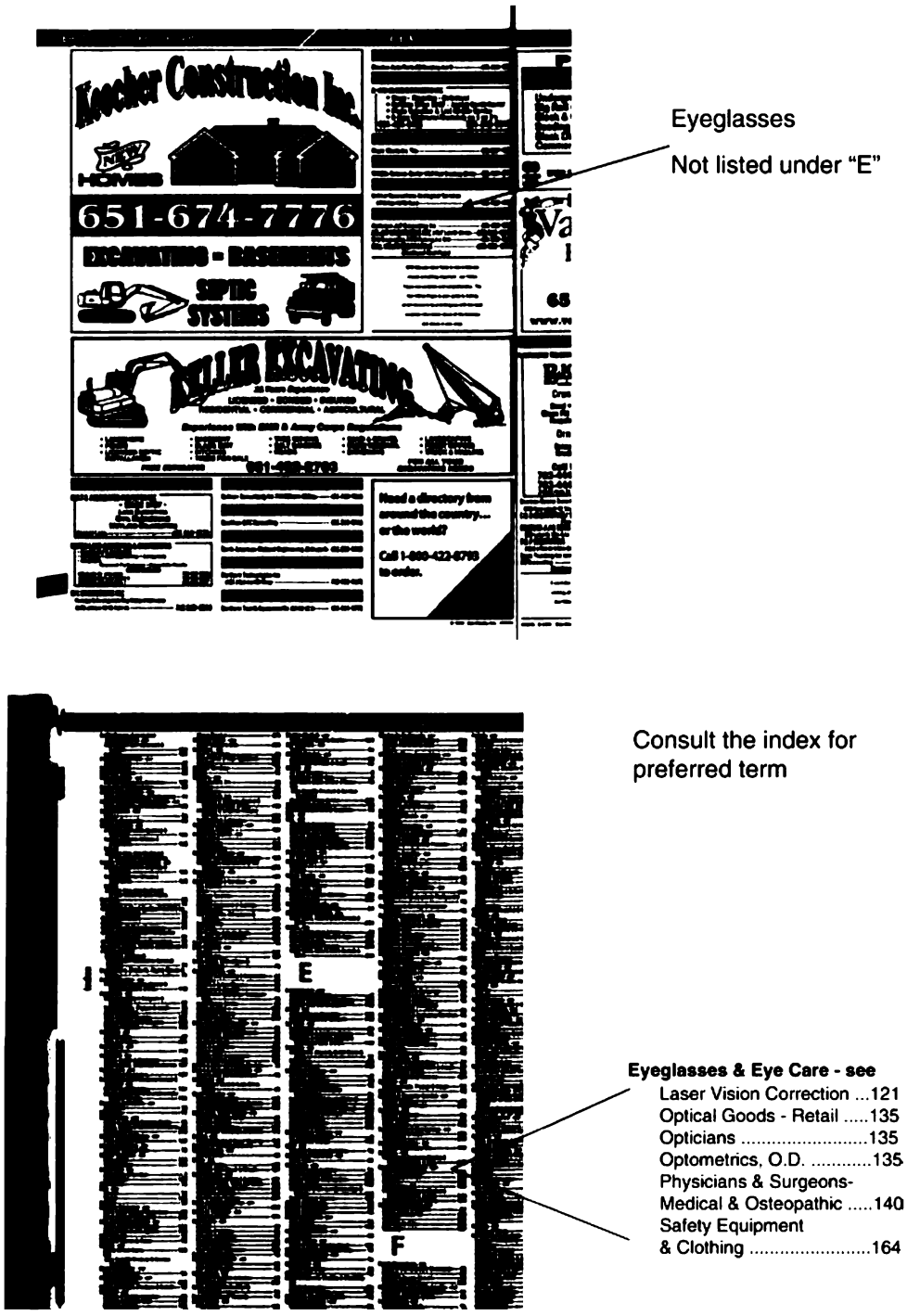


Figure A.5 Yellow Pages Example

Figure A.7 shows partial detail of one node of a functional taxonomy developed for a state government regulatory agency. The agency organizational structure is based on regulatory programs. Within the program areas are similar (repeated) functions and activities (e.g. permitting, compliance, and enforcement, etc.).

When the repeated functions and activities are universalized, the results are a "flatter" taxonomy. *This type of taxonomy is better suited to endure organizational shifts and changes*. In addition, the process of universalizing the functions and activities inherently results in broader and more generic naming conventions. This provides flexibility when adding new record types (transactions) because there will be fewer changes to the hierarchy structure.[54]

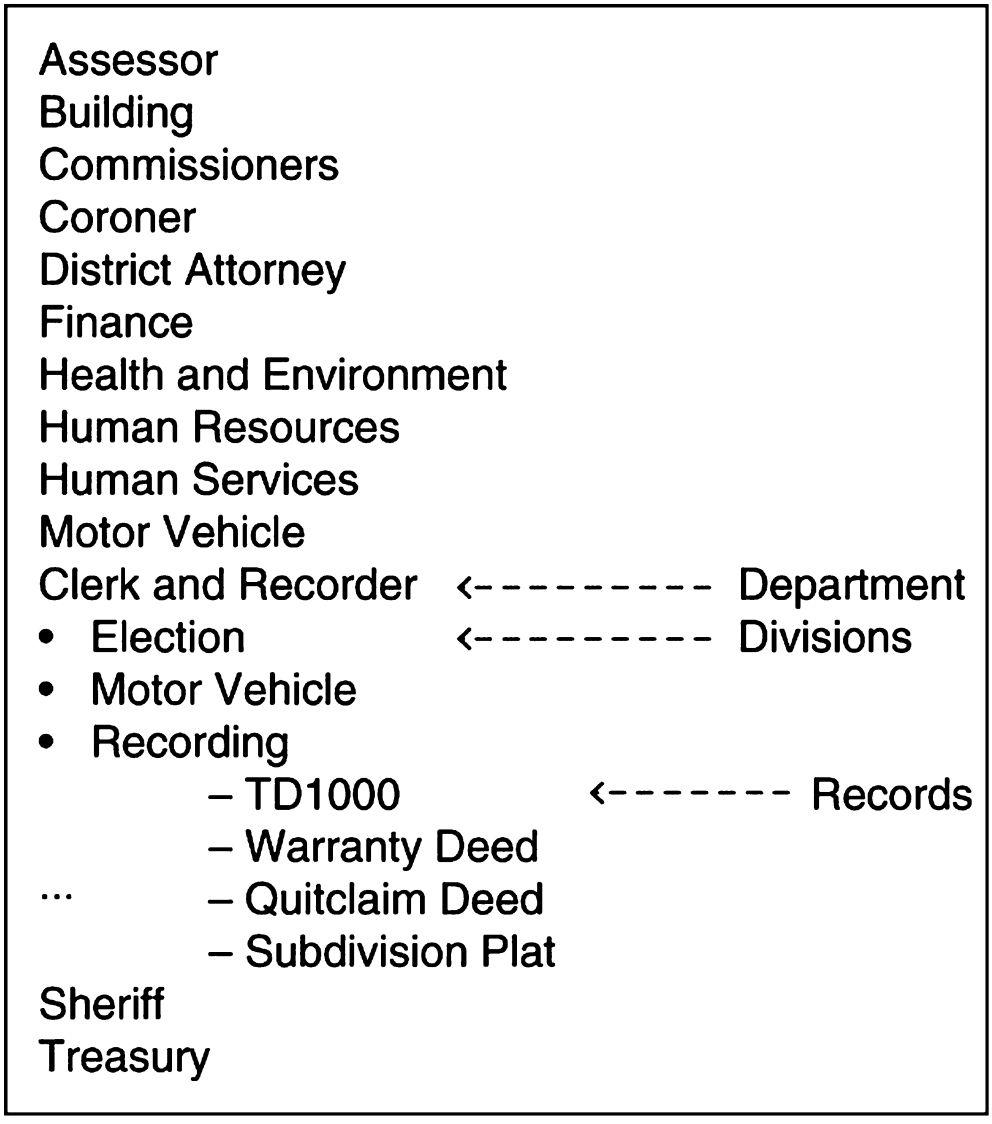


Figure A.6 Community Government Business-Unit Taxonomy

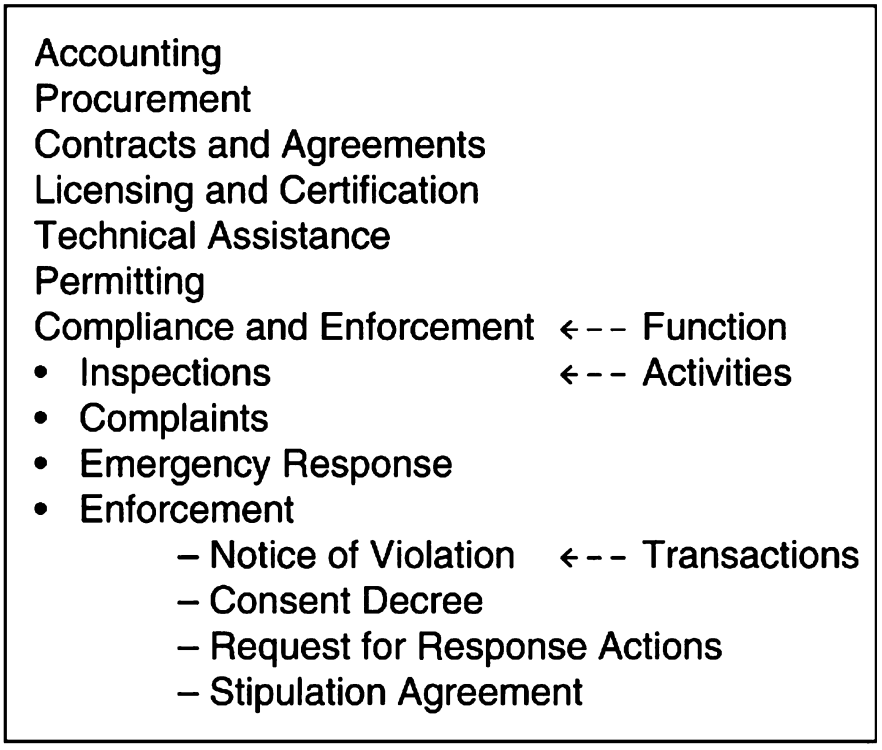
A functional taxonomy is better suited to endure organizational changes.

One disadvantage of a functional taxonomy is its inability to address case files (or project files). A case file is a collection of records that relate to a entity, person, or project. The records in the case file can be generated by multiple activities. For example, at the regulatory agency, enforcement files are maintained that contain records generated by enforcement activities (Notice of Violation, Consent Decree, etc.) and other ancillary, but related activities such as Contracting, Inspections, and Permitting.[55]

To address the case file issue at the regulatory Agency, metadata cross-referencing was used to provide a virtual case-file view of the records collection (see Figure A.8).

One disadvantage of a functional taxonomy is its inability to address case files (or project files).

A *hybrid* [taxonomy] is usually the best approach. There are certain business units that usually don't change over time. For example, accounting and human resources activities are fairly constant. Those portions of the taxonomy could be constructed in a business-unit manner even when other areas within the organization use a functional structure (see Figure A.9).[56]



Function	Activity
4. Permitting	4.1 Registration
	4.2 Application
	4.3 Public Notice
	4.4 Permit Development & Issuance
	4.5 Termination
5. Compliance and Enforcement	5.1 Inspections
	5.2 Complaints
	5.3 Emergency Response & Preparedness
	5.4 Monitoring Reporting
	5.5 Enforcement Actions

Figure A.7 State Government Regulatory Agency Functional Taxonomy

A hybrid approach to taxonomy design is usually the best.

Faceted taxonomies allow for multiple organizing principles to be applied to information along various dimensions. Facets can contain subjects, departments, business units, processes, tasks, interests, security levels, and other attributes used to describe information. There is never really one single taxonomy, but rather collections of taxonomies that describe different aspects of information. In the e-commerce world, facets are used to describe brand, size, color, price, and other context-specific attributes. Records management systems can also be developed with knowledge and process attributes related to the enterprise.[57]

Business Process Analysis

To establish the taxonomy, business processes must be documented and analyzed. There are two basic process analysis methods: top-down and bottom-up. In the top-down method, a high-level analysis of business functions is performed to establish the higher tiers. Detailed analyses are performed on each business process to "fill in" the lower tiers. The detailed analyses are usually conducted in a phased approach and the taxonomy is incrementally updated.

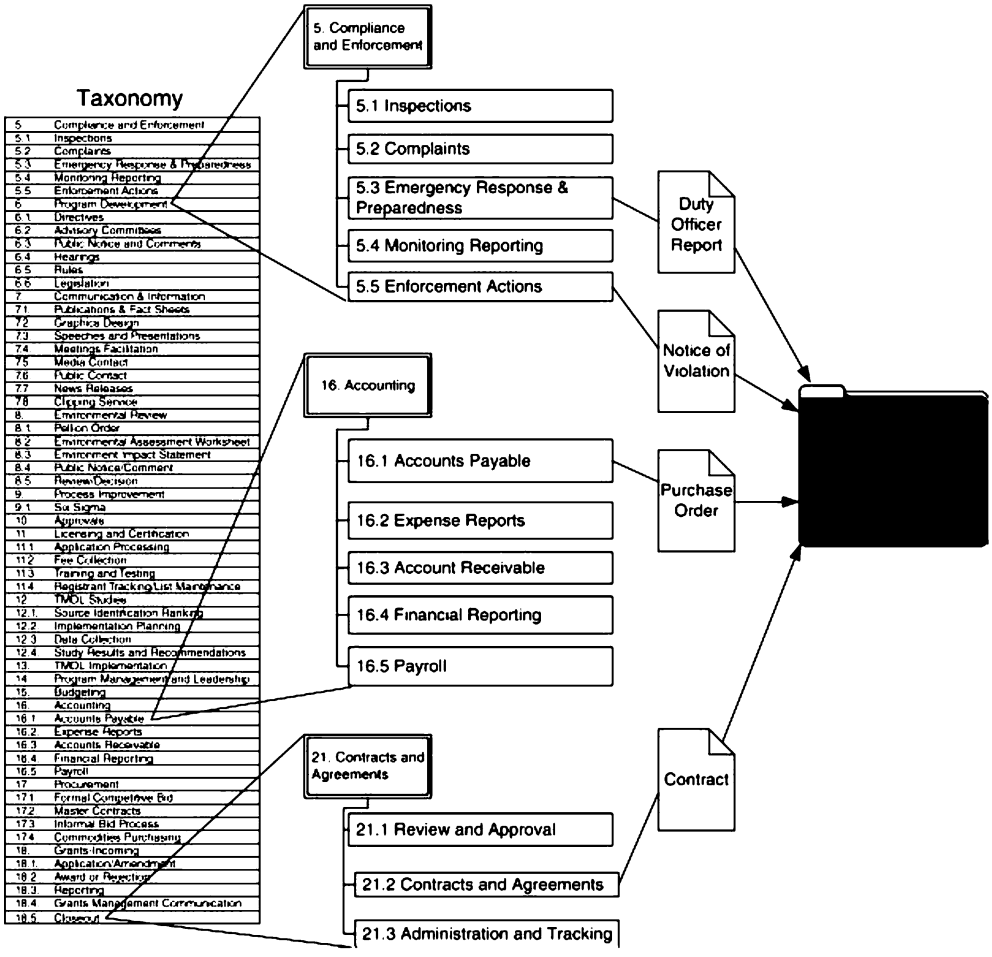


Figure A.8 Metadata Cross-Referencing within a Taxonomy

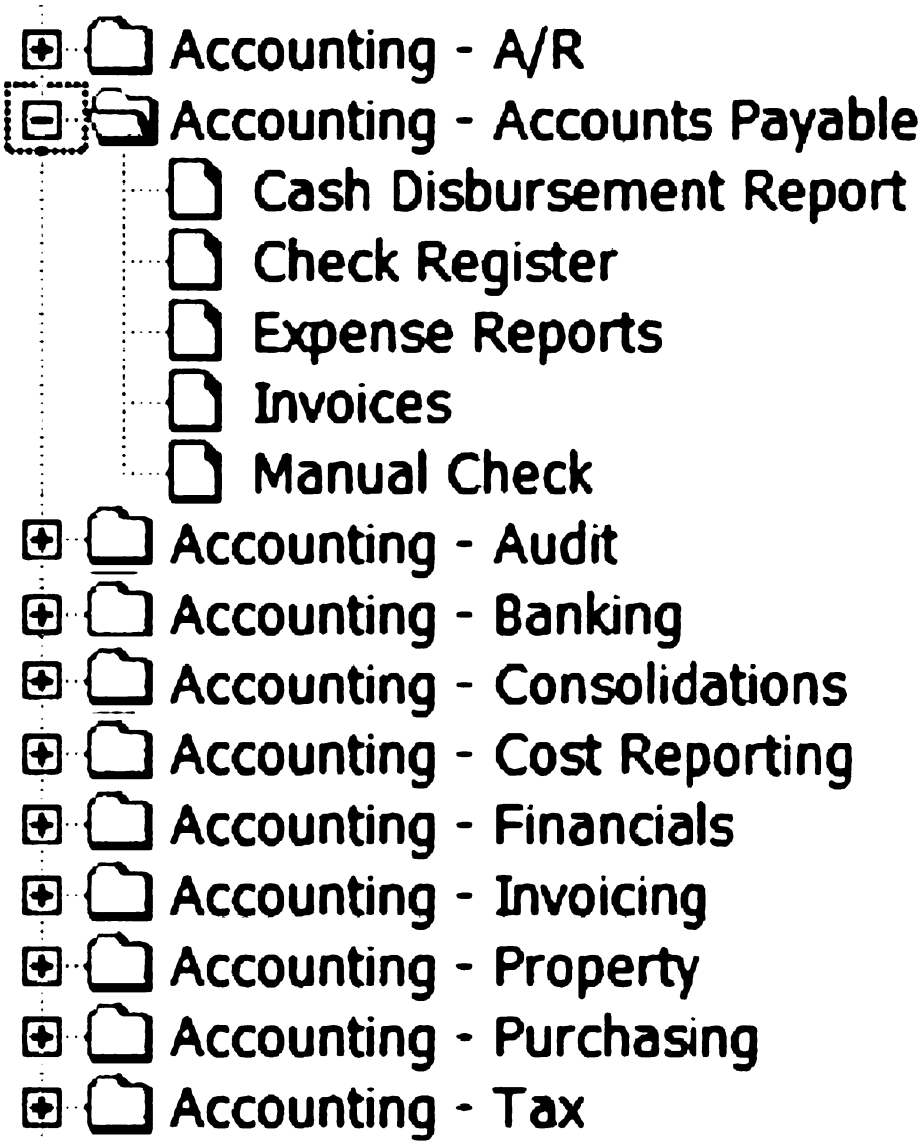


Figure A.9 Basic Accounting Business-Unit Taxonomy

> Business processes must be documented and analyzed to develop a taxonomy.

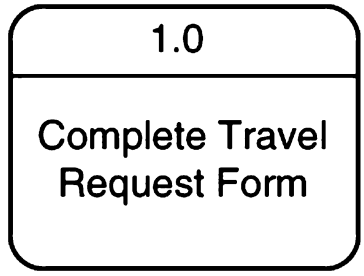
In order to use the bottom-up method, detailed analyses must be performed for all processes in one effort. Using this method ensures that there will be fewer modifications to the taxonomy. However, this is sometimes not feasible for organizations with limited resources. A phased or incremental approach is usually more budget-friendly and places fewer burdens on the organization's resources.

There are many diagramming formats and tools that will provide the details needed for the analysis. The most basic diagramming can be accomplished with a standard tool such as Visio® from Microsoft. There are also more advanced modeling tools that could be used to produce the diagrams that provide the functionality to statistically analyze process changes through simulation and provide information for architecture planning and other process initiatives within the organization.

Any diagramming format will suffice as long as it depicts the flow of data through the processes showing process steps, inputs, and outputs (documents), decision steps, organizational boundaries, and interaction with information systems. The diagrams should depict document movement within as well as between the

subject department and other departments or outside entities.

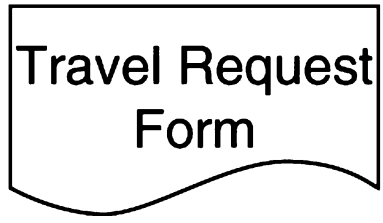
Figure A.10 uses a swim-lane type diagram. Each horizontal "lane" represents a participant or role. The flow of data and sequence of process steps is shown with lines (the arrows note the direction). Process steps are shown as boxes.



Decision steps are shown as diamonds.



Documents are depicted as a rectangle with a curved bottom line.



The first step is to review any existing business process documentation (e.g. business plans, procedures manuals, employee training manuals, etc.) to gain a better understanding of the functions and processes. This is done in advance of interviews to provide a base-level understanding to reduce the amount of time required of the interviewees.

Two different types of interviews (high-level and detailed business process) are conducted with key personnel from each department. The initial (high-level) interviews are conducted with a representative that will provide an overall high-level view of the department including its mission, responsibilities, and identification of the functional areas. This person will identify those staff that will provide details of the specific processes in each of the functional areas identified. For instance, if the department is Human Resources, functional areas of the department might include: Applicant Processing, Classification, Training, and Personnel File Management. It is expected that this first interview/meeting will last approximately one hour.

