

Data: Business Intelligence

6

CHAPTER

SECTION 6.1

Data, Information, and Databases

- The Business Benefits of High-Quality Information
- Storing Information Using a Relational Database Management System
- Using a Relational Database for Business Advantages
- Driving Websites with Data

SECTION 6.2

Business Intelligence

- Supporting Decisions with Business Intelligence
- The Business Benefits of Data Warehousing
- The Power of Big Data Analytics

CHAPTER OUTLINE

What's in IT for me?

This chapter introduces the concepts of information and data and their relative importance to business professionals and firms. It distinguishes between data stored in transactional databases and powerful business intelligence gleaned from data warehouses. Students who understand how to access, manipulate, summarize, sort, and analyze data to support decision making find success. Information has power, and understanding that power will help you compete in the global marketplace. This chapter will provide you with an overview of database fundamentals and the characteristics associated with high-quality information. It will also explain how the various bits of data stored across multiple, operational databases can be transformed in a centralized repository of summarized information in a data warehouse, which can be used for discovering business intelligence.

You, as a business student, need to understand the differences between transactional data and summarized information and the different types of questions you could use a transactional database to answer versus a data warehouse. You need to be aware of the complexity of storing data in databases and the level of effort required to transform operational data into meaningful, summarized information. You need to realize the power of information and the competitive advantage a data warehouse brings an organization in terms of facilitating business intelligence. Armed with the power of information, business students will make smart, informed, and data-supported managerial decisions.



Informing Information

Since the beginning of time, man has been using pictures and images to communicate, moving from caveman drawings to hieroglyphics to the Internet. Today, it is easier than ever to paint a picture worth 100,000 words, thanks to technological advances. The primary advantages are databases and data warehouses that capture enormous amounts of data. Informing means accessing large amounts of data from different management information systems. According to a recent analysis of press releases by *PR Newswire*, an article or advertisement that uses visual images can significantly improve the number of views a message generates. This can be a true competitive advantage in the digital age.

An infographic (or information graphic) displays information graphically so it can be more easily understood. Infographics cut straight to the point by presenting complex information in a simple visual format. Infographics can present the results of large data analysis, looking for patterns and relationships that monitor changes in variables over time. Because infographics can easily become overwhelming, users need to be careful not to display too much data or the resulting infographics can result in information overload. Effective infographics can achieve outstanding results for marketing, advertising, and public relations. According to *PR Newswire*, infographics gain the greatest competitive advantage when they have the following:

- ❖ Survey results that are too hard to understand in text format.
- ❖ Statistical data that are not interesting for readers.
- ❖ Comparison research where the impact can be far more dramatic when presented visually.
- ❖ Messages for multilingual audiences.
- ❖ Any information that can use a visual element to make it more interesting (see Figures 6.1, 6.2, and 6.3 for examples).¹