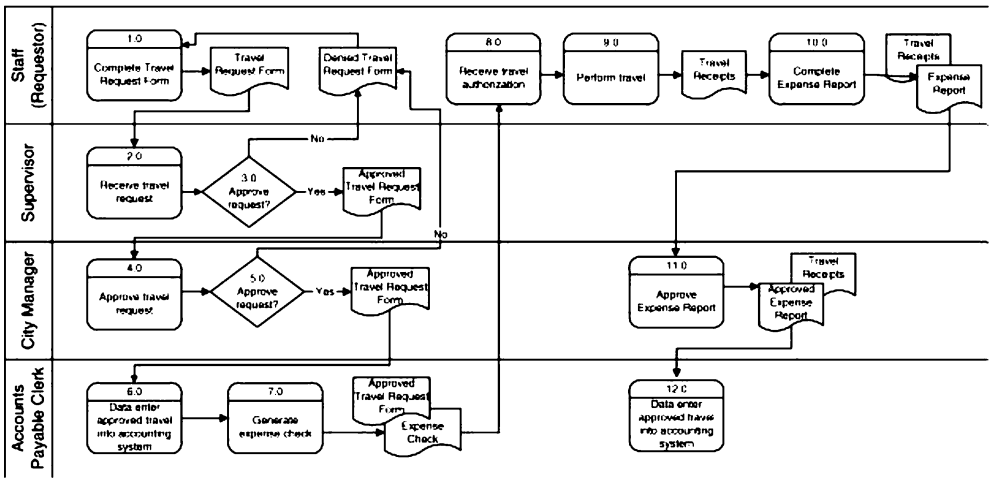


Figure A.10 Business Process Example—Travel Expense Process

Source: Blackburn Consulting.

The second interviews will be detailed interviews that will focus on daily processes performed in each functional area. For example, if the function is Human Resources Classification, the process may be the creation/management of position descriptions. It is only necessary to interview one person that represents a process—there is no need to interview multiple staff performing the same function. These second interviews will likely last one to two hours each, depending on the complexity of the process.

When there are processes that "connect" (e.g. the output from one process is the input to another), it is useful to conduct group interviews with representatives for each process. This often results in "aha" moments when an employee from one process finally understands why they are sending certain records to another process. It also brings to light **business process improvement** (BPI) opportunities. When employees understand the big picture process, they can identify unnecessary process steps and redundant or obsolete documents that can be eliminated.

One purpose of process analysis is to develop taxonomy facets that can be used to surface information for steps in the process. In some cases, process steps can directly inform the types of artifacts that are needed at a part of the process and therefore be used to develop content types in knowledge management use cases. This is related to records management in that knowledge management applications are simply another lens under which content can be viewed. Process analysis can also help determine the scope of metadata for content. For example, when developing an application to view invoices, if the process includes understanding line item detail, this will dictate a different metadata model than if the process sought only to determine whether invoices over a certain threshold were unpaid. Different processes, different use cases, different metadata.[58]

Taxonomy Testing: A Necessary Step

Once a new taxonomy is developed, it must be tested and piloted to see if it meets user needs and expectations. To attempt the rollout of a new taxonomy without testing it first is imprudent and will end up costing more time and resources in the long run. So budget the time and money for it.[59] Taxonomy testing is where the rubber meets the road; it provides real data to see if the taxonomy design has met user expectations and actually helps them in their work.

User testing provides valuable feedback and allows the taxonomist or taxonomy team to fine-tune the work they have done to more closely align the taxonomy with user needs and business objectives. What may have seemed an obvious term or category may, in fact, be *way off*. This may result from the sheer focus and myopia of the taxonomy team. So getting user feedback is essential.

> There's nothing better than getting quantitative feedback to see if you're hitting the mark with users.

There are many taxonomy testing tools that can assist in the design effort. Once an initial design is drafted, a

"low-tech" approach is to handwrite classification categories and document types on post-it notes or index cards. Then bring in a sampling of users and ask them to place the notes or cards in the proper category. The results are tracked and calculated.[60]

Software is available to conduct this card sorting in a more high-tech way, and more sophisticated software to assist in the development and testing effort, and to help to update and maintain the taxonomy.

Regardless of the method used, the taxonomy team or even IG team or task force needs to be the designated arbiter when conflicting opinions arise.

Taxonomy testing is not a one-shot task; with feedback and changes, you progress in iterations closer and closer to meeting user requirements, which may take several rounds of testing and changes.[61]

Taxonomies can be tested in multiple ways. User acceptance throughout the derivation process can be simple conference room pilots or validation, formal usability testing based on use cases, card sorting (open and closed), and tagging processes. Autotagging of content with target taxonomies is also an area that requires testing.[62]

Taxonomy Maintenance

After a taxonomy has been implemented, it will need to be updated over time to reflect changes in document management processes as well to increase usability. Therefore, users should have the opportunity to suggest changes, addition, deletions, and so on. *There should be a formal process in place to manage requests for changes*. A person or committee should be assigned the responsibility to determine how and if each request will be facilitated.

There must be guidelines to follow in making changes to the taxonomy. A US State Agency organization uses the following guidelines in determining taxonomy changes:

- The new term must have a definition, preferably provided by the proposer of the new term.
- It should be a term someone would recognize even if they have no background within our agency's workings; use of industry standard terminology is preferred.
- Terms should be mutually exclusive from other terms.
- Terms that can be derived using a combination of other terms or facilitated with metadata will not be added.
- The value should not be a "temporary" term—it should have some expectation to have a long lifespan.
- We should expect that there would be a significant volume of content that could be assigned the value—otherwise, use of a more general document type and clarification through the metadata on items is preferred: if enough items are titled with the new term over time to warrant reconsideration, it will be reconsidered.
- For higher-level values in the hierarchy, the relationship between parents and children (functions and activities) is always "is a kind of ..." Other relationships are not supported.
- Document type values should not reflect the underlying technology used to capture the content and should not reflect the format of the content directly.

> There should be a formal process in place to manage requests for taxonomy changes.

Social Tagging and Folksonomies

Social tagging is a method that allows users to manage content with metadata they apply themselves using keywords or metadata tags. Unlike traditional classification, which uses a controlled vocabulary, *social tagging keywords are freely chosen by each individual*.

Folksonomy is the term used for this free-form, social approach to metadata assignment.

Folksonomies are not an ordered classification system; rather, they are a list of keywords input by users that are ranked by popularity.[63]

A folksonomy uses free-form words to classify documents. A folksonomy approach is useful for potentially updating your taxonomy structure and improves the user search experience.

Taxonomies and folksonomies both have their place. *Folksonomies can be used in concert with taxonomies to nominate key terms for use in the taxonomy*, which contributes toward the updating and maintenance of the taxonomy while making the user experience better by utilizing their own preferred terms.

A combined taxonomy and folksonomy approach may provide for an optional "free-text metadata field" for social tags that might be titled "Subject" or "Comment." Then users could search that free-form, uncontrolled field to narrow document searches. The folksonomy fields will be of most use to a user or departmental area, but if the terms are used frequently enough, they may need to be added to the formal taxonomy's controlled vocabulary to benefit the entire organization.

In sum, taxonomy development, testing and maintenance is hard work—but it can yield significant and sustained benefits to the organization over the long haul by providing more complete and accurate information when knowledge workers make searches, better IG and control over the organization's documents, records, and information, and a more agile compliance and litigation readiness posture.

APPENDIX SUMMARY: KEY POINTS

- During an average workday, knowledge workers spend 15 to 25% of their timesearching for information, often due to poor taxonomy design.
- Taxonomies are hierarchical classification structures used to standardize the naming and organization of information using controlled vocabularies for terms.
- Taxonomies speed up the process of retrieving records because end-users can select from subject categories or topics.
- Taxonomies need to be considered from two main perspectives: navigation and classification.
- Poor search results, inconsistent or conflicting file plans, and the inability to locate information on a timely basis are indications that taxonomy work is needed.
- Metadata, which are the characteristics of a document expressed in data fields, must be leveraged in taxonomy design.
- Best practices dictate that taxonomy development includes designing the taxonomy structure and heuristic principles to align with user needs.
- There are three basic types of hierarchical taxonomies: subject, business-unit, and functional.
- A *hybrid* approach to taxonomy design is usually the best.
- A subject matter expert (SME) can be a valuable resource in taxonomy development. They should not be relied on too heavily, though, or the taxonomy may end up filled with esoteric jargon.
- A document inventory is conducted to gather detailed information regarding the documents managed.
- Business processes must be documented and analyzed to develop a taxonomy.
- User testing is essential and provides valuable feedback and allows the taxonomist or taxonomy team to fine-tune the work.
- Begin by using low-cost, simple tools for taxonomy development and migrate to more capable ones as your organization's needs grow and maintenance is required.
- A folksonomy uses free-form words to classify documents. A folksonomy approach is useful for potentially updating your taxonomy structure and improves the user search experience.

Endnotes

1. Cadence Group, "Taxonomies: The Backbone of Enterprise Content Management," August 18, 2006, www.cadence-group.com/articles/taxonomy/backbone.htm.

2. Delphi Group, "Taxonomy and Content Classification: Market Milestone Report," 2002, https://whitepapers.us.com/taxonomy-content-classification-market-milestone-report-white-paper-uga-edu.html (accessed September 14, 2018).

3. Ibid.

4. Cadence Group, "Taxonomies: The Backbone of Enterprise Content Management."

5. Daniela Barbosa, "The Taxonomy Folksonomy Cookbook, www.slideshare.net/HeuvelMarketing/taxonomy-folksonomy-cookbook (accessed September 14, 2018).

6. Ibid.

7. Montague Institute Review, "Your Taxonomy Is Your Future," February 2000, http://www.montague.com/review/articles/future.pdf.

8. The Free Library, "Creating Order out of Chaos with Taxonomies," 2005, www.thefreelibrary.com/Creating+order+out+of+chaos+with+taxonomies%3A+the+increasing+volume+of...-a0132679071 (accessed September 14, 2018).

9. Susan Cisco and Wanda Jackson, *Information Management Journal,* "Creating Order out of Chaos with Taxonomies" May/June 2005, www.arma.org/bookstore/files/Cisco.pdf.

10. Marcia Morante, "Usability Guidelines for Taxonomy Development," April 2003, www.montague.com/abstracts/usability.html.

11. Seth Earley, e-mail to author, September 10, 2012.

12. Ibid.

13. Cadence Group, "Taxonomies," 3.

14. Dam Coalition, "8 Things You Need to Know about How Taxonomy Can Improve Search," May 16, 2010, www.tech-speed.co.uk/dam/2010/05/17/8-things-you-need-to-know-about-how-taxonomy-can-improve-search.html (accessed September 14, 2018).

15. Ibid.

16. Seth Earley, e-mail to author, September 10, 2012.

17. National Archives of Australia, "AGLS Metadata Standard, Part 2—Usage Guide," Version 2.0, July 2010, www.agls.gov.au/.

18. Kate Cumming, "Metadata Matters," in *Managing Electronic Records,* eds. Julie McLeod and Catherine Hare (London: Facet Publishing, 2005), 34.

19. Minnesota State Archives, "Electronic Records Management Guidelines," March 12, 2012, www.mnhs.org/preserve/records/electronicrecords/ermetadata.html.

20. Ibid.

21. Kate Cumming, "Metadata Matters," 35.

22. Ibid.

23. "Understanding Metadata," NISO, https://groups.niso.org/apps/group_public/download.php/17443/understanding-metadata (accessed September 14, 2018).

24. Minnesota State Archives, "Electronic Records Management Guidelines."

25. Ibid.

26. Ibid.

27. The National Archives, "Requirements for Electronic Records Management Systems," 2002, http://webarchive.nationalarchives.gov.uk/+/http://www.nationalarchives.gov.uk/documents/metadatafinal.pdf (accessed September 21, 2018).

28. "ISO 23081-1:2006, Information and Documentation—Records Management Processes—Metadata for Records—Part 1: Principles," www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40832 (accessed September 21, 2018).

29. Carl Weise, "ISO 23081-1: 2006, Metadata for Records, Part 1: Principles," January 27, 2012, www.aiim.org/community/blogs/expert/ISO-23081-1-2006-Metadata-for-records-Part-1-principles.

30. Dublin Core Metadata Initiative, http://dublincore.org/metadata-basics/ (accessed September 21, 2018).

31. Diane Hillman, Dublin Core Metadata Initiative, User Guide, November 7, 2005, http://dublincore.org/documents/usageguide/.

32. Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set," Version 1.1, June 14, 2012, http://dublincore.org/documents/dces/.

33. Library of Congress, International Standard Maintenance Agency, www.loc.gov/z3950/agency/ (accessed September 14, 2018).

34. National Information Standards Organization (NISO), "ANSI/NISO Z39.50 2003 (R2009) Information Retrieval: Application Service Definition & Protocol Specification," https://www.niso.org/publications/ansiniso-z3950-2003-s2014-information-retrieval-application-service-definition (accessed September 24, 2018).

35. Jenn Riley, "Glossary of Metadata Standards," 2009–2010, http://jennriley.com/metadatamap/seeingstandards_glossary_pamphlet.pdf (accessed September 14, 2018).

36. Global Information Locator Service (GILS), "Initiatives," www.gils.net/initiatives.html (accessed September 14, 2018).

37. Ibid.

38. Adventures in Records Management, "The Business Classification Scheme," October 15, 2006, http://adventuresinrecordsmanagement.blogspot.com/2006/10/business-classification-scheme.html.

39. Seth Earley, e-mail to author, September 10, 2012.

40. National Archives of Australia, www.naa.gov.au (accessed September 14, 2018).

41. Adventures in Records Management, "The Business Classification Scheme."

42. Ibid.

43. Cisco and Jackson, "Creating Order out of Chaos with Taxonomies."

44. [4]Ibid.

45. Seth Earley, e-mail to author, September 10, 2012.

46. Hedden, "The Accidental Taxonomist," 10.

47. Seth Earley, e-mail to author, September 10, 2012.

48. Hedden, "The Accidental Taxonomist," 8.

49. NISO, Project ISO 25964, www.niso.org/workrooms/iso25964 (accessed September 14, 2018).

50. This section is reprinted with permission from Barb Blackburn, "Taxonomy Design Types," www.imergeconsult.com/img/114BB.pdf (accessed October 12, 2012); *e-Doc Magazine*, AIIM International (May/June 2006), 14 and 16.

51. Ibid.

52. Ibid.

53. Ibid.

54. Ibid.

55. Ibid.

56. Ibid.

57. Seth Earley, e-mail to author, September 10, 2012.

58. Ibid.

59. Stephanie Lemieux, "The Pain and Gain of Taxonomy User Testing," July 8, 2008, https://sethearley.wordpress.com/2008/07/08/the-pain-and-gain-of-taxonomy-user-testing.

60. Ibid.

61. Ibid.

62. Seth Earley, e-mail to author, September 10, 2012.

63. Tom Reamy, "Folksonomy Folktales," *KM World*, September 29, 2009, www.kmworld.com/Articles/Editorial/Feature/Folksonomy-folktales-56210.aspx.

APPENDIX B
Laws and Major Regulations Related to Records Management

United States

Records management practices and standards are delineated in many federal regulations. Also, there are a number of state statutes that have passed and in some cases they actually supersede federal regulations; therefore it is crucial to understand compliance within the state or states where an organization operates.

On the federal level, public companies must be vigilant in verifying, protecting, and reporting financial information to comply with requirements under Sarbanes-Oxley and the Gramm-Leach-Bliley Act (GLBA). Healthcare concerns must meet the requirements of HIPAA, and investment firms must comply with a myriad of regulations by the Securities and Exchange Commission (SEC) and National Association of Securities Dealers (NASD).

Following is a brief description of current rules, laws, regulators, and their records retention and corporate policy requirements. *(Note: This is an overview, and firms should consult their own legal counsel for interpretation and applicability.)*

Gramm-Leach-Bliley Act

The Financial Institution Privacy Protection Act of 2001 and Financial Institution Privacy Protection Act of 2003 (Gramm-Leach-Bliley Act) was amended in 2003 to improve and increase protection of nonpublic personal information. Through this Act, financial records be properly secured, safeguarded, and eventually completely destroyed so that the information cannot be further accessed.

Healthcare Insurance Portability and Accountability Act of 1996 (HIPAA)

HIPAA requires that security standards be adopted for: (1) controlling who may access health information; (2) providing audit trails for electronic record systems; (3) isolating health data, making it inaccessible to unauthorized access; (4) ensuring the confidentiality and safeguarding of health information when it is electronically transmitted to ensure it is physically, electronically, and administratively secure; and (5) meeting the needs and capabilities of small and rural healthcare providers.

PATRIOT Act (Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001)

The PATRIOT Act: (1) requires that the identity of a person opening an account with any financial institution is verified by the financial institution, and they must implement reasonable procedures to maintain identity information; and (2) provides law enforcement organizations broad investigatory rights, including warrantless searches.

Sarbanes-Oxley Act (SOX)

The key provisions of SOX require that: (1) public corporations implement extensive policies, procedures, and tools to prevent fraudulent activities; (2) financial control and risk mitigation processes be documented and verified by independent auditors; (3) executives of publicly traded companies certify the validity of the company's financial statements; and (4) business records must be kept for not less than five years.

SEC Rule 17A-4

SEC Rule 17A-4 requires that: (1) records that must be maintained and preserved and be available to be produced or reproduced using either micrographic media (such as microfilm or microfiche) or electronic storage media (any digital storage medium or system); and (2) original copies of all communications, such as interoffice memoranda, be preserved for no less than *three* years, the first two in an easily accessible location.

CFR Title 47, Part 42—Telecommunications

CFR Title 47, Part 42 requires that telecommunications carriers keep original records or reproductions of original records, including memoranda, documents, papers, and correspondence that the carrier prepared or that were prepared on behalf of the carrier.

CFR Title 21, Part 11—Pharmaceuticals

CFR Title 21, Part 11 requires: (1) controls are in place to protect content stored on both open and closed systems to ensure the authenticity and integrity of electronic records; and (2) generating accurate and complete electronic copies of records so that the Food and Drug Administration (FDA) may inspect them.

US Federal Authority on Archives and Records: National Archives and Records Administration (NARA)

The National Archives and Records Administration (nara.gov):

- Oversees physical and electronic recordkeeping policies and procedures of government agencies, requiring adequate and proper documentation on the conduction of US government business;
- Defines formal e-records as machine-readable materials created or received by an agency of the US federal government under federal law or in the course of the transaction of public business;
- Requires that organized records series be established for electronic records on a particular subject or function to facilitate the management of these e-records.

NARA regulations affecting Federal agencies and their records management programs are found in Subchapter B of 36 Code of Federal Regulations Chapter XII.[1,2]

- Part 1220—Federal Records; General
- Part 1222—Creation and Maintenance of Records
- Part 1223—Managing Vital Records
- Part 1224—Records Disposition Program
- Part 1225—Scheduling Records
- Part 1226—Implementing Disposition
- Part 1227—General Records Schedule
- Part 1228—Loan of Permanent and Unscheduled Records
- Part 1229—Emergency Authorization to Destroy Records
- Part 1230—Unlawful or Accidental Removal, Defacing, Alteration, or Destruction of Records
- Part 1231—Transfer of Records from the Custody of One Executive Agency to Another
- Part 1232—Transfer of Records to Records Storage Facilities
- Part 1233—Transfer, Use, and Disposition of Records in a NARA Federal Records Center
- Part 1234—Facility Standards for Records Storage Facilities
- Part 1235—Transfer of Records to the National Archives of the United States
- Part 1236—Electronic Records Management
- Part 1237—Audiovisual, Cartographic, and Related Records Management
- Part 1238—Microform Records Management
- Part 1239—Program Assistance and Inspections

- Part 1240–1249—[Reserved]

US Code of Federal Regulations

In the Code of Federal Regulations there are over 5,000 references to retaining records. The Code can be found online at: www.ecfr.gov.

Canada*

The National Standards of Canada for electronic records management are: (1) *Electronic Records as Documentary Evidence* CAN/CGSB-72.34–2005 ("72.34"), published in December 2005; and, (2) *Microfilm and Electronic Images as Documentary Evidence* CAN/CGSB-72.11–93, first published in 1979 and updated to 2000 ("72.11").[3] 72.34 incorporates all that 72.11 deals with and is therefore the more important of the two. Because of its age, 72.11 should not be relied upon for its "legal" content. However, 72.11 has remained the industry standard for "imaging" procedures—converting original paper records to electronic storage. The Canada Revenue Agency has adopted these standards as applicable to records concerning taxation.[4]

72.34 deals with these topics: (1) management authorization and accountability; (2) documentation of procedures used to manage records; (3) "reliability testing" of electronic records according to existing legal rules; (4) the procedures manual and the chief records officer; (5) readiness to produce (the "prime directive"); (6) records recorded and stored in accordance with "the usual and ordinary course of business" and "system integrity," being key phrases from the Evidence Acts in Canada; (7) retention and disposal of electronic records; (8) backup and records system recovery; and (9) security and protection. From these standards practitioners have derived many specific tests for auditing, establishing, and revising electronic records management systems.[5]

The "prime directive" of these standards states: "An organization shall always be prepared to produce its records as evidence."[6] *The duty to establish the "prime directive" falls upon senior management*:[7]

5.4.3 Senior management, the organization's own internal law-making authority, proclaims throughout the organization the integrity of the organization's records system (and, therefore, the integrity of its electronic records) by establishing and declaring:

a. The system's role in the usual and ordinary course of business.
b. The circumstances under which its records are made.
c. Its prime directive for all RMS [records management system] purposes, i.e. an organization shall always be prepared to produce its records as evidence. This dominant principle applies to all of the organization's business records, including electronic, optical, original paper source records, microfilm, and

other records of equivalent form and content.

Being the "dominant principle" of an organization's electronic records management system, the duty to maintain compliance with the "prime directive" should fall upon its senior management.

Because an electronic record is completely dependent upon its ERM system for everything, compliance with these National Standards and their "prime directive" should be part of the determination of the "admissibility" (acceptability) of evidence and of electronic discovery in court proceedings (litigation) and in regulatory tribunal proceedings.[8]

There are 14 legal jurisdictions in Canada: 10 provinces; 3 territories; and the federal jurisdiction of the Government of Canada. Each has an Evidence Act (the Civil Code in the province of Quebec[9]), which applies to legal proceedings within its legislative jurisdiction. For example, criminal law and patents and copyrights are within federal legislative jurisdiction, and most civil litigation comes within provincial legislative jurisdiction.[10]

The admissibility of records as evidence is determined under the "business record" provisions of the Evidence Acts.[11] They require proof that a record was made "in the usual and ordinary course of business," and of "the circumstances of the making of the record." In addition, to obtain admissibility for electronic records, most of the Evidence Acts contain electronic record provisions, which state that an electronic record is admissible as evidence on proof of the "integrity of the electronic record system in which the data was recorded or stored."[12] This is the "system integrity" test for the admissibility of electronic records. The word "integrity" has yet to be defined by the courts.[13]

However, by way of sections such as the following, the electronic record provisions of the Evidence Acts make reference to the use of standards such as the National Standards of Canada:

> For the purpose of determining under any rule of law whether an electronic record is admissible, evidence may be presented in respect of any standard, procedure, usage or practice on how electronic records are to be recorded or stored, having regard to the type of business or endeavor that used, recorded, or stored the electronic record and the nature and purpose of the electronic record.[14]

There are six areas of law and records and information management (RIM) applicable to paper and electronic records:

1. The laws of evidence applicable to electronic and paper records[15]
2. The National Standards of Canada concerning electronic records[16]
3. The records requirements of government agencies, such as the Canada Revenue Agency[17]

4. The electronic commerce legislation[18]
5. The privacy laws[19]
6. The guidelines for electronic discovery in legal proceedings[20]

These six areas are closely interrelated and are based upon very similar concepts. They all make demands of records systems and of the chief records officer or others responsible for records. *Therefore, a failure to satisfy the records management needs of any one of them will likely mean a failure to satisfy all of them*. Agencies that manage these areas of law look to the decisions of the courts to determine the requirements for acceptable records.

Each of these areas of law affects records and information management, just as they are affected by the laws governing the use of records as evidence in legal proceedings—the laws of evidence. These relationships make mandatory compliance with the "prime directive" provided by the national standards, which states: "an organization shall always be prepared to produce its records as evidence."[21]

United Kingdom

Regulations and Legislation Impacting Records Retention

"The following Acts and Statutory Instruments of the UK and Scottish Parliaments contain provisions that are relevant to records retention and disposal:"[22]

Acts of the UK Parliament

1957 c31 Occupiers Liability Act 1957

1969 c57 Employers' Liability (Compulsory Insurance) Act 1969

1970 c41 Equal Pay Act 1970

1970 c9 Taxes Management Act 1970

1973 c52 Prescription and Limitations (Scotland) Act 1973

1974 c37 Health and Safety at Work (etc.) Act 1974

1975 c65 Sex Discrimination Act 1975

1976 c74 Race Relations Act 1976

1980 c58 Limitation Act 1980

1992 c4 Social Security Contributions and Benefits Act 1992

1994 c30 Education Act 1994

1994 c23 Value Added Tax Act 1994

1995 c50 Disability Discrimination Act 1995

1998 c29 Data Protection Act 1998

Acts of the Scottish Parliament

2002 asp13 Freedom of Information (Scotland) Act 2002

Statutory Instruments of the UK Parliament

SI 1977/500 The Safety Representatives and Safety Committees Regulations 1977

SI 1981/917 The Health and Safety (First Aid) Regulations 1981

SI 1982/894 The Statutory Sick Pay (General) Regulations 1982

SI 1986/1960 The Statutory Maternity Pay (General) Regulations 1986

SI 1989/1790 The Noise at Work Regulations 1989

SI 1989/635 The Electricity at Work Regulations 1989

SI 1989/682 The Health and Safety Information for Employees Regulations 1989

SI 1991/2680 The Public Works Contracts Regulations 1991

SI 1992/2792 The Health and Safety (Display Screen Equipment) Regulations 1992

SI 1992/2793 The Manual Handling Operations Regulations 1992

SI 1992/2932 The Provision and Use of Work Equipment Regulations 1992

SI 1992/2966 The Personal Protective Equipment at Work Regulations 1992

SI 1993/3228 The Public Services Contracts Regulations 1993

SI 1993/744 The Income Tax (Employments) Regulations 1993

SI 1995/201 The Public Supply Contracts Regulations 1995

SI 1995/3163 The Reporting of Injuries, Diseases and Dangerous Occurrences Regulations 1995

SI 1996/1513 The Health and Safety (Consultation with Employees) Regulations 1996

SI 1996/341 The Health and Safety (Safety Signs and Signals) Regulations 1996

SI 1996/972 The Special Waste Regulations 1996

SI 1997/1840 The Fire Precautions (Workplace) Regulations 1997

SI 1998/1833 The Working Time Regulations 1998

SI 1998/2306 The Provision and Use of Work Equipment Regulations 1998

SI 1998/2307 The Lifting Operations and Lifting Equipment Regulations 1998

SI 1998/2573 The Employers' Liability (Compulsory Insurance) Regulations 1998

SI 1999/3242 The Management of Health and Safety at Work Regulations 1999

SI 1999/3312 The Maternity and Parental Leave (etc.) Regulations 1999

SI 1999/584 The National Minimum Wage Regulations 1998

SI 2002/2675 The Control of Asbestos at Work Regulations 2002

SI 2002/2676 The Control of Lead at Work Regulations 2002

SI 2002/2677 The Control of Substances Hazardous to Health Regulations 2002

Other Provisions

HMCE 700/21 HM Customs and Excise Notice 700/21: Keeping [VAT] records and accounts

IR CA30 Statutory Sick Pay Manual for Employers CA30

Australia

Archives Act

The Archives Act 1983 empowers the Archives to preserve the archival resources of the Australian Government—those records designated "national archives." Under the Act, it is illegal to destroy Australian Government records without permission from the Archives unless destruction is specified in another piece of legislation or allowed under a normal administrative practice.

The Act also establishes a right of public access to nonexempt Commonwealth records in the "open access period" (transitioning from 30 years to 20 years over the period 2011 to 2021 under amendments to the Act passed in 2010). Different open access periods exist for Cabinet notebooks (transitioning from 50 years to 30 years over the period 2011 to 2021) and records containing Census information (99 years).

Freedom of Information Act

The Freedom of Information Act 1982 gives individuals the legal right to access documents held by Australian Government ministers, departments, and most agencies, including Norfolk Island Government agencies. From November 1, 2010, the FOI Act also applies to documents created or held by contractors or subcontractors who provided services to the public or third parties on behalf of agencies.

The FOI Act applies to records that are not yet in the open access period under the Archives Act unless the document contains personal information (including personal information about a deceased person). The Archives Act regulates access to records in the open access period.

When a member of the public requests information, your agency must identify and preserve all relevant sources, including records, until a final decision on the request is made. The FOI Act also sets out how agencies may correct, annotate, or update records if a member of the public shows that any personal information relating to them is incomplete, incorrect, out of date, or misleading.

The FOI Act also establishes the Information Publication Scheme (IPS), which requires agencies subject to the FOI Act to take a proactive approach to publishing a broad range of information on their website. The IPS does not apply to a small number of security and intelligence agencies that are exempt from the FOI Act.

Australian Information Commissioner Act

The Australian Information Commissioner Act 2010 established the Office of the Australian Information Commissioner. The OAIC has three sets of functions. These are:

1. Freedom of information functions—protecting the public's right of access to documents under the amended *Freedom of Information Act* and reviewing decisions made by agencies and ministers under that Act.
2. Privacy functions—ensuring proper handling of personal information in accordance with the *Privacy Act 1988*.
3. Government and information policy functions, conferred on it by the *Australian Information Commissioner Act 2010*—these include strategic functions relating to information management and ensuring maximum coordination, efficiency and transparency in government information policy and practice.

As part of its government and information policy function, the OAIC is committed to leading the development and implementation of a national information policy framework to promote secure and open government. It aims to achieve this by driving public access to government information and encouraging agencies to proactively publish information.

Privacy Act

The Privacy Act 1988 regulates the handling of personal information by Australian Government agencies, ACT government agencies, ACT government agencies, Norfolk Island Government agencies, and a range of private and not-for-profit organizations. The *Privacy Act* regulates the way in which personal information can be collected, its accuracy, how it is kept secure, and how it is used and disclosed. It also provides rights to individuals to access and correct the information that organizations and government agencies hold about them. Records in the open-access period as defined in the *Archives Act 1983* are not covered by the Privacy Act. The Privacy Act also sets out requirements that may apply when an agency enters into a contract under which services are provided to the agency.

Evidence Act

The Evidence Act 1995 defines what documents, including records, can be used as evidence in a Commonwealth court.[23]

All agencies need to take account of evidence legislation. A court may need to examine records as evidence of an organization's decisions and actions. General advice on the impact of the *Evidence Act* is given in the publication Commonwealth Records in Evidence (pdf, 418kb).

Electronic Transactions Act

The Electronic Transactions Act 1999 encourages online business by ensuring that electronic evidence of transactions is not invalidated because of its format. This Act does not authorize the destruction of any Australian Government records, whether originals or copies. The obligations placed on agencies under the *Archives Act 1983* for the preservation and disposal of Commonwealth records continue to apply.

Financial Management and Accountability Act

The Financial Management and Accountability Act 1997 states that an APS employee who misapplies, improperly disposes of, or improperly uses Commonwealth records may be in breach of the *Financial Management and Accountability Act* (s. 41). Regulation 12 of the Act requires that the terms of approval for a proposal to spend money be recorded in writing as soon as practicable.

Australian Government records fall within the meaning of “public property” as defined in this Act.

Crimes Act

The Crimes Act 1914 outlines crimes against the Commonwealth. Several parts of the Act relate to records. For example, section 70 prohibits public servants (or anyone working for the Australian Government, including contractors and consultants) from publishing or communicating facts, documents, or information that they gain access to through their work unless they have permission to do so. This includes taking or selling records that should be destroyed.

This Act also makes it an offence for someone to intentionally destroy documents that they know may be required as evidence in a judicial proceeding.

Identifying Records Management Requirements in Other Legislation

Your agency [or business] needs to be aware of the legislation governing its own records practices.

Some legislative requirements apply to many agencies [and businesses]. For example, occupational health and safety legislation requires an organization to

keep certain types of records for prescribed periods of time. Requirements that apply to all agencies are included in the National Archives' Administrative Functions Disposal Authority.

Other legislative requirements may apply only to the particular business of one or a number of agencies.

Recordkeeping requirements may be stipulated in your agency's enabling legislation (legislation that established the agency) or in specific legislation that your agency is responsible for administering.[24]

Notes

1. NARA Records Management Guidance and Regulations, www.archives.gov/records-mgmt/policy/guidance-regulations.html (accessed October 17, 2012).

2. NARA Records Management Guidance and Regulations, www.archives.gov/about/regulations/subchapter/b.html (accessed October 17, 2012).

3. These standards were developed by the CGSB (Canadian General Standards Board), which is a standards-writing agency within Public Works and Government Services Canada (a department of the federal government). It is accredited by the Standards Council of Canada as a standards development agency. The Council must certify that standards have been developed by the required procedures before it will designate them as being National Standards of Canada. *72.34* incorporates by reference as "normative references": (1) many of the standards of the International Organization for Standardization (ISO) in Geneva, Switzerland. ("ISO," derived from the Greek word *isos* (equal) so as to provide a common acronym for all languages); and, (2) several of the standards of the Canadian Standards Association (CSA). The "Normative references" section of 72.34 (p. 2) states that these "referenced documents are indispensable for the application of this document." 72.11 cites (p. 2, "Applicable Publications") several standards of the American National Standards Institute/Association for Information and Image Management (ANSI/AIIM) as publications "applicable to this standard." The process by which the National Standards of Canada are created and maintained is described within the standards themselves (reverse side of the front cover), and on the CGSB's website (see "Standards Development"), from which website these standards may be obtained; online: www.ongc-cgsb.gc.ca.

4. The Canada Revenue Agency (CRA) informs the public of its policies and procedures by means, among others, of its *Information Circulars* (ICs), and *GST/HST Memoranda*. (GST: goods and services tax; HST: harmonized sales tax, i.e. the harmonization of federal and provincial sales taxes into one retail sales tax.) In particular, see: *IC05-1*, dated June 2010, entitled, *Electronic Record Keeping*, paragraphs 24, 26 and 28. Note that use of the National

Standard cited in paragraph 26, *Microfilm and Electronic Images as Documentary Evidence* CAN/CGSB-72.11-93 is mandatory for, "Imaging and microfilm (including microfiche) reproductions of books of original entry and source documents...." Paragraph 24 recommends the use of the newer national standard, *Electronic Records as Documentary Evidence* CAN/CGSB-72.34-2005, "To ensure the reliability, integrity and authenticity of electronic records." However, if this newer standard is given the same treatment by CRA as the older standard, it will be made mandatory as well. And similar statements appear in the GST Memoranda, *Computerized Records* 500-1-2, *Books and Records* 500-1. IC05-1. *Electronic Record Keeping*, concludes with the note, "Most Canada Revenue Agency publications are available on the CRA website, www.cra.gc.ca, under the heading 'Forms and Publications.'"

5. There are more than 200 specific compliance tests that can be applied to determine if the principles of 72.34 are being complied with. The analysts—a combined team of records management and legal expertise—analyze: (1) the nature of the business involved; (2) the uses and value of its records for its various functions; (3) the likelihood and risk of the various types of its records being the subject of legal proceedings, or of their being challenged by some regulating authority; and, (4) the consequences of the unavailability of acceptable records—for example, the consequences of its records not being accepted in legal proceedings. Similarly, in regard to the older National Standard of Canada, 72.11, there is a comparable series of more than 50 tests that can be applied to determine the state of compliance with its principles.

6. *Electronic Records as Documentary Evidence* CAN/CGSB-72.34-2005 ("72.34"), clause 5.4.3 c) at p. 17; and, *Microfilm and Electronic Images as Documentary Evidence* CAN/CGSB-72.11-93 ("72.11"), paragraph 4.1.2 at p. 2, *supra* note 49.

7. 72.34, Clause 5.4.3, *ibid.*

8. "Admissibility" refers to the procedure by which a presiding judge determines if a record or other proffered evidence is acceptable as evidence according the rules of evidence. "Electronic discovery" is the compulsory exchange of relevant records by the parties to legal proceedings prior to trial. As to the admissibility of records as evidence see: Ken Chasse, "The Admissibility of Electronic Business Records" (2010), 8 Canadian Journal of Law and Technology 105; and, Ken Chasse, "Electronic Records for Evidence and Disclosure and Discovery" (2011) 57 The Criminal Law Quarterly 284. For the electronic discovery of records see: Ken Chasse, "Electronic Discovery—*Sedona Canada* is Inadequate on Records Management—Here's *Sedona Canada* in Amended Form," *Canadian Journal of Law and Technology* 9 (2011): 135; and Ken Chasse, "Electronic Discovery in the Criminal Court System" *Canadian Criminal Law Review* 14 (2010): 111. See also note 18 *infra*, and accompanying text.

9. For the province of Quebec, comparable provisions are contained in Articles

2831-2842, 2859-2862, 2869-2874 of Book 7 "Evidence" of the Civil Code of Quebec, S.Q. 1991, c. C-64, to be read in conjunction with, An Act to Establish a Legal Framework for Information Technology, R.S.Q. 2001, c. C-1.1, ss. 2, 5-8, and 68.

10. For the legislative jurisdiction of the federal and provincial governments in Canada, see The Constitution Act, 1867 (U.K.) 30 and 31 Victoria, c. 3, s. 91 (federal), and s. 92 (provincial); at online: www.canlii.org/en/ca/laws/stat/30---31-vict-c-3/latest/30---31-vict-c-3.html.

11. The two provinces of Alberta and Newfoundland and Labrador do not have business record provisions in their Evidence Acts. Therefore "admissibility" would be determined in those jurisdictions by way of the court decisions that define the applicable common law rules; such decisions as, *Ares v. Venner*, [1970] S.C.R. 608, 14 D.L.R. (3d) 4 (S.C.C.), and decisions that have applied it.

12. See for example, the Canada Evidence Act, R.S.C. 1985, c. C-5, ss. 31.1-31.8; Alberta Evidence Act, R.S.A. 2000, c. A-18, ss. 41.1-41.8; (Ontario) Evidence Act, R.S.O. 1990, c. E.23, s. 34.1; and the (Nova Scotia) Evidence Act, R.S.N.S. 1989, c. 154, ss. 23A-23G. The Evidence Acts of the two provinces of British Columbia and Newfoundland and Labrador do not contain electronic record provisions. However, because an electronic record is no better than the quality of the record system in which it is recorded or stored, its "integrity" (reliability, credibility) will have to be determined under the other provincial laws that determine the admissibility of records as evidence.

13. The electronic record provisions have been in the Evidence Acts in Canada since 2000. They have been applied to admit electronic records into evidence, but they have not yet received any detailed analysis by the courts.

14. This is the wording used in, for example, s. 41.6 of the Alberta Evidence Act, s. 34.1(8) of the (Ontario) Evidence Act; and, s. 23F of the (Nova Scotia) Evidence Act, *supra* note 10. Section 31.5 of the Canada Evidence Act, *supra* note 58, uses the same wording, the only significant difference being that the word "document" is used instead of "record." For the province of Quebec, see sections 12 and 68 of, An Act to Establish a Legal Framework for Information Technology, R.S.Q., chapter C-1.1.

15. *Supra* notes 54 to 59 and accompanying texts.

16. *Supra* notes 49 and 52 and accompanying texts.

17. *Supra* note 50 and accompanying text.

18. All 14 jurisdictions of Canada have electronic commerce legislation except for the Northwest Territories. See for example, the Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, Parts 2 and 3; Ontario's Electronic Commerce Act, 2000, S.O. 2000, c. 17; and, British Columbia's Electronic Transactions Act, R.B.C. 20001, c. 10. The concept of "system

integrity" in the Evidence Acts (*supra* note 58 and accompanying text), is also found in the electronic commerce legislation. See for example, s. 8 of the Ontario Electronic Commerce Act, 2000, under the heading, "Legal Requirement re Original Documents."

19. For example, Part 1, "Personal Information Protection," of the federal Personal Information Protection and Electronic Documents Act (PIPEDA), S.C. 2000, c. 5, which applies within provincial legislative jurisdiction as well as federal, until a province enacts its own personal information protection Act (a PIPA"), which displaces it in the provincial sphere. British Columbia, Alberta, and Quebec are the only provinces that have done so.

20. The dominant guideline for electronic discovery in Canada is, *The Sedona Canada Principles—Addressing Electronic Discovery*; online: The Sedona Conference, Canada, January 2008: www.thesedonaconference.com/content/miscFiles/canada_pincpls_FINAL_108.pdf or www.thesedonaconference.org/dltForm?did=canada_pincpls_FINAL_108.pdf and, E-Discovery Canada website, hosted by LexUM (at the University of Montreal), online: www.lexum.umontreal.ca/e-discovery. And see also the law journal articles concerning electronic discovery cited in note 54 *supra*.

21. *Supra* notes 52 and 53 and accompanying texts.

22. "Information Governance Record Retention Guidance, www.rec-man.stir.ac.uk/rec-ret/legislation.php (accessed October 17, 2012).

23. www.comlaw.gov.au/Details/C2012C00518(accessed November 30, 2012).

24. National Archives of Australia, www.naa.gov.au/records-management/strategic-information/standards/recordslegislation.aspx (accessed October 17, 2012).

* This section was contributed by Ken Chasse J.D., LL.M., member of the Law Society of Upper Canada (Ontario) and of the Law Society of British Columbia, Canada.

APPENDIX C
Laws and Major Regulations Related to Privacy

United States

Note: This list is representative and not to be considered an exhaustive listing. State laws (such as the California Consumer Privacy Act) and industry regulations may apply to your organization. Consult your legal counsel for definitive research.