FIGURE 6.1

Hotels.com Travel Infographic

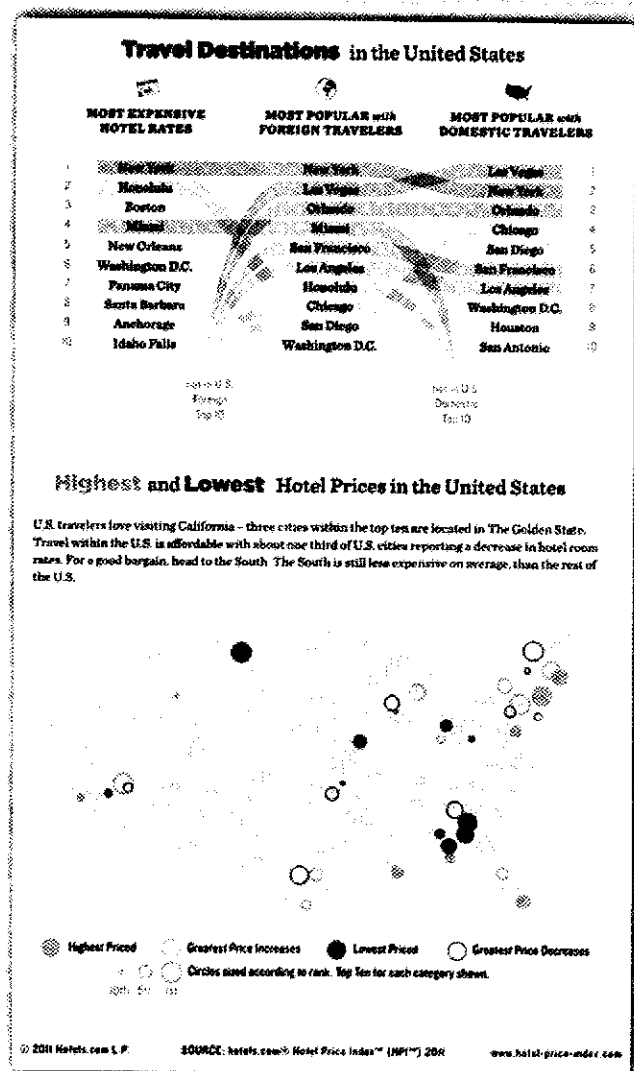
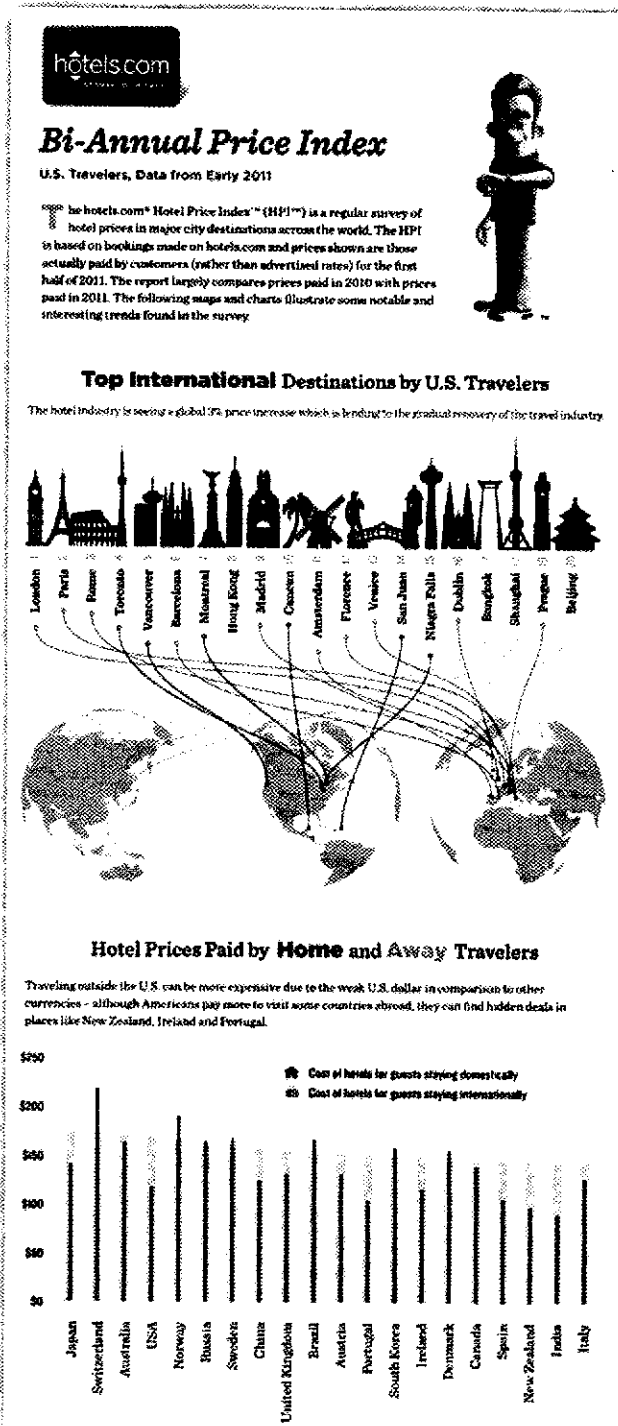