Americans with Disabilities Act (ADA)[1]

Cable Communications Policy Act of 1984 (Cable Act)

California Consumer Privacy Act (AB-375)

California Senate Bill 1386 (SB 1386)

Children's Internet Protection Act of 2001 (CIPA)

Children's Online Privacy Protection Act of 1998 (COPPA)

Communications Assistance for Law Enforcement Act of 1994 (CALEA): official CALEA Web site

Computer Fraud and Abuse Act of 1986 (CFAA)

Computer Security Act of 1987: superseded by the Federal Information Security Management Act (FISMA)

Consumer Credit Reporting Reform Act of 1996 (CCRRA): modifies the Fair Credit Reporting Act (FCRA)

Controlling the Assault of Non-Solicited Pornography and Marketing (CAN-SPAM) Act of 2003

Driver's Privacy Protection Act of 1994

Electronic Communications Privacy Act of 1986 (ECPA)

Electronic Freedom of Information Act of 1996 (E-FOIA)

Electronic Funds Transfer Act (EFTA)

Fair and Accurate Credit Transactions Act (FACTA) of 2003

Fair Credit Reporting Act of 1999 (FCRA)

Family Education Rights and Privacy Act of 1974 (FERPA; aka the Buckley Amendment)

Federal Information Security Management Act (FISMA)

Federal Trade Commission Act (FTCA)

Gramm-Leach-Bliley Financial Services Modernization Act of 1999 (GLBA)

Privacy Act of 1974: including U.S. Department of Justice Overview

Privacy Protection Act of 1980 (PPA)

Right to Financial Privacy Act of 1978 (RFPA)

Telecommunications Act of 1996

Telephone Consumer Protection Act of 1991 (TCPA)

Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001 (USA PATRIOT Act)

Video Privacy Protection Act of 1988: discussion and overview

European Union General Data Protection Regulation (GDPR)

The European Union General Data Protection Regulation (GDPR) went into effect May 25, 2018. It is the most significant change in data privacy not only affecting the EU, but it has global implications. It will fundamentally change the personal data is governed across industries.[2]

From the EUGDPR.org website:

The aim of the GDPR is to protect all EU citizens from privacy and data breaches in today's data-driven world. Although the key principles of data privacy still hold true to the previous directive, many changes have been proposed to the regulatory policies; the key points of the GDPR as well as information on the impacts it will have on business can be found below.

Increased Territorial Scope (Extraterritorial Applicability)

Arguably the biggest change to the regulatory landscape of data privacy comes with the extended jurisdiction of the GDPR, as it applies to all companies processing the personal data of data subjects residing in the Union, regardless of the company's location. Previously, territorial applicability of the directive was ambiguous and referred to data process 'in context of an establishment.' This topic has arisen in a number of high-profile court cases. GDPR makes its applicability very clear—it applies to the processing of personal data by controllers and processors in the EU, regardless of whether the processing takes place in the EU or not. The GDPR also applies to the processing of personal data of data subjects in the EU by a controller or processor not established in the EU, where the activities relate to: offering goods or services to EU citizens (irrespective of whether payment is required) and the monitoring of behavior that takes place within the EU. Non-EU businesses processing the data of EU citizens also have to appoint a representative in the EU.

Penalties

Organizations in breach of GDPR can be fined up to 4% of annual global turnover (revenue) or €20 million (whichever is greater). This is the maximum fine that can

be imposed for the most serious infringements, for example, not having sufficient customer consent to process data or violating the core of Privacy by Design concepts. There is a tiered approach to fines, for example, a company can be fined 2% for not having their records in order (article 28), not notifying the supervising authority and data subject about a breach or not conducting impact assessment. It is important to note that these rules apply to both controllers and processors—meaning 'clouds' are not exempt from GDPR enforcement.

Consent

The conditions for consent have been strengthened, and companies are no longer able to use long illegible terms and conditions full of legalese. The request for consent must be given in an intelligible and easily accessible form, with the purpose for data processing attached to that consent. Consent must be clear and distinguishable from other matters and provided in an intelligible and easily accessible form, using clear and plain language. It must be as easy to withdraw consent as it is to give it.

Data Subject Rights

Breach Notification

Under the GDPR, breach notifications are now mandatory in all member states where a data breach is likely to "result in a risk for the rights and freedoms of individuals." This must be done within 72 hours of first having become aware of the breach. Data processors are also required to notify their customers, the controllers, "without undue delay" after first becoming aware of a data breach.

Right to Access

Part of the expanded rights of data subjects outlined by the GDPR is the right for data subjects to obtain confirmation from the data controller as to whether or not personal data concerning them is being processed, where and for what purpose. Further, the controller shall provide a copy of the personal data, free of charge, in an electronic format. This change is a dramatic shift to data transparency and empowerment of data subjects.

Right to Be Forgotten

Also known as Data Erasure, the right to be forgotten entitles the data subject to have the data controller erase his/her personal data, cease further dissemination of the data, and potentially have third parties halt processing of the data. The conditions for erasure, as outlined in article 17, include the data no longer being relevant to original purposes for processing, or a data subject withdrawing consent. It should also be noted that this right requires controllers to compare the subjects' rights to "the public interest in the availability of the data" when considering such requests.

Data Portability

GDPR introduces data portability—the right for a data subject to receive the personal data concerning them—which they have previously provided in a 'commonly used and machine readable format' and have the right to transmit that data to another controller.

Privacy by Design

Privacy by design as a concept has existed for years, but it is only just becoming part of a legal requirement with the GDPR. At its core, privacy by design calls for the inclusion of data protection from the onset of the designing of systems, rather than an addition. More specifically, 'The controller shall ... implement appropriate technical and organisational measures ... in an effective way... in order to meet the requirements of this Regulation and protect the rights of data subjects.' Article 23 calls for controllers to hold and process only the data absolutely necessary for the completion of its duties (data minimization), as well as limiting the access to personal data to those needing to act out the processing.

Data Protection Officers

Under GDPR it is not necessary to submit notifications/registrations to each local DPA of data processing activities, nor is it a requirement to notify/obtain approval for transfers based on the Model Contract Clauses (MCCs). Instead, there are internal record keeping requirements, as further explained below, and DPO appointment is mandatory only for those controllers and processors whose core activities consist of processing operations which require regular and systematic monitoring of data subjects on a large scale or of special categories of data or data relating to criminal convictions and offences. Importantly, the Data Protection Officer:

- Must be appointed on the basis of professional qualities and, in particular, expert knowledge on data protection law and practices.
- May be a staff member or an external service provider.
- Contact details must be provided to the relevant DPA.
- Must be provided with appropriate resources to carry out their tasks and maintain their expert knowledge.
- Must report directly to the highest level of management.
- Must not carry out any other tasks that could result in a conflict of interest.

Major Privacy Laws Worldwide, by Country

Note: Privacy laws are in a state of rapid change and updating. This list is representative and not to be considered an exhaustive listing. State or provincial laws and industry regulations may apply to your organization. Consult your legal counsel for definitive research.

Argentina. Personal Data Protection Act of 2000

Australia. Privacy Act of 1988; Privacy Amendment (Notifiable Data Breaches) Act 2017

Austria. Data Protection Act 2000, Austrian Federal Law Gazette part I No. 165/1999 (Datenschutzgesetz 2000 or DSG 2000)

Belgium. Belgium Data Protection Law and Belgian Data Privacy Commission Privacy Blog

Brazil. 2018 Data Protection Bill of Law; Privacy currently governed by Article 5 of the 1988 Constitution

Bulgaria. Bulgarian Personal Data Protection Act

Canada. Privacy Act—July 1983; Personal Information Protection and Electronic Data Act (PIPEDA) of 2000 (Bill C-6)

Chile. Act on the Protection of Personal Data, August 1998

Colombia. Law 1266 of 2008: (in Spanish) and Law 1273 of 2009 (in Spanish)

Czech Republic. Act on Protection of Personal Data (April 2000) No. 101

Denmark. Act on Processing of Personal Data, Act No. 429, May 2000

Estonia. Personal Data Protection Act of 2003. June 1996, Consolidated July 2002

European Union. European Union Data Protection Directive of 1998; EU Internet Privacy Law of 2002 (Directive 2002/58/EC) *Finland*. Act on the Amendment of the Personal Data Act (986) 2000

France. Data Protection Act of 1978 (revised in 2004)

Germany. Federal Data Protection Act of 2001

Greece. Law No. 2472 on the Protection of Individuals with Regard to the Processing of Personal Data, April 1997

Guernsey. Data Protection (Bailiwick of Guernsey) Law of 2001

Hong Kong. Personal Data Ordinance (the Ordinance)

Hungary. Act LXIII of 1992 on the Protection of Personal Data and the Publicity of Data of Public Interests

Iceland. Act of Protection of Individual; Processing Personal Data, January 2000

Ireland. Data Protection (Amendment) Act, Number 6, of 2003

India. Information Technology Act of 2000; 2018 Data Protection Bill of Law

Italy. Processing of Personal Data Act, January 1997; Data Protection Code of 2003

Japan. Personal Information Protection Law (Act) Law for the Protection of Computer Processed Data Held by Administrative Organs, December 1988

Korea. Act on Personal Information Protection of Public Agencies Act on Information and Communication Network Usage

Latvia. Personal Data Protection Law, March 2000

Lithuania. Law on Legal Protection of Personal Data, June 1996

Luxembourg. Law of August 2002 on the Protection of Persons with Regard to the Processing of Personal Data

Malaysia. Common Law Principle of Confidentiality Personal Data Protection Bill Banking and Financial Institutions Act of 1989 Privacy Provisions

Malta. Data Protection Act (Act XXVI of 2001), Amended March 22, 2002, November 15, 2002 and July 15, 2003

Mexico. Federal Law for the Protection of Personal Data Possessed by Private Persons (Spanish)

Morocco. Data Protection Act

Netherlands. Dutch Personal Data Protection Act 2000 as amended by Acts dated April 5, 2001, Bulletin of Acts, Orders and Decrees 180, December 6, 2001

New Zealand. Privacy Act, May 1993; Privacy Amendment Act, 1993; Privacy Amendment Act, 1994

Norway. Personal Data Act (April 2000)–Act of April 14, 2000 No. 31 Relating to the Processing of Personal Data (Personal Data Act)

Philippines. Data Privacy Act of 2011 (There is also a recognized right of privacy in civil law and a model data protection code.)

Romania. Law No. 677/2001 for the Protection of Persons Concerning the Processing of Personal Data and the Free Circulation of Such Data

Poland. Act of the Protection of Personal Data (August 1997)

Portugal. Act on the Protection of Personal Data (Law 67/98 of 26 October)

Singapore. E-commerce Code for the Protection of Personal Information and Communications of Consumers of Internet Commerce

Slovak Republic. Act No. 428 of July 3, 2002, on Personal Data Protection

Slovenia. Personal Data Protection Act, RS No. 55/99

South Africa. Electronic Communications and Transactions Act, 2002

South Korea. Act on Promotion of Information and Communications Network Utilization and Data Protection of 2000

Spain. Organic Law 15/1999 of December 13 on the Protection of Personal Data

Switzerland. Federal Law on Data Protection of 1992

Sweden. Personal Data Protection Act (1998: 204), October 24, 1998

Taiwan. Computer Processed Personal Data Protection Law (public institution applicability only)

Thailand. Official Information Act, B.E. 2540 (1997) (for state agencies)

United Kingdom. UK Data Protection Act 1998; Privacy and Electronic Communications (EC Directive) Regulations 2003

Vietnam. Law on Electronic Transactions 2008

Notes

1. Information Shield, "United States Privacy Laws," www.informationshield.com/usprivacylaws.html (accessed October 18, 2013).
2. "GDPR Key Changes" https://eugdpr.org/the-regulation/gdpr-faqs/ (accessed December 24, 2018).

GLOSSARY

access control list
In systems—such as ERM, EDRMS, or document management systems—a list of individuals authorized to access, view, amend, transfer, or delete documents, records, or files. Access rights are enforced through software controls.

accountability
The assigned responsibility for records management at a senior level to ensure effective governance with the appropriate level of authority.

adverse inference
Generally a legal inference, adverse to the concerned party, made from a party's silence or the absence of requested evidence.

application programming interface (API)
A way of standardizing the connection between two software applications. They are essentially standard hooks that an application uses to connect to another software application.

archival information package (AIP)
One of three types of information packages that can be submitted in the OAIS preservation model.

archive
Storing information and records for long term or permanent preservation. With respect to e-mail, in a compressed and indexed format to reduce storage requirements and allow for rapid, complex searches (this can also done for blogs, social media or other applications). Archiving of real-time applications like email can only be deemed reliable with record integrity if it is performed immediately, in real time.

artificial intelligence
A branch of computer science that aims to create intelligent machines that can perform human-like thinking tasks in a rapid and automated way, programming computers to solve problems using knowledge, reasoning, perception, learning, and planning. AI is increasingly being used in IG programs for finding relevant information in e-discovery, and classifying unstructured information, as well as assisting in compliance tasks.

ARMA International
Association for Records Managers and Administrators, the U.S.-based nonprofit organization for records managers with a network of international chapters.

authentication, authorization, and audit (or accounting) (AAA)
A network management and security framework that controls computer system logons and access to applications that enforces IG policies and audits usage.[1]

autoclassification

Setting predefined indices to classify documents and records and having the process performed automatically by using software, rather than human intervention. A strong trend toward autoclassification is emerging due to the impact of "Big Data" and rapidly increasing volumes of documents and records.

backup

A complete spare copy of data for purposes of disaster recovery. Backups are nonindexed mass storage and cannot substitute for indexed, archived information that can be quickly searched and retrieved (as in archiving).

best practices

Those methods, processes, or procedures that have been proven to be the most effective, based on real-world experience and measured results.

Big Data

It is high volume, variety, and velocity of data that is too large to manage in traditional relational databases.

blockchain

An aggregation of appended records, "blocks," that are kept in sequence as a list of records, cryptographically linked—that is, each block contains a cryptographic hash of the previous block, a timestamp, and transaction data. A blockchain is more secure than previous records databases, and represents an open, distributed ledger that records and verifies transactions efficiently and permanently.

business activities

The tasks performed to accomplish a particular business function. Several activities may be associated with each business function.

business case

A written analysis of the financial, productivity, auditability, and other factors to justify the investment in software and hardware systems, implementation, and training.

business classification scheme

Also referred to as a BCS, the overall structure an organization uses for organizing, searching, retrieving, storing, and managing documents and records in ERM. The BCS must be developed based on the business functions and activities. A file plan is a graphic representation of the BCS, usually a "hierarchical structure consisting of headings and folders to indicate where and when records should be created during the conducting of the business of an office." In other words *the file plan links the records to their business context*.

business driver

Is a key factor that motivates an organization to undertake a project or program to address that business need.

business functions

Basic business units such as accounting, legal, human resources, and purchasing.

business intelligence

The set of techniques and tools for the transformation of raw data into meaningful and useful information for business analysis purposes. Business intelligence has generally focused mainly on structured data held within relational databases.

business processes
A coordinated set of collaborative and transactional work activities carried out to complete work-steps.

business process improvement (BPI)
Analyzing and redesigning business processes to streamline them and gain efficiencies, reduce cycle times, and improve auditability and worker productivity.

business process outsourcing (BPO)
It is the practice of contracting a third party to perform specific business functions. Often BPO engages businesses outside the home country of the primary business, in order to lower costs.

business process management
Is the analysis, refinement and improvement of automated work steps to reduce cycle times, costs, and labor to speed processing and improve its accuracy.

business process management system (BPMS)
A superset of workflow software, and more: BPMS software offers five main capabilities:[2]

1. Puts existing and new application software under the direct control of business managers.
2. Makes it easier to improve existing business processes and create new ones.
3. Enables the automation of processes across the entire organization, and beyond it.
4. Gives managers “real-time” information on the performance of processes.
5. Allows organizations to take full advantage of new computing services.

BYOD
Is a ‘bring your own device’ which is a policy whereby organizations allow employees to bring their own smartphone, tablet, or laptop to use in the workplace.

capture
Capture components are often also called input components. There are several levels and technologies, from simple document scanning and capture to complex information preparation using automatic classification.

case records
Case records are characterized as having a beginning and an end, but are added

to over time. Case records generally have titles that include names, dates, numbers, or places.

change management
Methods and best practices to assist an organization and its employees in implementing changes to business processes, culture, and systems.

classification
Systematic identification and arrangement of business activities and/or records into categories according to logically structured conventions, methods, and procedural rules represented in a classification system. A coding of content items as members of a group for the purposes of cataloging them or associating them with a taxonomy.

CIA
triad is a cyber-security principle or model whereby security policies are formed using the confidentiality, integrity, and accessibility of information as key considerations.

cloud computing
Cloud computing refers to the provision of computational resources on demand via a network. Cloud computing can be compared to the supply of electricity and gas, or the provision of telephone, television, and postal services. All of these services are presented to the users in a simple way that is easy to understand without the users' needing to know how the services are provided. This simplified view is called an abstraction. Similarly, cloud computing offers computer application developers and users an abstract view of services, which simplifies and ignores much of the details and inner workings. A provider's offering of abstracted Internet services is often called The Cloud.

Code of Federal Regulations (CFR)
"The Code of Federal Regulations (CFR)," issued annually, is the codification of the general and permanent rules published in the *Federal Register* by the departments and agencies of the federal government. It is divided into 50 titles that represent broad areas subject to federal regulation. The 50 subject matter titles contain one or more individual volumes, which are updated once each calendar year, on a staggered basis."[3]

cold site
A cold site is simply an empty computer facility or data center that is ready with air-conditioning, raised floors, telecommunication lines, and electric power. Backup hardware and software will have to be purchased and shipped in quickly to resume operations. Arrangements can be made with suppliers for rapid delivery in the event of a disaster.

compliance monitoring
Being regularly apprised and updated on pertinent regulations and laws and examining processes in the organization to ensure compliance with them. In a records management sense, this involves reviewing and inspecting the various facets of a records management program to ensure it is in compliance. Compliance monitoring can be carried out by an internal audit, external

organization, or records management and must be done on a regular basis.

computer memory
Solid state volatile (erasable) storage capability built into central processing units of computers. At times memory size can be increased by expanding it to the computer's hard drive or external magnetic disks.

consensus mechanism
The way a group makes a decision jointly. Often referred to in blockchain applications like Bitcoin, where users need to constantly update their history of transactions in order to reflect new transactions and wallet balances.

content
In records, the actual information contained in the record; more broadly, content is information, for example, content is managed by ECM systems, and may be email, e-documents, Web content, report content, and so on.

content services
A newer definition of cloud-based content management services offered by Gartner. A content services platform is a set of services and microservices, embodied either as an integrated product suite or as separate applications, that share common APIs and repositories, to exploit diverse content types and to serve multiple constituencies and numerous use cases across an organization.

controlled vocabulary
Set, defined terms used in a taxonomy.

corporate compliance
The set of activities and processes that result in meeting and adhering to all regulations and laws that apply to an organization.

dark data
Unknown data; data that has accumulated but is not used to derive insights or for decision making.

data analytics
The process and techniques for the exploration and analysis of business data to discover and identify new and meaningful information and trends that allow for analysis to take place.

data cleansing (or data scrubbing)
The process of removing corrupt, redundant, and inaccurate data in the data governance process.

data governance
A collection of practices and processes that help to ensure the formal management of data assets within an organization. Data governance often includes other concepts such as data stewardship, data quality, and others to help an enterprise gain better control over its data assets, including methods, technologies, and behaviors around the proper management of data. It also deals with security and privacy, integrity, usability, integration, compliance, availability, roles and responsibilities, and overall management of the internal and external data flows within an organization.

data governance framework
A logical structure for classifying, organizing, and communicating complex activities involved in making decisions about and taking action on enterprise data. The framework or system sets the guidelines and rules of engagement for business and management activities, especially those that deal with or result in the creation and manipulation of data.

data integrity
The overall completeness, accuracy, consistency, and trustworthiness of data.

Data stewardship
It is where individuals are assigned responsibility for the accuracy, integrity, and accessibility of data in the management and oversight of an organization's data assets.

data loss prevention (DLP)
Data loss prevention (DLP; also known as data *leak* prevention) is a computer security term referring to systems that identify, monitor, and protect data in use (e.g. endpoint actions), data in motion (e.g. network actions), and data at rest (e.g. data storage) through deep content inspection, contextual security analysis of transaction (attributes of originator, data object, medium, timing, recipient/destination, and so on) and with a centralized management framework. Systems are designed to detect and prevent unauthorized use and transmission of confidential information.

decentralized architecture
Characterized by no single point of failure. All system information is instantly and constantly replicated to all nodes in a network.

declaration
Assignment of metadata elements to associate the attributes of one or more record folder(s) to a record, or for categories to be managed at the record level, providing the capability to associate a record category to a specific record.

de-duplication
The process of identifying and eliminating redundant occurrences of data.

defensible deletion
Managing the life cycle of information in a standardized and routine way whereby information that has met its life cycle retention requirement is fully and completely deleted. Defensible deletion is the practice of methodically deleting electronically stored information (ESI) when it is no longer useful. This reduces the storage costs and legal risks.

descriptive analytics
A reactive type of data analytics that provides real-time analysis of incoming data.

Designing and Implementation of Recordkeeping Systems (DIRKS)
An Australian framework or methodology consisting of eight steps developed by the Archives Authority of New South Wales, included in ISO 15489, the international standard for records management. Roughly analogous to the

Generally Accepted Recordkeeping Principles® developed by ARMA in the United States.

destruction
The process of eliminating or deleting records, beyond any possible reconstruction.

destruction certificate
Issued once the destruction of a record is complete, which verifies it has taken place, who authorized the destruction, and who carried it out. May include some metadata about the record.

destructive retention policy
Permanently destroying documents or e-documents (such as e-mail) after retaining them for a specified period of time.

diagnostic analytics
A reactive type of data analytics that provides insights into past performance.

digital preservation
See *long-term digital preservation*.

disaster recovery (DR)/business continuity (BC)
The planning, preparation, and testing set of activities used to help a business plan recover from any major business interruption, and to resume normal business operations.

discovery
May refer to the process of gathering and exchanging evidence in civil trials; or, to discover information flows inside an organization using data loss prevention (DLP) tools.

dissemination information package (DIP)
One of three types of information packages that can be submitted in the OAIS preservation model.

disposition
The range of processes associated with implementing records retention, destruction, or transfer decisions, which are documented in disposition authorities or other instruments.

distributed ledger
A decentralized database that provides a consensus of replicated, shared, and synchronized digital transactions geographically spread across many remote computers.

document
Recorded information or object that can be treated as a unit.

document analytics
Detailed usage statistics on e-documents, such as time spent viewing, which pages were viewed and for how long, number of documents printed, where printed, number of copies printed, and other granular information about how and where a document is accessed, viewed, edited, or printed.

Document labeling
It involves adding a tag or label to easily identify a document type or class, such as vital records or confidential documents.

document imaging
Scanning and digitally capturing images of paper documents.

document life cycle
The span of a document's use, from creation, through active use, storage, and final disposition, which may be destruction or preservation.

document life cycle security (DLS)
Providing a secure and controlled environment for e-documents. This can be accomplished by properly implementing technologies including information rights management (IRM) and data loss prevention (DLP), along with complementary technologies like digital signatures.

document management
Managing documents throughout their life cycle from creation to final disposition, including managing revisions. Also called document lifecycle management.

document type
A term used by many software systems to refer to a grouping of related records.

early case assessment
The process of attempting to quickly surface key electronically stored information (ESI), paper documents, and other potential evidence early on in a legal matter. The data gathered during early case assessment is then used to help estimate risk and guide case strategy, such as decisions to go to trial or settle.

e-discovery
Discovery in civil litigation or government investigations that deals with the exchange of information in electronic format (often referred to as electronically stored information or ESI). These data are subject to local rules and agreed-upon processes, and are often reviewed for privilege and relevance before being turned over to opposing counsel.

e-document
An electronic document, that is, a document in digital form.

Electronic Code of Federal Regulations (e-CFR)
"It is not an official legal edition of the CFR. The e-CFR is an editorial compilation of CFR material and Federal Register amendments produced by the National Archives and Records Administration's Office of the Federal Register (OFR) and the Government Printing Office."[4]

electronic document and records management system (EDRMS)
Software that has the ability to manage documents and records.

electronic record
Information recorded in a form that requires a computer or other machine to process and view it and that satisfies the legal or business definition of a record.

electronic records management (ERM)
Electronic records management is the management of electronic and nonelectronic records by software, including maintaining disposition schedules for keeping records for specified retention periods, archiving, or destruction. (*For* enterprise rights management, *see* Information Rights Management [IRM].)

electronically stored information (ESI)
A term coined by the legal community to connote any information at all that is stored by electronic means; this can include not just e-mail and e-documents but also audio and video recordings, and any other type of information stored on electronic media. ESI is a term that was created in 2006 when the US Federal Rules of Civil Procedure (FRCP) were revised to include the governance of ESI in litigation.

e-mail and e-document encryption
E-mail and e-document encryption refers to encryption or scrambling (and often authentication) of e-mail messages, which can be done in order to protect the content from being read by unintended recipients.

emulation
Software that mimics the behavior of another computer environment, often used to maintain compatibility of dated software or records in digital preservation. There can be legal concerns that arise regarding fidelity.

Encryption
It is the process of encoding or scrambling a message or information so that only authorized parties with the proper encryption key may access the information in a readable format.

enterprise content management (ECM)
Software that manages unstructured information such as e-documents, document images, e-mail, word processing documents, spreadsheets, Web content, and other documents; most systems also include some records management capability.

enterprise mobility management (EMM)
Software that enables the secure use of mobile devices and applications. Allows for IT to add and update apps to enable knowledge workers to complete work on mobile devices.

enterprise process analytics
Enterprise process analytics provides digital feedback on the status or various business processes in an organization, usually represented is a dashboard format so that management may understand the efficiency of business operations.

Enterprise risk management
It is the process of identifying and assessing the relative seriousness and likelihood of risks and organization faces, and crafting counter-measures to reduce the risks or their impact.

event-based disposition
A disposition instruction in which a record is eligible for the specified disposition (transfer or destroy) upon when or immediately after the specified event occurs. No retention period is applied and there is no fixed waiting period as with timed or combination timed-event dispositions. Example: *Destroy when no longer needed for current operations*.

expected value (EV)
Expected value is a calculation of the potential financial value of the impact of a risk, multiplied by the percentage likelihood that the risk event will occur.

faceted search
Faceted search helps those searching for information to narrow their options and more quickly find the information they are looking for.

faceted taxonomy
Faceted taxonomies allow for multiple organizing principles to be applied to information along various dimensions. Facets can contain subjects, departments, business units, processes, tasks, interests, security levels, and other attributes used to describe information. There is never really one single taxonomy but rather collections of taxonomies that describe different aspects of information.

Federal Rules of Civil Procedure (FRCP)—Amended 2006
In US civil litigation, the FRCP governs the discovery and exchange of electronically stored information (ESI), which includes not only e-mail but all forms of information that can be stored electronically.

file plan
A file plan is a graphic representation of the business classification scheme (BCS), usually a "hierarchical structure consisting of headings and folders to indicate where and when records should be created during the conducting of the business of an office. In other words *the file plan links the records to their business context*."

folksonomy
The term used for a free-form, social approach to metadata assignment. Folksonomies are not an ordered classification system; rather, they are a list of keywords input by users that are ranked by popularity.[5]

functional retention schedule
Groups records series based on business functions, such as financial, legal, product management, or sales. Each function or grouping is also used for classification. Rather than detail every sequence of records, these larger functional groups are less numerous, and are easier for users to understand.

Generally Accepted Privacy Principles
A set of 10 Generally Accepted Privacy Principles, developed jointly by the Canadian Institute of Chartered Accountants and the American Institute of Certified Public Accountants through the AICPA/CICA Privacy Task Force. These principles can be used to guide the privacy aspects of an information

governance program.

Generally Accepted Recordkeeping Principles®

A set of eight Generally Accepted Recordkeeping Principles®, also known as "The Principles" within the records management community,[6] published in 2009 by US-based ARMA International to foster awareness of good recordkeeping practices and to provide guidance for records management maturity in organizations. These principles and associated metrics provide an **information governance** (IG) framework that can support continuous improvement.

governance, risk management, and compliance (GRC)

GRC is a high-level risk assessment set of tools to help senior and executive management assess the relative risks an organization faces, in the areas of compliance and governance.

governance model

A framework or model that can assist in guiding governance efforts. Examples include using a SharePoint governance model, the information governance reference model (IGRM), MIKE2.0, and others.

guiding principles

In developing a governance model, for instance for a SharePoint deployment, the basic principles used to guide its development. May include principles such as accountability (who is accountable for managing the site, who is accountable for certain content), who has authorized access to which documents, and whether or not the governance model is required for use, or to be used optionally as a reference.

Healthcare Insurance Portability and Accountability Act (HIPAA)

HIPAA was enacted by the US Congress in 1996. According to the Centers for Medicare and Medicaid Services (CMS) website, Title II of HIPAA, known as the administrative simplification (AS) provision, requires the establishment of national standards for electronic health care transactions and national identifiers for providers, health insurance plans, and employers.

hot site

A hot site is one that has identical or nearly identical hardware and operating system configurations, and copies of application software, and receives live, real-time backup data from business operations. In the event of a business interruption, the IT and electronic vital records operations can be switched over automatically, providing uninterrupted service.

identity and access management (IAM)

Sometimes referred to loosely as single sign-on, IAM is a framework of business processes, policies, and technologies that facilitates the management of electronic or digital identities. With an IAM framework in place IT managers can control user access to critical information within their organizations. IAM software offers role-based access control, which lets system administrators regulate access to systems or networks based on the roles of individual users within the enterprise.

IG Process Maturity Model (IGPMM)
A model from the Compliance, Governance, and Oversight Council that measures the maturity of 22 enterprise processes of IT, RIM, Legal, Privacy & Security, and Business units (the five key impact areas of the IG Reference Model) on four levels. Bringing these processes to maturity can reduce the risks and costs associated with evolving compliance and privacy regulations, increasingly expensive legal discovery, ever-growing storage requirements, and new data security challenges.

infonomics
The discipline that assigns "economic significance" to information and provides a framework to manage, measure, and monetize information.

information footprint
The total size of the amount of information an organization manages.

information asset risk mitigation plan
An information asset risk mitigation plan delineates the key risks an organization faces, and develops countermeasures to offset and reduce those risks.

information governance (IG)
IG is a subset of corporate governance and is an all-encompassing term for how an organization manages the totality of its information. IG "encompasses the policies and leveraged technologies meant to dictate and manage what corporate information is retained, where and for how long, and also how it is retained (e.g. protected, replicated, and secured). Information governance spans retention, security and life cycle management issues."[7] IG is an ongoing program that helps organizations meet external compliance and legal demands and internal governance rules.

information life cycle
The span of the use of information, from creation, through active use, storage, and final disposition, which may be destruction or preservation.

information management
The collection and management of information from one or more sources and the distribution of that information to one or more audiences. The process of collecting, storing, managing, and maintaining information in all its forms. IM broadly incorporates policies and procedures for centrally managing and sharing information among different individuals, organizations, and/or information systems throughout the information life cycle. Information management may also be called information asset management.

information and communications technology
ICT is a term that refers generally to information and communication technologies.

information map or data map
A graphic diagram that shows where information is created, where it resides, and the path it takes.

information quality
The accuracy, reliability, and quality of the content of information systems.

information rights management (IRM)
Information rights management (IRM) is often referred to as enterprise rights management. IRM applies to a technology set that protects sensitive information, usually documents or e-mail messages, from unauthorized access. IRM is technology that allows for information (mostly in the form of documents) to be remote controlled. This means that information and its control can be separately created, viewed, edited, and distributed. IRM is sometimes also referred to as enterprise digital rights management (E-DRM). This can cause confusion because digital rights management (DRM) technologies are typically associated with business-to-consumer systems designed to protect rich media such as music and video.

information technology
Technologies used to manage digital information.

inherited metadata
Automatically assigning certain metadata to records based on rules that are established in advance and set up by a system administrator.

Internet of Things
The IoT includes Internet-enabled connections to sensors, processors, and devices that collect, send, and act on data they acquire from their surrounding environments.

inventorying records
A descriptive listing of each record series or system, together with an indication of location and other pertinent data. It is not a list of each document or each folder but rather of each series or system."[8]

ISO30300:2011 Information and documentation – Management systems for records – Fundamentals and vocabulary
Defines terms and definitions applicable to the standards on management systems for records (MSR) prepared by ISO/TC 46/SC 11. It also establishes the objectives for using a MSR, provides principles for a MSR, describes a process approach and specifies roles for top management.

ISO/TR 18128:2014 Information and documentation – Risk assessment for records processes and systems
Assists organizations in assessing risks to records processes and systems so they can ensure records continue to meet identified business needs as long as required.

jukebox (optical disk jukebox)
Optical disc autochanger units for mass storage that use robotics to pick and mount optical disks, and remove and replace them after use; dubbed a "jukebox" for their similarity in mechanics to jukebox units for playing vinyl records, and later, CDs.

key performance indicators (KPIs)

Metrics that measure progress toward achieving key business objectives. Organizations use KPIs at multiple levels to evaluate their success at reaching targets.

keyword search

Searching for a particular word or phrase that describes the contents of an e-document or information, or a Web page. Keywords represent shortcuts that sum up an entire email, document, or Web page. Keywords form part of a Web page's metadata and help search engines match a page to with an appropriate search query.

knowledge management (KM)

The accumulation, organization, and use of experience and "lessons learned" that can be leveraged to improve future decision-making efforts. Often involves listing and indexing subject matter experts, project categories, reports, studies, proposals and other intellectual property sources or outputs that is retained to build corporate memory. Good KM systems help train new employees and reduce the impact of turnover and retirement of key employees.

legal hold

Also known as a preservation order or hold order, a legal hold is a temporary suspension of the company's document retention destruction policies for the documents that may be relevant to a lawsuit or that are reasonably anticipated to be relevant. It is a stipulation requiring the company to preserve all data that may relate to a legal action involving the company. A litigation hold ensures that the documents relating to the litigation are not destroyed and are available for the discovery process prior to litigation. The legal hold process is a foundational element of information governance.

legal hold notification (LHN)

The process of notifying employees of certain date ranges and topics or categories of information that must be preserved and not modified or deleted in preparation for litigation.

limitation period

The length of time after which a legal action cannot be brought before the courts. Limitation periods are important because they determine the length of time records must be kept to support court action [including subsequent appeal periods]. It is important to be familiar with the purpose, principles, and special circumstances that affect limitation periods and therefore records retention."[9]

long-term digital preservation (LTDP)

The managed activities, methods, standards and technologies used to provide long-term, error-free storage of digital information, with means for retrieval and interpretation, for the entire time span the information is required to be retained.

machine learning

A category of AI algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. It is a branch of AI based on the idea that systems can learn from data, identify

patterns, and make decisions with minimal human intervention.

master retention schedule
A retention schedule which includes the retention and disposition requirements for records series that cross business unit boundaries. *The master retention schedule contains all records series in the entire enterprise.*

meaningful use
In the context of health IT, meaningful use is a term used to define minimum US government standards for electronic health records (EHR), outlining how clinical patient data should be exchanged between healthcare providers, between providers and insurers, and between providers and patients. It is typically estimated to be about 40% of the capabilities of the EHR software app. Meaningful use ensures that the certified EHR technology is connected in a manner that provides for the electronic exchange of health information to improve the quality of care.

metadata
Metadata is short descriptive information about an email, document or database such as its author, date and time created, length, language, and business unit owner.

migration
Migration is the transfer of a file from one storage medium to another, to ensure its future readability. An example is moving files form old floppy disks, to optical discs. Maintaining file integrity, accuracy, and readability is of paramount importance.

mobile device management (MDM)
A type of security software used to monitor, manage, and secure employees' mobile devices that are deployed across multiple mobile service providers and across multiple mobile operating systems being used in the organization. Allows for remote wiping of lost devices, software upgrades *en masse,* and overall management of mobile devices within a network.

negotiated procurement
A negotiated procurement is where a consultant or other third party negotiates the purchase of hardware, software, and/or services on behalf of a client, without putting the project out to bid.

NENR
It is non-erasable, non-rewritable storage which is unalterable. NENR is often used in financial institutions to prevent adulteration of information. NENR storage includes Write-Once, Read-Many (WORM) tape and optical media, disk and disk-and-tape blended.

OAIS
Reference Model for an Open Archival Information System describes how to prepare and submit digital objects for long-term digital preservation (LTDP) and retrieval but does not specify technologies, techniques, or content types. The OAIS Reference Model defines an archival information system as an archive, consisting of an organization of people and systems that has accepted

the responsibility to preserve information and make it available and understandable for a Designated Community (i.e. potential users or consumers), who should be able to understand the information. Thus, the context of an OAIS-compliant digital repository includes producers who originate the information to be preserved in the repository, consumers who retrieve the information, and a management/organization that hosts and administers the digital assets being preserved. The OAIS Information Model employs three types of information packages: A Submission Information Package (SIP), an Archival Information Package (AIP), and a Dissemination Information Package (DIP). An OAIS-compliant digital repository preserves AIPs and any PDI associated with them. A Submission Information Package encompasses digital content that a Producer has organized for submission to the OAIS. After the completion of quality assurance and normalization procedures, an Archival Information Package is created, which as noted previously is the focus of preservation activity. Subsequently, a Dissemination Package is created that consists of an AIP or information extracted from an AIP that is customized to the requirements of the Designated Community of users and consumers.

optical character recognition (OCR)
OCR is the process of using optical reading technologies to read data from a paper form or documents. Optical disc is a highly durable storage medium similar to DVD that uses lasers to read information.

optical disk
Round, platter-shaped storage media written to using laser technologies. Optical disk drives use lasers to record and retrieve information, and optical media has a much longer useful life (some purported to be 100 years or more) than magnetic.

Penetration testing
It is also called pen testing or ethical hacking, is the practice of testing a computer system, network or web application to find security vulnerabilities that an attacker could exploit.

pattern search
A search for patterns or concrete concepts when searching a corpus of e-documents. It attempts to find the best match (the solution that has the lowest error value). Often used in e-discovery phase of litigation.

personally identifiable information (PII)
Information about individuals that identifies them personally, such as Social Security number, address, credit card information, health information, and the like. PII is subject to privacy laws.

phishing
Phishing is a way of attempting to acquire sensitive information such as user names, passwords, and credit card details by masquerading as a trustworthy entity in an electronic communication. Communications purporting to be from popular social websites, auction sites, online payment processors, or IT

administrators are commonly used to lure the unsuspecting public. Phishing is typically carried out by e-mail or instant messaging, and it often directs users to enter details at a fake website that looks and feels almost identical to the legitimate one. Phishing is an example of social engineering techniques used to fool users, and it exploits the poor usability of current web security technologies.

policy
A high-level overall plan, containing a set of principles that embrace the general goals of the organization and are used as a basis for decisions. Can include some specifics of processes allowed and not allowed.

predictive analytics
A proactive type of data analytics that provides a forecast of what might happen.

predictive coding
The use of keyword search, filtering, and sampling to automate portions of an e-discovery document review. The goal of predictive coding is to reduce the number of irrelevant and nonresponsive documents that need to be reviewed manually.

prescriptive analytics
A proactive type of data analytics that formulates rules and recommendations based on historic data and other forward-looking data points.

preservation description information (PDI)
In the LTDP process, adhering to the OAIS reference model, description information such as provenance, context, and fixity.

privacy awareness training
A privacy training program for employees that raises privacy literacy and awareness in organizations. May use animations in short vignettes and may be gamified to improve user engagement.

process-enabled technologies
Information technologies that automate and streamline business processes. Process-enabled technologies are often divided into two categories: workflow automation or business process management. The two technologies have a significant amount in common. Indeed it is fair to say that a good deal of the technology that underpins business process management concepts has its roots in the late 1980s and early 1990s and stems from the early efforts of the workflow community.

project charter
A document that formally authorize**s a** project to move forward. "A project charter dramatically reduces the risk of a project being cancelled due to lack of support or perceived value to the company. It documents the overall objectives of the project and helps manage the expectations."[10]

project manager
The person primarily responsible for managing a project to its successful

completion.

project plan
Includes the project charter and project schedule and a delineation of all project team members and their roles and responsibilities.

project schedule
A listing of project tasks, subtasks, and estimated completion times.

policy
A high-level overall plan, containing a set of principles that embrace the general goals of the organization and are used as a basis for decisions. Can include some specifics of processes allowed and not allowed.

provenance
In records management, provenance is information about who created a record and what it is used for.

record category
A description of a particular set of records within a file plan. Each category has retention and disposition data associated with it, applied to all record folders and records within the category.

records appraisal
The process of assessing the value and risk of records to determine their retention and disposition requirements. Legal research is outlined in appraisal reports. This may be accomplished as a part of the process of developing the records retention schedules, as well as conducting a regular review to ensure that citations and requirements are current.

Ransomware as a service (RaaS)
RaaS is an approach to ransomware where back hat hackers sell ransomware through a cloud-based platform.

records integrity
The accuracy and consistency of records, and the assurance that they are genuine and unaltered.

records management (RM) or records and information management (RIM))
The field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use, and disposition of records, including processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records. A set of instructions allocated to a class or file to determine the length of time for which records should be retained by the organization for business purposes, and the eventual fate of the records on completion of this period of time.

records retention schedule/records retention period
A records retention schedule spells out how long different types of records are to be held, and how they will be archived or disposed of at the end of their life cycle. It considers legal, regulatory, operational, and historical requirements.[11]

records series

A group or unit of identical or related records that are normally used and filed as a unit and that can be evaluated as a unit or business function for scheduling purposes.[12]

responsibility assignment (RACI) matrix
A RACI matrix, or responsibility assignment matrix spells out who is Responsible, Accountable, Consulted, and Informed in a particular project or program.

refreshment
The process of copying stored e-records to new copies of the same media, to extend the storage life of the record by using new media.

return on investment (ROI)
"A performance measure used to evaluate the efficiency of an investment .... To calculate ROI, the benefit (return) of an investment is divided by the cost of the investment; the result is expressed as a percentage or a ratio."[13]

risk management framework
The structured process used to identify potential risks and threats an organization faces and to develop offsetting countermeasures to minimize the impact of these risks, as well as the control mechanisms to effectively monitor and evaluate this strategy.

risk profile
A listing of risks an organization faces and their relative likelihood; used as a basic building block in enterprise risk management that assists executives in understanding the risks associated with stated business objectives, and allocating resources, within a structured evaluation approach or framework.

secure sockets layer (SSL)/transport layer security (TLS)
Secure sockets layer (SSL) and transport layer security (TLS) are cryptographic protocols that provide communications security over the Internet. SSL and TLS encrypt the segments of network connections above the transport layer, using symmetric cryptography for privacy and a keyed message authentication code for message reliability.

senior records officer (SRO)
The leading records manager in an organization; may also be titled chief records officer or similar.

service-level agreement (SLA)
The service or maintenance contract that states the explicit levels of support, response time windows or ranges, escalation procedures in the event of a persistent problem, and possible penalties for nonconformance in the event the vendor does not meet its contractual obligations.

service-oriented architecture (SOA)
An IT architecture that separates infrastructure, applications, and data into layers.