FIGURE 6.2

Emerson's Food Waste
Infographic

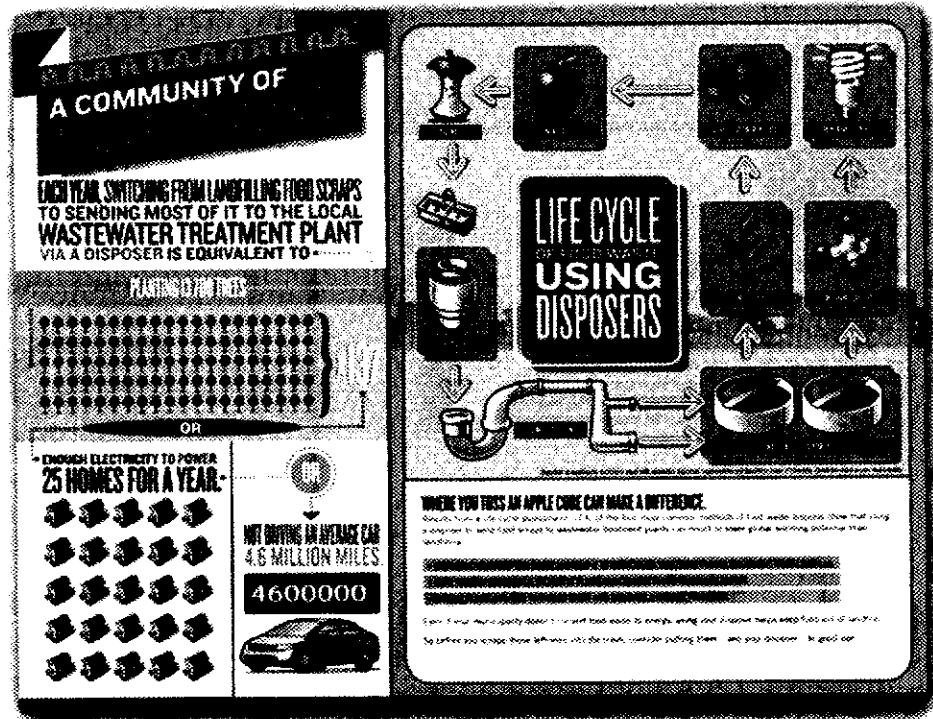


FIGURE 6.3

Changes in Consumer Action
on the Environment



2011

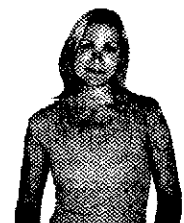
Changes in Consumer Action on the Environment

Nearly 75% of Americans believe that a manufacturer who takes steps to reduce their impact on the environment is making a good business decision.

1995–2011

In 1995, less than 10% of Americans said they were familiar with environmental issues and concerns. In 2011, that number is now nearly 70%.

Almost half of Americans are unsure of how to address these problems in the future.



In the last 20 years, the number of Americans who say they recycle has doubled and nearly 60% say they do so on a regular basis.

LEARNING OUTCOMES

- 6.1 Explain the four primary traits that determine the value of information.
- 6.2 Describe a database, a database management system, and the relational database model.
- 6.3 Identify the business advantages of a relational database.
- 6.4 Explain the business benefits of a data-driven website.

THE BUSINESS BENEFITS OF HIGH-QUALITY INFORMATION

Information is powerful. Information can tell an organization how its current operations are performing and help it estimate and strategize about how future operations might perform. The ability to understand, digest, analyze, and filter information is key to growth and success for any professional in any industry. Remember that new perspectives and opportunities can open up when you have the right data that you can turn into information and ultimately business intelligence.

Information is everywhere in an organization. Managers in sales, marketing, human resources, and management need information to run their departments and make daily decisions. When addressing a significant business issue, employees must be able to obtain and analyze all the relevant information so they can make the best decision possible. Information comes at different levels, formats, and granularities. **Information granularity** refers to the extent of detail within the information (fine and detailed or coarse and abstract). Employees must be able to correlate the different levels, formats, and granularities of information when making decisions. For example, a company might be collecting information from various suppliers to make needed decisions, only to find that the information is in different levels, formats, and granularities. One supplier might send detailed information in a spreadsheet, whereas another supplier might send summary information in a Word document, and still another might send a collection of information from emails. Employees will need to compare these differing types of information for what they commonly reveal to make strategic decisions. Figure 6.4 displays the various levels, formats, and granularities of organizational information.

Successfully collecting, compiling, sorting, and finally analyzing information from multiple levels, in varied formats, and exhibiting different granularities can provide tremendous insight into how an organization is performing. Exciting and unexpected results can include potential new markets, new ways of reaching customers, and even new methods of doing business. After understanding the different levels, formats, and granularities of information, managers next want to look at the four primary traits that help determine the value of information (see Figure 6.5).