Six Sigma
Six sigma is a disciplined, statistical-based, data-driven approach and

continuous improvement methodology for eliminating defects in a product, process or service.

smishing
SMS is short message service, which is texting on smartphones and mobile devices. SMiShing is a security attack in which the user is tricked into downloading a Trojan horse, virus, or other malware onto his cellular phone or other mobile device. SMiShing is short for SMS phishing.

social engineering
The term used for a broad range of malicious activities accomplished through human interactions. It uses psychological manipulation to trick users into making security mistakes or giving away sensitive information.

social tagging
A method that allows users to manage content with metadata they apply themselves using keywords or metadata tags. Unlike traditional classification, which uses a controlled vocabulary, *social tagging keywords are freely chosen by each individual*. This can help uncover new categories of documents that are emerging, and helps users find information using their terms they believe are relevant.

solid state disk drive
A solid state drive is storage that has no moving parts, similar to computer memory.

spoliation
The loss of proven authenticity of a record. Can occur in the case of e-mail records if they are not captured in real time, or they have been edited in any way.

structured data
A collection of records or data that is stored in a computer; records maintained in a database or application.

subject matter expert (SME)
A person with deep knowledge of a particular topical area. SMEs can be useful in the consultation phase of the taxonomy design process.

subject records (or topic or function records)
Subject records (also referred to as topic or function records) "contain information relating to specific or general topics and that are arranged according to their informational content or by the function/activity/transaction they pertain to."[14]

submission information package (SIP)
One of three types of information packages that can be submitted in the OAIS preservation model.

technology-assisted review (TAR)
Technology Assisted Review (TAR) is a process of having computer software electronically classify documents based on input from expert reviewers, in an effort to expedite the organization and prioritization of the document

collection. typically during the e-discovery process.

taxonomy
A hierarchical structure of information components, for example, a subject, business-unit, or functional taxonomy, any part of which can be used to classify a content item in relation to other items in the structure.

text analytics
The process of deriving high-quality information from text, typically through the devising of patterns and trends through means such as statistical pattern learning. Also known as text mining.

text mining
Performing detailed full-text searches on the content of document. See *text analytics*.

thesaurus
In taxonomies, a thesaurus contains all synonyms and definitions, is used to enforce naming conventions in a controlled vocabulary, for example, *invoice* and *bill* could be terms that are used interchangeably.

time-/date-based disposition
A disposition instruction specifying when a record shall be cut off and when a fixed retention period is applied. The retention period does not begin until after the records have been cut off, for example: destroy after two years.

total cost of ownership (TCO)
TCO is the calculation of to total cost of a computing system including hardware, software, maintenance, and other related costs throughout the lifespan of a computing system.

transfer
Moving records from one location to another, or change of custody, ownership, and/or responsibility for records.

unstructured information
Records that are not expressed in numerical rows and columns but rather are objects such as image files, e-mail files, Microsoft Office files, and so forth. Structured records are maintained in databases.

usage (records)
The purpose a record is used for, i.e. its primary use.

vulnerability assessment
A vulnerability assessment is the process of defining, identifying, classifying and prioritizing vulnerabilities in computer systems, applications and network infrastructures and providing the organization with an understanding of the threats and how to counter them.

vital records
Vital records are mission-critical records that are necessary for an organization to continue to operate in the event of disruption or disaster and cannot be recreated from any other source. Typically, they make up about 3%-5% of an organization's total records. They are the most important records to be

protected, and a plan for disaster recovery (DR)/business continuity (BC) must be in place to safeguard these records.

warm site
A warm site has the hardware and operating systems the main data center has, and likely the applications, but needs data loaded to go online and resume processing when a main site is damaged or compromised.

workflow, workflow automation, and workflow software
Software that can route electronic folders through a series of worksteps to speed processing and improve auditability. Not to be confused with business process management systems (BPMS), which have more robust capabilities.

WORM
Write Once Read Many optical disk storage media that is nonerasable, and can only be written to one time.

Notes

1. TechTarget.com, "Authentication, Authorization, and Accounting," http://searchsecurity.techtarget.com/definition/authentication-authorization-and-accounting (accessed December 5, 2012).

2. John O'Connell, Jon Pyke, and Roger Whitehead, *Mastering Your Organization's Processes* (Cambridge, UK: Cambridge University Press, 2006), 14.

3. The U.S. Government Printing Office (GPO), "Code of Federal Regulations," www.gpo.gov/help/index.html#about_code_of_federal_regulations.htm (accessed April 22, 2012).

4. National Archives and Records Administration, "Electronic Code of Federal Regulations," October 2, 2012, http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl.

5. Tom Reamy, "Folksonomy Folktales," *KM World*, September 29, 2009, www.kmworld.com/Articles/Editorial/Feature/Folksonomy-folktales-56210.aspx.

6. ARMA International, "How to Cite GARP," www.arma.org/garp/copyright.cfm (accessed May 8, 2012). This chapter was contributed by Charmaine Brooks, CRM.

7. Kathleen Reidy, "The Rise of Information Governance," *Too Much Information: The 451 Take on Information Management* (blog), August 5, 2009, http://blogs.the451group.com/information_management/2009/08/05/the-rise-of-information-governance/.

8. U.S. National Archives and Records Administration, "Disposition of Federal

Records: A Records Management Handbook," www.archives.gov/records-mgmt/publications/disposition-of-federal-records/chapter-3.html (accessed April 3, 2012).

9. Government of Alberta, "Developing Retention and Disposition Schedules," p. 122.

10. Rita Mulcahy, "Project Management Crash Course: What Is a Project Charter?" October 28, 2009, www.ciscopress.com/articles/article.asp?p=1400865.

11. National Archives, "Frequently Asked Questions about Records Scheduling and Disposition," updated June 6, 2005, www.archives.gov/records-mgmt/faqs/scheduling.html#whysched.

12. University of Toronto Archives, "Glossary," www.library.utoronto.ca/utarms/info/glossary.html (accessed September 10, 2012).

13. Investopedia website, "Return on Investment," www.investopedia.com/terms/r/returnoninvestment.asp#axzz2E6SXDDOc (accessed December 4, 2012).

14. Ibid.

ABOUT THE AUTHOR

Robert F. Smallwood, MBA, CIP, IGP, is founder of the Institute for Information Governance, a specialty training and consulting practice, and CEO, publisher, and cofounder of *Information Governance World* magazine. Mr. Smallwood has over 35 years of experience in the information technology industry and holds an MBA in International Business from Loyola University of New Orleans. He is recognized as the world's leading author and trainer in IG. He consults with Fortune 500 companies and governments to assist them in making technology decisions and implementations. Some of his past research and consulting clients include Abbott Labs, Kirkwood and Ellis LLP, NASA, Novartis Pharmaceuticals, Pepsi, and Verizon. He has published more than 100 articles and given more than 50 conference presentations on documents, records, and content management. He is the author of *Managing Electronic Records: Methods, Best Practices, and Technologies* (John Wiley & Sons, 2013), *Safeguarding Critical E-Documents* (John Wiley & Sons, 2012), *Managing Social Media Business Records, Taming the Email Tiger,* and several other books, including a novel, a theatrical play, and the first published personal account of Hurricane Katrina.

ABOUT THE MAJOR CONTRIBUTORS

Lori J. Ashley is a Wisconsin-based consultant, writer, and educator dedicated to helping clients improve the performance of their record and information management practices and controls. An experienced business strategist and organizational development specialist, she has codeveloped four continuous improvement methodologies aimed at jump-starting collaboration among stakeholders who share accountability for effective and efficient life cycle management of valued records and information assets.

Barbara Blackburn, CRM, is an electronic records management consultant who assists organizations in defining, researching, selecting, and implementing cost-effective solutions. She assists clients in preparing for technology deployment by providing strategic planning and developing recordkeeping programs and taxonomies. Ms. Blackburn has expert taxonomy design skills and has taught AIIM's Electronic Records Management and Electronic Content Management certification classroom courses.

Baird Brueseke, CIP, IGP, is an IG consultant and trainer with a focus on cybersecurity issues, a cofounder of *Information Governance World* magazine. He is a cybersecurity policy and controls expert, with special expertise in vulnerability assessments and gap analysis. With over 30 years' experience in the technology field, Mr. Brueseke is coauthor of a patent on cloud-based instruction, and a frequent speaker and instructor on cybersecurity topics. He is a cofounder and a board member of the Cyber-Security Institute of San Diego and worked on developing the NIST cybersecurity framework. He also was the cofounder of Wheb systems, which developed and marketed the first Windows-based electronic forms processing system that could read handwriting. Wheb, through growth and acquisitions, became Captiva, a company with over $400 million in revenues. Mr. Brueseke is active in AIIM International and has served as the San Diego chapter president.

Barclay T. Blair is an advisor to Fortune 500 companies, software and hardware vendors, and government institutions and is an author, speaker, and internationally recognized authority on information governance. He has led several high-profile consulting engagements at the world's leading institutions to help them globally transform the way they manage information. Mr. Blair is the president and founder of ViaLumina and Executive Director and co-founder of the IG Initiative.

Charmaine Brooks, CRM, is a principal with IMERGE Consulting, Inc., and has over 35 years of experience in records and information management and content management. Ms. Brooks is a certified trainer and has taught AIIM classroom courses on ERM and provided many workshops for ARMA. Formerly a records manager for a leading worldwide provider of semiconductor memory solutions and a manager in a records management software development company, today Ms.

Brooks provides clients, small and large, public and private, with guidance in developing records management and information governance programs.

Monica Crocker, CRM, PMP, CIP, is a corporate records manager for Wells Fargo. Ms. Crocker has also been an information management consultant for 20 years, defining content and records management best practices for organizations across the United States. Her expertise includes SharePoint governance, enterprise strategies for content management, records management, electronic discovery, taxonomy design, project management, and business process redesign. Ms. Crocker is a recipient of AIIM's Distinguished Service Award.

Charles M. Dollar, now retired, is an internationally recognized archival educator, consultant, and author who draws on more than three decades of knowledge and experience in working with public and private sector organizations to optimize the use of information technologies to satisfy legal, regulatory, business, and cultural memory recordkeeping requirements for digital preservation. He is codeveloper of a capability maturity model for long-term digital preservation that incorporates the specifications of ISO 15489, ISO 14721, ISO 18492, and ISO 16363.

Mark Driskill is a lifelong California resident. He holds a BA in American history and a Master's in archives and records management from San José State University. His professional interests include the philosophy of information management, corporate archives, and the social aspects surrounding our digital footprints. Driskill currently works as a freelance writer/editor and contributing editor to *IG World* magazine.

Patricia Franks, PhD, CA, CRM, IGP, FAI, is a certified records manager and information governance professional and the coordinator for the Master of Archives and Records Administration degree program in the School of Library and Information Science at San José State University. She served as the team lead for both the ANSI/ARMA standard released in January of 2011, *Implications of Web-based Collaborative Technologies in Records Management*, and the 2012 technical report, *Using Social Media in Organizations*. Her latest publication, *Records and Information Management* (ALA Neal-Schuman, 2013) offers insight into a range of topics affecting records and information management professionals.

Darra Hoffman, J.D., M.S.L.S., is completing her PhD at University of British Columbia, School of Library, Archival, and Information Studies, where she is a research assistant with Interpares Trust and Blockchain at UBC. Her research focuses on the intersection between records, technology, and human rights, with a particular emphasis on blockchain. Her research is supported by a Canada Graduate Scholarship and a Killam Doctoral Scholarship.

Randolph Kahn, Esq., is the founder of Kahn Consulting, one of the premier information governance advisory firms. The Kahn Consulting team has provided consulting services to major global organizations including, advising the US and foreign governments, courts systems, and major multinational corporations on a

wide variety of information issues, including e-communications strategies, social media policy, records management programs implementation, and litigation response processes. Mr. Kahn is a highly sought-after speaker and a two-time recipient of the Britt Literary Award. He has authored dozens of published works, including *Chucking Daisies*, his new book on defensible disposition; *Email Rules*; *Information Nation: Seven Keys to Information Management Compliance*; *Information Nation Warrior*; and *Privacy Nation*. He is a cofounder of the Council for Information Auto-Classification and has been expert witness and an advocate in many industry organizations. Mr. Kahn is an attorney who attained his JD degree from Washington University in St. Louis, Missouri. Mr. Kahn has taught at George Washington University.

Doug Laney, the "Father of Infonomics," is a solutions-focused senior executive and board member with more than 35 years of success across the nonprofit, telecommunications, retail, insurance, higher education, IT, and IT services industries. Leveraging extensive experience in data strategy and innovation, he is a valuable advisor for an organization seeking guidance in monetizing data. His broad areas of expertise include business analysis, startup leadership, data management, and business intelligence. He is currently Principal Data Strategist with Caserta, a data warehousing and analytics consulting firm out of New York. Throughout his executive career, Doug has held leadership positions with Gartner, the University of Illinois at Urbana–Champaign, the GIES College of Business, Forbes, and Deloitte Consulting.

Barry Murphy is Head of Marketing, AWS Customer Enablement at Amazon Web Services, and Principal Consultant of Murphy Insights. He has performed product marketing for a variety of technology provider clients. He has served as Product & Content Marketing Manager at Mirakl, and Head of Product Marketing & Alliances at X1. Mr. Murphy was a cofounder of eDJ Group, Inc., and a thought leader in information governance, e-discovery, records management, and content archiving. Previously, he was director of product marketing at Mimosa Systems, a leading content archiving and e-discovery software provider. He joined Mimosa after a highly successful stint as principal analyst for e-discovery, records management, and content archiving at Forrester Research. Mr. Murphy received a BS from the State University of New York at Binghamton and an MBA from the University of Notre Dame. He is an active member of both AIIM and ARMA.

Andrew Ysasi is an executive in the information services and storage industries. Since 2009, Mr. Ysassi's leadership at Kent Record Management led to a 100% increase in revenue and increasing the valuation of the business by 15 times. He was able to increase profitability, cash flow, and strengthen the balance sheet by streamlining operations to client needs, launching new services, and by building an industry-recognized sales and marketing strategy. To improve chances of success, he became a subject matter expert on leadership, information governance, records management, business development, information privacy, information security, project management, and sales management. Mr. Ysasi has received over 40 written recommendations from professional colleagues, advanced certifications, a

Master of Science in Administration degree, is a sought-after speaker, and is respected by his colleagues and clients across the United States. Recently, Ysasi has conducted some short, content-rich, and entertaining Pecha-Kucha and Ignite presentations at national conferences.

In addition to his role as Executive Director at Kent Records, he is a former adjunct professor at Davenport University and taught global project management and the technology capstone courses. He started IG GURU®, which is an Information Governance News and Community website. He is the founder of the career coaching company Admovio®, where his resume review work and career advice is published on CIO.com. He founded myPACT to enable individuals to showcase themselves to employers and has a patent pending for the technology. He has served as a volunteer test writer on the Institute of Certified Records Managers (ICRM) Exam Development Committee (EDC) since 2012, and he is the Regent for the EDC for the 2017-2019 term. Andrew was also elected to the PRISM International Board of Directors through 2019. Further, he has served on local ARMA boards throughout his career, was Western Michigan ARMA's chapter member of the year in 2013, and was an author on the second edition of the ARMA RIM Core Competencies. Andrew is a Certified Records Manager (CRM), a Certified Information Privacy Manager (CIPM), a Certified Information Privacy Professional (CIPP), Fellow of Information Privacy (FIP), Project Management Professional (PMP), and Certified Information Security Manager (CISM), and was part of the inaugural group to achieve the Information Governance Professional (IGP) certification. Andrew also holds various certifications from Microsoft and CompTIA.

INDEX

3G/4G interoperability, 322

Abatan, Peter, 271
Accenture, IG failure, 13–14
Access, 419
  authorization, absence, 346–347
  control methods, 32
  GAPP criterion, 45
  rights policies, development/updating, 115
Access control, 95
  identity access management, usage, 254–255
Accountability, 234
  GAR Principle component, 35, 37–38, 88
Accounting business-unit taxonomy, 454f
Actionable plans, development, 82
Active Directory Federation Services (ADFS), 354
Administration, decentralization, 267
Administrative metadata, 440
Adverse events, financial/operational impact (determination), 58
Adverse inference, 139
Agent-based solutions, usage, 264
Agreed-on trigger events, automated capture, 203
Amazon Machine Learning, 363
American Institute of Certified Public Accountants (AICPA), 45
American National Standards Institute (ANSI), 93, 449
American Recovery and Reinvestment Act, 241
Analytics, 357, 359–363
  leverage, 79, 92
  technologies, 155
Anonymous social networks, usage, 302

Anthem, Inc., IG failure, 12–13
Anticipation (executive sponsor purpose), 70
Anti-virus (mobile computing trend), 323
Apple iOS, usage, 326
Application programming interface (API), 218, 345
Applications catalogue, 188–189
Approval (executive sponsor purpose), 70
Archival information package (AIP), 398–403
Archival storage, 398, 415
Archive
  disposition method, 41
  management, 188
Archives Act, 469
ArchivesSocial, 310
Archiving technology, 289
ARMA International, 88, 118, 171, 198
  GAR Principles, 91
Artificial intelligence (AI), 131, 166, 357, 363–366
  deployment, 325
  software, 147
Artificial intelligence as a service (AIaaS), 363
AS 8015, principles, 96
Asia-Pacific Economic Cooperation (APEC), 243–244
Asia, privacy, 243–244
Asset management, 95
Associates in Psychiatry and Psychology (APP), IG failure, 11
Audience, identification, 441
Auditing, 267
Auditor, usage, 150
Australasian Digital Recordkeeping Initiative (ADRI), 96

Australia
Archives Act, 469
Crimes Act, 471
Electronic Transactions Act, 471
Evidence Act, 471
Financial Management and Accountability Act, 471
Freedom of Information Act (FOIA), 469–470
Information Commissioner Act, 470
Privacy Act, 470
records management, laws/regulations, 469–471
Australian ERM standards, 101–102
Australian Government Locator Service (AGLS), 102
Australian Information Commissioner Act, 470
Australian records standards, development, 102
Australian standards, 102
Authentication, authorization, and audit (AAA) controls, 342
Auto-categorization analytics software, implementation, 116
Auto-classification, 155, 157, 365, 434
Automatic archiving, usage, 288–291
Automation, 126
Availability (GAR Principle component), 35, 40, 88
Azure Active Directory service, 354

Backups, 11
Backward compatibility, 418
Barnhart, Brent, 300
Best practices, 15, 91–92, 315–316, 330
Better Business Bureau, 232

Big Data, 75, 216, 358, 372
challenges, 88
data debris/information, value (loss), 79–80, 92
impact, 5–7, 137–138
opportunities/challenges, 122–123
tools/techniques, 128
Bitstream, 399
readability (maintenance), device/media removal (usage), 418
Blockchain, 357, 366–372
interrelated solution layers, 371f
solution, 369–371, 370f
technology, 366
Blogs, 308, 314
Blue Pill root technique, 346
Blueprints, protection, 271–272
Botnets, impact, 344
Brand equity, preservation/protection, 116
Breaches, 56
Bring-your-own-device (BYOD), 324–325, 331, 340, 349
British Standards Institute (BSI), 92–93, 449
Broadcast Wave Format (BWF), 409
Broad network access, 337
Brown, Jerry, 243
Budget (executive sponsor purpose), 70
Build and maintain (Navy Yard rebirth phase), 127, 128

Business, 76–77, 196, 212, 276, 397
activity, support (e-mail documentation), 198
goals/objectives (support), IG driving programs (creation), 82–83
IM use, best practices, 292–293
objective alignment, usage, 115–116
plans, change, 430
running, professionalism, 168
terms, usage (standardization), 212
Business Associates (BA), 240
Business classification scheme (BCS), 446–447
Business continuity (BC), 7, 87, 164
Business continuity management (BCM), 104
Business intelligence (BI), 131, 214
Business processes, 55, 126, 453–456
automation, 166
example, travel expense process, 456f
redesign, 79, 80, 92
Business process improvement (BPI) opportunities, 456
Business process management system (BPMS) software, 34
Business units, 26, 115
candidates, 117
data quality, accountability assignment, 21, 215

California Consumer Privacy Act (CCPA), 3, 14, 161, 242–243
Canada, records management (laws/regulations), 466
Canada Revenue Agency, 99, 466
Canadian Institute of Chartered Accountants (CICA), 45
Canadian Records Retention Database (excerpt), 201f
Capture, 39
Center for Internet Security (CIS), 250–251, 349–350
Database Server Benchmarks, 224
Mobile Companion Guide, release, 349
Certificate authority (CA), certificate issuance, 261

Certification and accreditation (C&A) report, 348
Certification checklists, 93
Certified Ethical Hacker (CEH), 252
Certified Information Privacy Manager (CIPM), 239
Change management (CM), 21, 33, 396, 426–429
Channel messaging, solutions, 294–295
Chief Information Governance Officer (CIGO), impact, 106, 425
Chief Information Security Officer (CISO), 64
Chief knowledge officer (CKO), 72
Children's Online Privacy Protection Act (COPPA), 242
Chipotle Mexican Grill, IG failure, 11–12
Choice/consent (GAPP criterion), 46
CIS Top 20, 62t, 116
Civil litigation, requirements, 164
Classification, 162, 435
  micro-classification, avoidance, 185
Clean (Navy Yard rebirth phase), 127
Clinton, Hillary (email usage), 288
Closure, 203–204
CloudBurst, 346
Cloud computing, 165, 336–341
  documents/records, management, 351
  hacking, 344
  information, 335, 341–342
  insider threats, 343
  rogue intrusions, 344
  security threats, 341–350, 352
  solutions, 351–352
Cloud connection, insecure points, 344–345
Cloud Controls Matrix (Cloud Security Alliance), 250
Cloud deployment models, 339
Cloud file storage, 340

Cloud Machine Learning, 363
Cloud Security Alliance, 341
Cloud services, 344
Code of Federal Regulations (CFR), 58, 200–201, 464–465
Collection
  GAPP criterion, 46
  limitation principle, 233
  policy, 401
Committee of Sponsoring Organizations of the Treadway Commission (COSO), 23
Communications, 95
  encryption, 325
  plan, creation, 441–442
  systems/network communications failure, 396
Community cloud, 339
Community government business-unit taxonomy, 452f
Complexity fees, 124
Compliance, 95, 126, 165, 270
  culture, maintenance, 425
  GAR Principle component, 35, 39–40, 88
  IG best practices, 223–225
  management, audit, 206
  research, 200–201
  risks/costs, excess, 117
Compliance, Governance and Oversight Council (CGOC), 6, 79, 88, 118, 137
Component obsolescence, 396
Computer-aided design (CAD), 266, 271–272
Computer-assisted review, 135
Computer systems, failure, 396
Conceptual data modeling approach, 219
Confidential data, large print files (security issues), 258–259
Confidential documents, theft, 13
Confidential e-documents, challenge, 256–257

Confidentiality Integrity Availability (CIA) triad, 45
Confidentiality National Health Service (NHS), code of practice, 245
Confidential stream messaging, 275–276
Consensus mechanism, 367
Consultative Committee for Space Data Systems, OAIS development, 397
Consumer review networks, usage, 301
ContentandCode, 304
Content control, 316
  models, 315
Content, impact, 195
Content management principles, 315–316
Content services, 163
Contextual enforcement, 273
Continuous improvement, 33–34, 430
Continuous process improvement (CPI), 429
Controlled vocabulary, 436
Controlling the Assault of Non-Solicited Pornography and Marketing Act (CAN-SPAM), 242
Control Objectives for Information and (related) Technology (COBIT®), 22–23, 220–221
Corporate governance, 8
Corporate memory, preservation, 167
Corporate objectives, 115, 116
Corporate-owned, personally enabled (COPE) devices, 349
Cost factors, 125f
Cost sources, 125–126
Council of Information Auto-Classification, "Information Explosion," 152
Covered Entities (CE), 240
Credit card information (PCI), 30
Crimes Act, 471
Cross-functional communications, promotion, 315
Cross-functional IG council, establishment, 78, 92
Cross-functional IG team, creation, 145

Cross-functional mobile strategy team, formation, 331
Cross-mapping scheme, 446
Culture, establishment, 145
Current state capabilities, 410
Customer relationship management (CRM), usage, 327
Custom taxonomies, prebuilt taxonomies (contrast), 448–449
CyberArk, 248
Cyberattacks, proliferation, 246
Cybersecurity, 116, 245, 376
  assessments, 250–252
  considerations/approaches, 253–254
Cyber-security process, accountability, 56

Dashboards, usage, 124
Da Silva Moore v Publicis Groupe, 155

Data
access, right, 477
analytics, 357–358
architecture, 217–218
asset, uniqueness (recognition), 21
breach, 214, 477
categories, 219f
cleansing/scrubbing, 20, 213
debris, deletion, 79–80, 92
growth, 152–154
information, value (loss), 79–80
integration (data modeling approach), 219
management, 398
map, 195
modeling, 218–219, 218f
portability, 477
privacy, 169–171, 190
protected data, data security design pattern, 273
protection, 26, 190, 225, 478
quality, 215, 233
securing, approaches, 272–274
stewardship, 20
type, 236
value, computation, 214
Database
compliance, IG best practices, 223–225
security, 223–225, 270
strengthening, 224
Database activity monitoring (DAM), usage, 224–225
Data Controller (DC), 239

Data governance, 19–21, 213–215
efforts, focus, 79, 92
Information governance/IT governance, difference, 19
Data Governance Institute (DGI) data governance framework, 215, 216f
Data loss prevention (DLP), 26, 130, 224, 263–265
IRM, combination, 270
technology, 181, 262–265
Data Protection Act, 172, 245
Data Protection Officer, 478
Data quality, 21, 79
Decentralized architecture, 367
Decision-based processes, 181
Decision making
executive sponsor purpose, 70
improvement, analytics (leverage), 79, 92
De-duplication, 20, 213
Deep learning, 363
Defense in depth, 254
Defensibility, end state, 156–157
Defensible deletion, 145
Defensible disposal, 151
Defensible disposition, 135, 146, 153, 157
Deleted files, location, 258
Deposit agreements, 402
Descriptive analytics, 357, 358
Descriptive metadata, 440
Design and Implementation of Recordkeeping Systems (DIRKS), 96, 171
Device
control methods, 269
media renewal, 415
Diagnostic analytics, 357, 358
Digital content, indexing, 155

Digital information assets, preservation, 81, 92
Digital preservation, 393
capabilities, 410f, 411–412
capability maturity model, 412f
capability performance metrics, 416–417
infrastructure, 413–414
performance metrics, 417t
policy, 413
processes/services, 415–416
requirements, 409
standards, 397–403
strategies, 413, 417–419
techniques, 417–419
Digital records, 98, 155
Digital repository, 414–415, 419–420
Digital signatures, 34, 261–262
Digital subject access request (dSAR), 26
Digitization, 126
Direct observation, usage, 180
Disaster recovery (DR), 7
Discard (disposition method), 41
Disclosure (third parties) (GAPP criterion), 47
Discovery
costs, increase, 153
phase, 135
Disposition, 126, 162, 192
capabilities, 203
event-based disposition, prerequisites, 202–203
event-based retention scheduling, 201–202
final disposition criteria, 203–204
improvement, 156
timing, 205

Disposition (GAR Principle component), 36, 41–42, 88
Dissemination information package (DIP), 398, 414
Distributed ledger technology (DLT), 367
Dixon, Helen, 238
DLM Forum, 101
Document life cycle security (DLS) technology, 248, 342
Documents
  analytics, 275
  document type, term (usage), 196
  encryption, 262
  governance efforts, 31
  integrity/trustworthiness, 166
  labeling, 274–275
  management, 7, 351
  records, contrast, 442
Dominant principle, 466
Dublin Core Element Set, 444
Dublin Core Metadata Initiative (DCMI), 102, 442–444

Early case assessment (ECA), 74, 83
  e-discovery phase, 147
Economic environment, survey, 76–77

E-discovery, 10, 125, 135
assistance, technologies (usage), 147–151
collection, 63, 365
costs, 74, 124
dangers, 378
Federal Rules of Civil Procedure, 135–137
information governance (IG), 143–146, 143t–144t
issues, 377–380
legal case, 139
life cycle, 144
plan, development/execution, 143
planning, guidelines, 142–143
process, steps, 140
readiness, 145–146
review, 365
techniques, 140
usage, 146
E-Discovery Reference Model (EDRM), 140–143, 141f
E-documents, 6
external access, cessation, 258
protection, 257
repository-based approaches, limitations, 256–257
securing, challenge, 256–257
security, 80, 265
Elastic Compute Cloud service, usage, 34
Electronically stored information (ESI), 136, 140, 164, 199
forms, 313
inventorying, 142
search, 377
Electronic Code of Federal Regulations (e-CFR), 201
Electronic Communications Privacy Act (ECPA), 244–245
Electronic content, migration, 400

Electronic document and records management systems (EDRMS), 97, 101, 447
Electronic records
  inventory survey form, 178f–179f
  management, challenges, 395
  survey, 415
Electronic Records as Documentary Evidence (CAN/CGSB-72.34-2017), 99
Electronic records management (ERM), 31, 80, 161, 433, 438
  agreed-on trigger events, automated capture, 203
  Australian ERM, 101–102
  benefits, 166–167
  Canadian standards, 99–100
  importance, 173
  intangible benefits, 167–168
  legal considerations, 99–100
  retention/disposition capabilities, 203
  software, usage, 34
  system, acquisition/implementation, 174
  system/application, 314
  usage, 80
Electronic records management systems (ERMS), 101
Electronic Transactions Act, 471

E-mails, 287
archiving, 289
bucket approach, 291
destructive retention, 80, 147, 199–200, 290–291, 921
documentation, 198
encryption, 259, 273
information governance, usage, 285
integrity/admissibility (preservation), automatic archiving (usage), 288–291
messages, 80, 92, 197
records, 197–199, 290
retention period, 147, 199
risk, organizations (employee exposure), 286
social media, contrast, 305
Embedded protection, 268–270
Emerging technologies, leveraging/governing, 357
Empirical metrics, 401
Employees
social media security threat, 306–307
training, 116
Emulation, 418
Encryption, 92, 269, 342
Endpoint management, unification, 325
Enforcement, technology application, 14–15
Enterprise content management (ECM), 33, 256–257
ERM component, 163
software, 314
Enterprise data, 219, 273
Enterprise file synch and share (EFSS), 33
Enterprise information security practices, direct connection, 56
Enterprise mobility management (EMM), 321, 325, 350
Enterprise resource planning (ERP) database, 120
Enterprise, social media (usage), 304–305

Environmental Protection Agency (EPA), 167
Environmental security, 95
Equity value, protection, 56
E-records
  disposition, event-based retention scheduling, 201–202
  inventory, challenges, 172
  inventorying, 168–169
  management, 96–98
  retention, legal issue, 287–288
  shadow copies, 172
Etsy, Dan, 129
European Broadcasting Tech 3285: 2011 Broadcast Wave Format, 409
European Research Cluster on the Internet of Things, 374
European Union (EU) General Data Protection Regulation (GDPR), 3, 31, 235–239, 476–478
  compliance, 238, 364
  consent, 477
  data subject rights, 477–478
  enforcement/precedent setting, 238–239
  fine, 26
  impact, 118, 161
  penalties, 476–477
  requirements, 187
Event-based disposition, prerequisites, 202–203
Evidence Act, 100, 466–467, 471
Evidence-based practice (EBP) protocols, AI assistance, 365
Executive sponsors
  purposes, 70
  recruitment, 214
  role, 70–71
Executive sponsorship, 33
  importance, 78, 92
Expectation management (executive sponsor purpose), 70

Expense report, 437
Extended enterprise, 257–259
Extensibility, 444
eXtensible Markup Language (XML), 400, 405, 407
External factors, survey/evaluation, 75–81
Extract, Transform, and Load (ETL) capabilities, 219
Extraterritorial applicability, 476

Facebook, social media site, 300, 309
Facet-driven mechanism, 149–150
Faceted search, usage, 445
Faceted taxonomies, 452
Fair and Accurate Credit Transaction Act, 240
Fair Credit Reporting Act (FCRA), 240
Fair Information Practices (FIPS), 232–233
Federal Bureau of Investigation, IG failure, 13
Federal Deposit Insurance Corp. (FDIC), 330
Federal Information Security Management Act, 348
Federal Rules of Civil Procedure (FRCP), 14, 135–137, 164, 380
  Rule 34, 313
  rules, details, 138–139
Federal Trade Commission (FTC), 240
Federal Wiretap Act, 244
File analysis classification and remediation (FACR), 361, 365
Files
  deletion, tracking, 258
  delivery, response time (problem), 349
  plans, 446–447
  protection, 255–256
  remediation, 365
  response time, problem, 349
  series, 196
Filtering technologies, 148

Final disposition criteria, 203–304
Financial Institution Privacy Protection Act, 463
Financial Management and Accountability Act, 471
Firefox browser, usage, 309
Flexibility, decrease, 93
Flynn, Nancy, 275
Folksonomies, 458
Food and Drug Administration (FDA), records inspection, 464
Ford Motor Company, IG failure, 13
Formal IG Program Charter, creation, 78, 92
Freedom of Information Act (FOIA), 172, 191, 193, 245, 469–470
Free-text metadata field, option, 458
Full cost accounting (FCA), 123–124
Full disk encryption (FDE), 260, 269
Fully managed devices, 349
Functional taxonomy, implementation, 116
Function records, 197
Future costs, 124

Geithner, Timothy, 288
General and administrative costs, 124
Generally Accepted Privacy Principles (GAPP), 45, 231–232
  criteria, 46–48
Generally Accepted Recordkeeping Principles® (GAR Principles), 35–42, 79, 171, 193
  assessment/improvement roadmap, 42, 43t–44t
  IGRM, impact, 91
  improvement areas, 42t
  levels, 36t
  review, 88
Generation Gmail entry, 304
Global Aerospace, Inc., Et al. v. Landow Aviation, LP, et al., 148
Global Information Locator Service (GILS), 444–445

Global positioning systems (GPS), 322
"Going forward" strategy, implementation, 21
Goodgle Android, apps (usage), 328
Governance, risk management, and compliance (GRC), 77
Government oversight, increase, 163–164
Gramm-Leach-Bliley Act (GLBA), 463
Graphics image format (GIF), compression algorithms, 408
Gruman, Galen, 337

Hacker intrusion events, number (reduction), 63
Hacking, 346–347
  protection, 326
Health care, AI (usage), 364–365
Health Information Act in Canada, 191
Health Information Technology for Economy and Clinical Health (HITECH), 241
Health Insurance Portability and Accountability Act (HIPAA), 240–242, 463–464
  fine, 26
  Privacy Rule, inclusion, 241
Heat map, 60
High-level strategic plans, creation, 61–62
Hijacking, 346–347
HootSuite, 301
HP Security Research study, 379
Human error, impact, 396
Human resources security, 95
Human review (predictive coding component), 148–149
Hybrid cloud, 339
Hypertext transfer protocol secure (https), 325

IBM Watson, 363
Identity access management, usage, 254–255
Identity and access management (IAM), 254
Imaging (disposition method), 41
Implementation frameworks, 93

Indirect costs, 124
Individual participation principle, 234
Industrial Internet Consortium (IIC), 382
Industry
  best practices, 77–80, 430
  industry-specific best practices, 212
  regulation, increase, 163–164
Infonomics, 10, 117, 244
Infonomics (Laney), 4, 30, 117, 214
Information
  accessibility, 32
  baseline, 130
  breaches, 342
  cap-and-trade system, 130
  classification, technology (usage), 155–156
  confidence, obtaining, 26
  control, 32–33
  costs, calculation, 121–122
  creation/usage, mapping, 26
  custodians, 154
  delivery platforms, changes, 165
  effectiveness, impact, 152
  environment, changes, 119–121
  footprint, 6
  full cost accounting, 123–124
  growth, 151–152
  harvesting/leveraging, 26
  integrity, 31
  life cycle, 90
  loss, 341–342
  management, 80, 151–158, 216–219
  models, 129–130

monetization, 26
organization/classification, 31–32, 433
privacy, 32, 229–231
quality, 213–214
retrieval, 150
safeguarding, 167
search/access/collaboration, 126
security expert, hiring, 330
stakeholder, participation, 29
systems, acquisition/development/maintenance, 95
units (tracking/trading), internal accounting system (creation), 130
value, 117, 127–129, 145
Information and communication technologies (ICT), 75
Information Asset Register (IAR), 185–189
Information asset risk planning
audit/review/adjustment, 55
countermeasure, 55
impact, assessment, 55
management, 55
metrics, establishment, 55
plan, execution, 55
policy, creation, 55
process, benefits, 56
responsibilities, assignation, 55
steps, 55

Information assets
access/use, 190
confidentiality, 190
dates, 190
description, 190
ownership, 190
problems/issues, risk/impact, 191
protection/preservation, 56
register, approach, 185–189
retention, 190
risk mitigation plan, 61–62
risk mitigation plan, auditing, 65
survey questions, 190–191
tracking, 185–189
Information governance (IG), 25, 94–96, 115, 142
accountability, importance, 34
approach, customization, 212
artificial intelligence, role, 363–366
awareness, 164
best practices, 69, 78–80, 91–92, 223–225
business case, 212
business conditions, survey, 76–77
challenges, 372–375
common terms, usage, 26
continuous improvement, 33–34
debt, increase, 382
decisions, 30
defining, 7–9
development, 4–5
economic environment, survey, 76–77
E-discovery, relationship, 143–146
efforts, standards (relevance), 93–98

enforcement, 255–256
executive sponsor role, importance, 70–71
executive sponsorship, 33
external factors, survey/evaluation, 75–81
failures, 11–14
formal IG Program Charter, creation, 78, 92
framework, best practices/standards selections (impact), 105
framework/maturity model, usage, 79, 92
functions, 211, 229
good business, 9–10
guidelines, 351–352, 434
IG-enabled organization, appearance, 130–132
impact, 135
imperative, 3
importance, 361
industry best practices, survey/determination, 77–80
input, gaining, 83
IoT trustworthiness, contrast, 384
issues, 376–377
IT governance/data governance, differences, 19
legal functions, 135
legal hold process, 144
legal/regulatory/political factors, analysis, 77
monitoring/accountability, 425–426
monitoring/auditing, 33
ongoing program, 9, 78–79
plan, organizational strategic plan (alignment), 73–75
policies, 14–15, 48–49, 80, 87, 92, 394
principles, 29
process, steps, 128t
proof of concept, e-discovery (usage), 146
requirements, 165

RIM functions, relationship, 161
Sedona Conference® Commentary, 29–30, 87
Smallwood information governance principles, 30–34
strategic plan, 81–83
strategic planning, 69
strategy, information (synthesis/fusion), 81–82
team, 72–73
threats/concerns, 348–349
usage, 285, 299, 319, 335
Information Governance for Healthcare Professionals, 118
Information Governance Framework (IGF), 380
Information Governance Process Maturity Model (IGPMM), 79, 89, 119

Information governance (IG) program, 3
analytics, role, 360
auditing, effectiveness, 80, 92
best practices, 92
business considerations, 119
conformance/performance, measurement, 80
creation, 82–83
efforts, 211
elements, 80
establishment, 29
guidelines, 30–34
impact, 25–26
implementation, 29, 142
independence, 29
information stakeholder participation, 29
initiation, 118
maintenance, 425
objectives, 115, 116
organizational strategy, development, 78
piloting, business unit candidates, 117
SAT, impact, 249
strategic objectives, 29
Information Governance Reference Model (IGRM), 88–91, 89f
Center, 90–91
diagram, interpretation, 90–91
Outer Ring, 90
Information life cycle management (ILM), 217
Information rights management (IRM), 14, 62t, 265–268
characteristics, 266–268
DLP, combination, 270
software, 14, 32, 64, 265
technology, 80, 92, 342

Information risk
management, summary, 65
planning, summary, 65
reduction, 116
Information risk planning process, 56–59
steps, 56–59
vulnerabilities/threats (identification), formal process (conducting), 56–58
Information security (InfoSec), 32, 94–96
assessments/awareness training, 248–253
improvement/optimization, 80
incident management, 95
management, 95
organization, 95
penetration testing (pen test), 252–253
principles, 45–48
Information Security NHS Code of Practice, 245
Information technology (IT), 115
changes, 162
concerns, 287
department/provider, dependence, 165
functions, 211
governance, 19, 21–25, 220–223
infrastructure, deployment, 336
network diagram, 176
platform, migration, 156
processes, control objectives (mapping), 221
security practices, assessment, 116
trends, analysis, 75–76
Information Technology Infrastructure Library (ITIL), 22, 24, 222–223
Inherited metadata, 438
Innovation, security (contrast), 328–330
In-person interviews, conducting, 179

Insider threats, 247–248, 343
Instant messaging (IM), 63, 291–292
advice, 294
attachments, 294
business IM use, best practices, 292–293
contacts, restriction, 294
information governance, usage, 285
messages, archiving, 293
monitoring, technology (usage), 293
policies, enforcement, 294
screen names, privacy, 294
social media, contrast, 305
Institute for Security and Open Methodologies (ISECOM), 251–252
Intangible benefits (ERM), 167–168
Integration, data modeling steps, 218f
Integrity (GAR Principle component), 35, 39, 88
Intellectual entities, 404
Intellectual property (IP), 153, 247
Internal accounting system, creation, 130
Internal price lists, securing, 272
Internal Revenue Service (IRS), Elastic Compute Cloud service (usage), 348
International Association of Privacy Professionals (IAPP), 45–46, 239
International Council on Archives (ICA), 96
International Data Corporation (IDC), 151
International metadata standards/guidance, 442–446
International Organization of Standardization (ISO), 93, 161, 394, 449
Internet-based applications, integration, 304
Internet-based networks, usage, 302

Internet of Things (IoT), 75, 231, 357
challenges, 379–380
contracts system, 375
cybersecurity, 376
data, 381–382
definitions, 373
Detect Derive Decide Do (4 Ds), 380
devices, 325, 378
E-discovery issues, 377–380
growth, 373
information governance (IG), 372–377
privacy, 376
publicity, 376
risks, 376–377
states, 383
trustworthiness, 380–386
Internet of Things European Research Cluster (IERC), IoT definition, 373
Interoperability support, 93
Interviewees, selection, 181
Interviews, 182, 182f–183f
Inventory
conducting, 179–182
goals, definition, 173
information/elements, collection, 175–176
management support, executive sponsor, 175
project, goals, 173–174
results, analysis/verification, 183
scope, defining, 173
scoping, 174–175
surveys, distribution/collection, 179
Irish Data Protection Commission (DPC), 238
ISO13008:2012 (Information and documentation), 98

ISO 14721:2012 (Space Data and Information Transfer Systems), 103
ISO14721:2012 “Space Data and Information Transfer Systems,” 103
ISO 14721/ISO 16363, 413–415
ISO 15489 Records Management, 102
  definitions/relevance, 442
ISO 16175, 97
ISO 16363:2012 (Space Data and Information Transfer Systems), 103, 401–403
ISO 19005-1:2005 “Document Management,” 102
ISO 19005 (PDF/A) Document Management, 406–407
ISO 22301:2012 (Societal Security), 104
ISO 28500: 2009—WebARChive, 409
ISO 30300:2011 (Information and Documentation), 97
ISO 38500, 24–25, 95–96, 223
ISO/IEC 13818: 2000 Motion Picture Expert Group, 409
ISO/IEC 15444:2000—Joint Photographic Engineers Group, 408
ISO/IEC 15498: 2004 Information Technology-Computer Graphics, 408
ISO/IEC20000, 222
ISO/IEC 27001:2013, 94–95
ISO Technical Specification 23081-1:2006 Information and Documentation, 442–443
ISO/TR 13028:2010 (Information and documentation), 98
ISO/TR 18128:2014, 94
ISO TR 18492 (2005) (Long-Term Preservation of Electronic Document Based Information), 103, 400–401

Jailbreaking, 326
Joint Photographic Engineers Group (JPEG), 400, 408

Kelly, John, 328
Keyword search capabilities, 148
Knorr, Eric, 337
Knowledge
  capture/transfer, 126
  leveraging, 362–363

Knowledge Exchange (KX), 13–14
Knowledge management (KM), 72, 131, 445
Koirtchinksy, Kostya, 346

Laney, Doug, 4, 30, 117, 214, 357
Laptops, IG (usage), 326
Layoffs, trade secrets (securing), 270–271
Legacy electronic records, acquisition, 418
Legal costs, 124
Legal hold notification (LHN), 31–33
  initiation process, 144–145
  process, 80, 92, 130
Legal hold process, 144
Legal hold programs, building, 146
Legal review team, 150
Leming, Briton Reynold, 186, 189
Leverage technology, 143
Library of Congress subject headings (LCSH), 449, 450f
Lightweight directory access protocol (LDAP), 268
LinkedIn, social media site, 300, 309
Linux Foundation, 369
Litigation, 117, 153
Logan, Debra, 34, 211
Logical data modeling approach, 219
Long-term archival records, 200
Long-term business planning, 402
Long-Term Digital Preservation (LTDP), 7, 32, 39, 102–104
  defining, 393–394
  factors, 394–396
  methods, 162
  policies, 204
Long-Term Digital Preservation Capability Maturity Model (CMM)
  scope, 412–416

Long-Term Digital Preservation Capability Maturity Model (DPCMM), 409–412
Long-term evolution (LTE) (mobile computing trend), 322
Lossy compression, irreversibility, 408

Machine learning, 74, 147
  predictive coding component, 149
Madrid Resolution 2009, 234–235
Malicious insider, 247
Malware, impact, 344
Management (GAPP criterion), 46
Management location, 120
Management systems for records (MSR), 97
Manning, Bradley (Chelsea), 307
Many-to-many associations, 437
Marketing professionals, tools (usage), 300t–301t
Masking, deployment, 225
Master Data Management (MDM), 211, 213, 216–217, 225
Meaning-based search, usage, 148
Meaningful use, benefits, 365
Media renewal, 400
Medicare reimbursements, 365
Metadata, 433, 436
  application, 438f
  cross-referencing, 454f
  governance, implementation/maintenance, 441
  governance/standards/strategies, 438–440
  issues, 441–442
  management, 79–80, 439
  schema, implementation, 116
  taxonomy, relationship, 437–438
  terms, creation, 79
  types, 440–441

Metrics
access, 124
determination, 63–64
empirical metrics, 401
establishment, 55
requirement, 80
usage, 116
variation, 63–64
Micro-classification, avoidance, 185
Micro-electromechanical systems (MEMS), 372, 374
Microsoft Active Directory (AD), 268
Microsoft Cognitive Services, 363
Microsoft Windows Office Desktop, security, 257–258
Migration, 126, 418
refreshment/replication/repackage/transformation, 399
Mission-critical process protection, 104
Mission statement, 401
Mitigation efforts (risk reduction measurement), metrics (determination), 63–64
Mobile applications
innovation/security, contrast, 328–330
securing, best practices (usage), 330
security, building, 326–329
threats, understanding, 327–328
Mobile application vetting (MAV), 350
Mobile Companion Guide, release, 349
Mobile computing, 165
information governance (IG), 325–326
security risks, 323–324
trends, 322–323
Mobile data, securing, 324
Mobile device management (MDM), 321, 324–325

Mobile devices
budgeting/expense control, 331
communications/training plan, development, 332
device/data security issues, 331–332
information governance, usage, 319
legal aspects/liability issues, 331
policies, development, 330–332
policy requirement details, 331
Mobile guide, CIS controls, 349–350
Mobile strategy, goals (clarification), 331
Mobile threat defense (MTD), 350
Model Contract Clauses (MCCs), 478
Model Requirements for Management of Electronic Records (MoReq2), 101
Monetize (Navy Yard rebirth phase), 127, 128–129
Monitoring/auditing tools, deployment, 224–225
Monitoring/enforcement (GAPP criterion), 48
MoReq2010, MoReq2 requirements unbundling, 101
Motion Picture Expert Group (MPEG-2), 409
Multitenancy, issues, 345–346

National Archives and Records Administration (NARA), 93, 98, 168, 194–195, 465
guidelines, 351
Office of the Federal Register (OFR), 201
social media strategy, 302
National Association of Realtors, 369
National Association of Securities Dealers (NASD), 266, 463
National Cultural AudioVisual Archives (NCAA), 420
National Institute of Standards and Technology (NIST), 93, 336, 339
Cybersecurity Framework, 250
National Security Agency (NSA), 246
Natural disaster, impact, 396
Navigation, importance, 435
Naylor, David, 363

Near-preservation-ready formats, electronic records (acquisition/transformation), 417–418
Neighbors, knowledge, 347–348
Nerney, Chris, 306
Netbooks, IG (usage), 326
Net neutrality, 231
Network infrastructure, storage (relationship), 126
NextRequest, 311
NHS Care Record Guarantee for England, 245
Noncompliance fines/sanctions, 57
Non-Invasive Data Governance (Seiner), 19
Nonmalicious insider, 248
Non-record information, categorization/scheduling, 32–33
Notice (GAPP criterion), 46
Noto, Anthony, 307

Offensive Security Certified Profession (OSCP), 252–253
Office365, IG (usage), 352–354
Office for Civil Rights (OCR), 240
Office of the Australian Information Commissioner (OAIC), 470
On-demand self-service, 337
One-to-many associations, 437
Open access period, 469
Open Archival Information System (OAIS), 397
  core, functional model, 398
  Information Model, 398
  Reference Model, 397–398, 399f
Open Archival Initiative Protocol for Metadata Harvesting (OAI-PMH), 443
Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), 443
Openness principle, 234
Open-Source Security Testing Methodology Manual, 251–252

Open standard technology-neutral formats, 400, 405–409, 406t, 414
adoption, 417
migration, 418–419
Open Web Application Security Project (OWASP), 252
Operating systems (OS), 319, 346
Operational technology (OT), 374
Operations management, 95
Opt-in/opt-out jurisdiction, 230
Organisation for Economic Co-operation and Development (OECD) privacy principles, 233–234
Organizational "defense in depth," 56
Organizational goals/objectives (support), actionable plans (development), 82
Organizational infrastructure, 401–402
Organizational strategic plans, IG plan (alignment), 73–75
Organizational strategy, development, 78
Organizations
information glut, 153
reputation/brand/equity value, protection, 56
Originality, 187
Ownership, 120, 187

PageFreezer, cloud storage, 310
Password protection, 325
Pattern search, usage, 148
Payment Card Industry Data Secured Standard (PCI-DSS), 270
Peck, Andrew, 155
Penetration testing (pen test), 252–253
Perimeter security, limitations, 253–254
Personal archiving, 80, 92, 289
Personal data, usage, 186–187, 235–236
Personal digital assistants (PDAs), 337
Personal information, divulging, 294

Personally identifiable information (PII), 12, 30, 190, 229–230, 243
data deletion, 379
handling, 63
protection, 244
storage, 349
Phishing, 285
Physical data modeling approach, 219
Physical security, 95
Planning/control (executive sponsor purpose), 70
Policy
creation/management, 267
management, compliance, 126
Portable computers, IG (usage), 326
Portable Document Format (PDF), 407
Portable storage devices, IG (usage), 326
Prebuilt taxonomies, custom taxonomies (contrast), 448–449
Predictive analytics, 357, 358
Predictive coding, 74, 135
components, 148–149
impact, 149
usage, 147–149
Prescriptive analytics, 357, 359
Preservation
metadata, 416, 419, 441
planning, 398
policies, 402
preservation-ready formats, electronic records (acquisition), 417
strategic plan, 401
Preservation description information (PDI), 398, 399

PREservation Metadata Information Strategies (PREMIS)
Data Model, 404f
PREMIS-based data dictionary, 416
Preservation Metadata Standard, 404–405
Primary Trustworthy Digital Repository Authorisation Body (PTAB), 420
Prime directive, 466
Principle of Least Privilege (POLP), 45
Privacy, 45–48, 56, 229–232, 478
Asia, privacy, 243–244
infonomics, relationship, 244
information privacy, 32, 229–231
IoT, 376
laws, 244–245, 475, 478–480
OECD privacy principles, 233–234
organizational "defense in depth," 56
process, accountability, 56
programs, 239
regulations, 475
requirements, measurement/enforcement, 56
United States privacy, 240–244
Privacy Act, 470
Privacy Awareness Training (PAT), 107, 130
Privacy compliance, 10
Private cloud, 339
Privileged access, verification, 225
Probability, determination, 55
Productivity gains/losses, 124
Program communications/training, 106–107
Program controls, monitoring/auditing/enforcement, 107
Programs/service staff, interview, 181
Progress (measurement), metrics (usage), 116
Project Management Book of Knowledge (PMBOK), 23

PRONON program, 406
Propagation-based mechanism, 150
Protected health information (PHI), 30, 230, 240, 324
Protected process/data, data security design pattern, 273
Protection (GAR Principle component), 35, 39, 88
Provenance, 189
Public cloud, 339
Public health initiatives, AI contribution, 365
Public key infrastructure (PKI), usage, 368
Public Record Office (PRO), 101
Public Records Act, 245
Public social media solutions, 309–310
Purge (disposition method), 42
Purpose specification principle, 233
Push-button applications (mobile computing trend), 323

Quality (GAPP criterion), 47
Quality assurance support, 93
Quest Software, 292

Radio frequency identification (RFID), 322
Rallo, Artemi, 235
Ransomware, 56–57, 214
Ransomware-as-a-service, 57
Really simple syndication (RSS), 299

Records
appraisal, 40, 184
classification, 195, 447
creation/growth, control, 167
creators/owners, 413
destruction, proving, 205
disposal, 204–205
documents, contrast, 442
findability, improvement, 80
grouping rationale, 196, 446
identification, 315
integrity, 39
inventory/classification, 195–196
location, determination, 179
preservation, threats, 396–397
producers/stakeholders, proactive engagement, 419
provenance, 195
record-free e-mail, usage, 260
scheduling, decisions, 195–196
series, 196–197
storage, centralized policies/systems (absence), 153
taxonomy, 447
threshold determinations, 314
value, appraisal, 184
Records and information management (RIM), 31, 35, 77
big bucket approach, 170
functions, IG (relationship), 161
perspective, 152

Records inventory
information, presence, 184
process, steps, 173
purposes, 172–173
steps, 173–183
survey form, 176f–177f
creation, 176–178
Records management (RM), 96–98, 351
business rationale, 163–164
challenges, 165–166, 305
data privacy management, relationship, 169–171
functionality, 316, 349
issues, 335
laws/regulations, 463
micro-classification, avoidance, 185
policies/procedures, access, 38
policy, adoption/compliance, 184–189
processes, 88
programs, 35, 88
requirements, identification, 471–472
space, 161
standards, 101–102
strategy, 153
technologies, assimilation, 167
Records Management NHS Code of Practice, 245
Records retention
guidelines, 314–315
policy, creation/implementation, 142–143
regulations/legislation, impact, 468

Records Retention Schedule (RRS), 31, 40, 83, 365, 447
development, 168, 192
steps, 194–195
IG program element, 80
mapping, 447f
sample, 194f
Redundant, outdated, and trivial (ROT)
data storage, 361
information, 4, 5, 76
records, 377
Reference data management (data modeling approach), 219
Relationships, understanding, 186
Repository-based approaches, limitations, 256–257
Resource pooling, 337–338
Responsibility assignment matrix (RACI matrix), 105–106
Responsible, accountable, consulted, informed (RACI) matrix, 64
Retention, 162, 188
automation, 205
capabilities, 203
compliance research, 200–201
event-based retention scheduling, 201–202
GAR Principle component, 36, 40, 88
legal limitation periods, 200
legal requirements, 200–201
periods, 203–204
policy, extension, 143
records retention schedule, development, 192
schedule, 204–207, 315
scheduling, principles, 191–192
Retention schedules, 193–195
Return on investment (ROI), 10, 146, 157, 304, 306

Risk
adverse findings, reduction, 63
assessment, 62t, 65, 214–215
avoidance, 315
events, materialization (likelihood), 58–59
identification, 55, 65
levels, evaluation, 65
management, 10, 94
map, 60
mitigation, 61–62, 64–65
probabilities, evaluation, 65
profile, creation, 59–61
recognition, 145
reductions (measurement), metrics (determination), 63–64
Rogue intrusions, 344
Rooting, 326
Rules-driven mechanism, 149
Rutowska, Joanna, 346

Sampling, 148
predictive coding component, 149
Sarbanes-Oxley Act (SOX), 163–164, 289, 464
Scalable Vector Graphics (SVG): 2003—W3C Internet Engineering Task Force, 408
Scotland, Parliament (acts), 468
Search capabilities, 316
Search results (improvement), taxonomies (impact), 436–437
Secure communications, record-free e-mail (usage), 260
Secure printing, 258
Secure sockets layer (SSL), 325
Securities and Exchange Commission (SEC)
Rule 17A-4, 464
rules, 463

Security
benefits, 166
building, 326–329
classifications, 186
concerns, 165
e-document security, provision, 80
enterprise information security practices, direct connection, 56
functions, 229
information security, 32
innovation, contrast, 328–330
issues, 258–259, 378–379
mobile computing trend, 322–323
perimeter security, limitations, 253–254
policy, 95
processes, integration/automation, 225
risk management, 403
safeguards principle, 233
Security Awareness Training (SAT), 31, 76, 82
program, 116, 248
provision, 63
sessions, 320
success, 249
Security Information Event Monitoring (SIEM), usage, 249
Security Technical Implementation Guides (STIGs), 224
Sedona Conference®, 29–30, 87, 313
Seiner, Robert, 19
Senior records officer (SRO), 72–73
Sensitive data, encryption, 330
Service-level agreements (SLAs), 346
Service-oriented architecture, 338
SHA128 hashing process, 354
SharePoint, capabilities, 275

SharePoint, IG, 352–354
Sharing economy networks, usage, 302
Short Message Service (SMS), 320
Short-term business planning, 402
Shred (disposition method), 41
Smallwood information governance principles, 30–34
Smarsh, 310–311
Smartphones
  applications, 322
  IG, usage, 325–326
Smishing, 320
SnapChat applications, 377
Snowden, Edward, 246, 307
Social Care Record Guarantee for England, 245, 276
Social engineering, 321

Social media, 165
archiving, 309–311, 310t
categories, 300t–301t, 303
e-mail/IM, contrast, 305
IG framework, 311
information governance, 299, 311
litigation considerations, 313–315
management software, 310t
planning stage, records (identification), 315
policy, 306, 312
posts, legal risks, 307–308
publication process, establishment, 315
public social media solutions, 309–310
records management, 313–316
risks, 306306–
technology, leverage, 304–305
tools, usage, 302t, 303t
types, 299–302
usage, 304–305
Social shopping networks, usage, 301
Social tagging, 434, 458
Software development life cycle (SDLC), 330
Software tools, usage, 180
Solid state drives (SSDs), 396
mobile computing trend, 322
Space planning, 188
Spam, problem, 344
SQL databases, usage, 268

Stakeholder
confidence, 104
consultation, 31
input, gaining, 83
participation, 29
Standard General Markup Language (SGML), 407
Standard industry classification (SIC) codes, 219
Standards
benefits/risks, 93
confusion, 93
considerations, 92–93
development/promotion, benefits, 93
downside considerations, 93
European standards, 101–102
UK standards, 101–102
Standards Council of Canada, 93
State government regulatory agency functional taxonomy, 453f
Steering committee, establishment, 78, 92
Storage
costs, increase, 153
encryption, 325
media, failure, 396
network infrastructure, relationship, 126
Stored Communications and Transaction Records Act (SCTRA), 244–245
Strategic plans, development, 63
Stream messaging, 276
Structural metadata, 440
Structured data, 5
unstructured information, contrast, 360
Subject matter, asset tag, 188
Subject matter experts (SMEs), usage, 436, 441
Subject records, 197

Submission information package (SIP), 398–403
Systems/network communications failure, 396

Tablets, IG (usage), 326
Tagged image file format (TIFF), 400, 407
Taxonomy (taxonomies), 162, 433, 446–447
  accounting business-unit taxonomy, 454f
  community government business-unit taxonomy, 452f
  faceted taxonomies, 452
  maintenance, 457–458
  metadata, 437–438, 437f, 454f
  necessity, 435–436
  prebuilt taxonomies, custom taxonomies (contrast), 448–449
  records grouping, 446
  records retention schedule, mapping, 447f
  state government regulatory agency functional taxonomy, 453f
  structure, metadata (application), 438f
  testing, 457
  thesaurus use, 449
  types, 449–453
Team messaging, solutions, 294–295
Technical infrastructure, 403
Technical metadata, 440
Technology
  application, 257–259
  change, 430
  emerging technologies, leveraging/governing, 357
  focus, 212
  importance, 155
  sharing, issues, 345–346
  usage, 155–156

Technology-assisted review (TAR), 82, 135, 143, 151
mechanisms, 149–150
exercise, 150
Teisch, Rachel, 376
Terminations, trade secret (securing), 270–271
Territorial scope, increase, 476
Text analytics, 148–149
Text mining, 434, 445–446
The National Archives (TNA), 101
Theoretical basis, impact, 93
Thesaurus, 438
Thin clients, 269, 273
Thin device, data security design pattern, 273
Third-party disclosure (GAPP criterion), 47
Third-party possession, 273
Threat identification/assessment, 104
Threat, recovery planning, 104
Topic records, 197
Total cost of ownership (TCO), 123–124
Trade secrets, securing, 270–271
Transaction, e-mail documentation, 198
Transitory records, retention, 204
Transparency (GAR Principle component), 35, 38, 88
Trigger events, 202–203
Trump, Ivanka (personal e-mail account), 288
Twitter, social media site, 300, 309

Unified end-user commuting management (UEM) platforms, 325
United Kingdom
Parliament, statutory instruments, 468–469
records management, laws/regulations, 468–469

United States
Department of Defense 5015.2 *Design Criteria Standard for Electronic Records Management Software Application*, 98
Government Protection Profile for Authorization Server for Basic Robustness Environments, 256
privacy, 240–244, 475–476
records management, laws/regulations, 463
regulations, 242
United States National Archives (NARA), social media tools (usage), 302t
Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act (PATRIOT Act), 464
Unmanaged devices, 350
Unstructured information, 5
challenges/factors, 120
location, 124
ownership, cost (calculation), 124–126
structured data, contrast, 360
volume, 119
Up-front costs, 124
Use limitation principle, 233
User assistance/compliance, 165–166
Use/retention/disposal (GAPP criterion), 46–47
User testing, feedback (usage), 457

Val IT®, 23, 222
Value information, asset status, 30
Virtual private network (VPN), 325, 331
software, 323
Visio, usage, 455
Vulnerabilities/threats (identification), formal process (conducting), 56–58
Vulnerability assessment, 251

Warwick Business School, 304
Web 2.0, 165
social media, types, 299–302

WebARChive (WARC), 409

Web sites/applications, 299

Whole disk encryption (WDE), 326

Winner, Reality Leigh, 307

Workflow, 148–149

Workplace, personal archiving (absence), 289

Worldwide Interoperability for Microwave Access (WiMax) (mobile computing trend), 322

World Wide Web Consortium (W3C) Internet Engineering Group, 407

Written records management plan, 173

Yellow Pages, text (example), 449–450, 451f

Zornes, Aaron, 211

Zubulake, Laura, 139

Zubulake v. UBS Warburg, 139

Zuckerberg, Mark, 238

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.