Information Type: Transactional and Analytical

As discussed previously in the text, the two primary types of information are transactional and analytical. Transactional information encompasses all of the information contained within a single business process or unit of work, and its primary purpose is to support daily operational tasks. Organizations need to capture and store transactional information to perform operational tasks and repetitive decisions such as analyzing daily sales reports and production schedules to determine how much inventory to carry. Consider Walmart, which handles more than 1 million customer transactions every hour, and Facebook, which keeps track of 400 million active users (along with their photos, friends, and web links). In addition, every time a cash register rings up a sale, a deposit or withdrawal is made from an ATM, or a receipt is given at the gas pump, the transactional information must be captured and stored.

LO 6.1: Explain the four primary traits that determine the value of information.

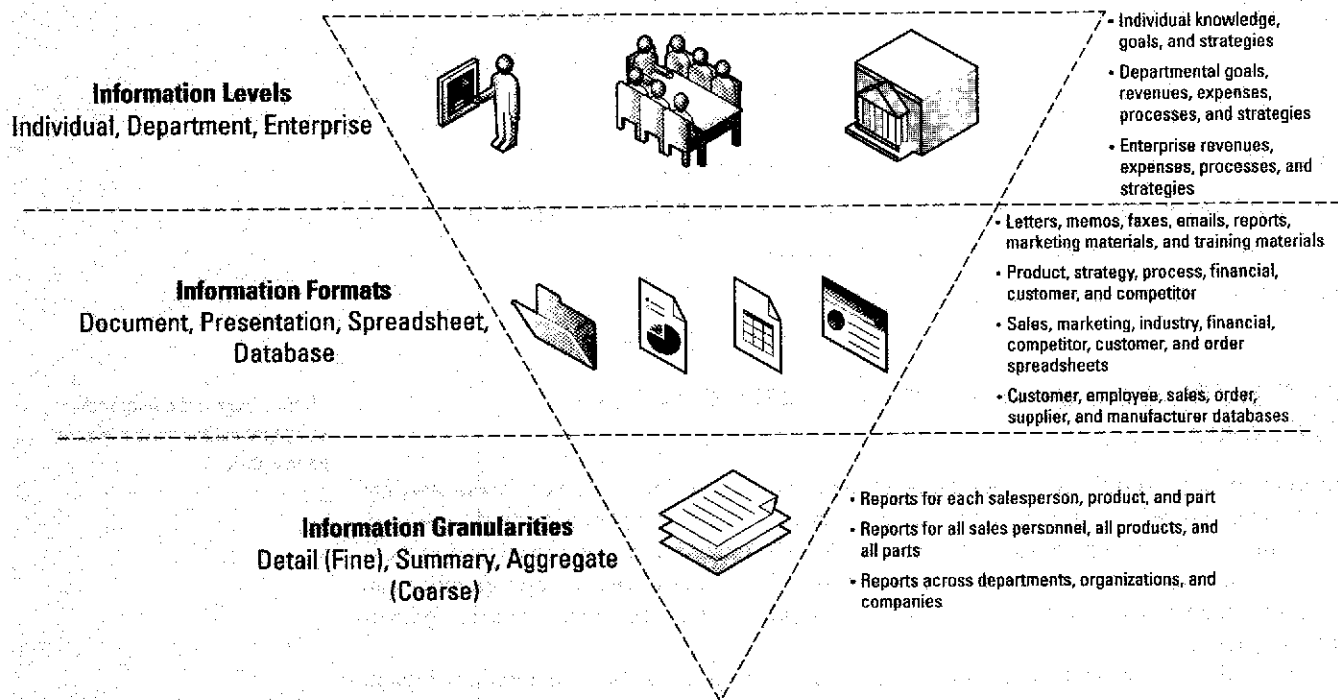


FIGURE 6.4

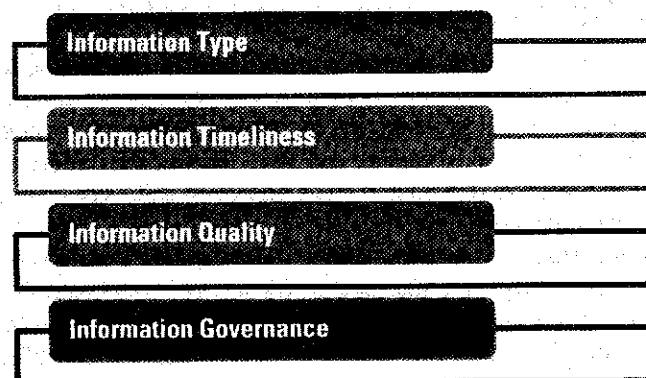
Levels, Formats, and Granularities of Organizational Information

Analytical information encompasses all organizational information, and its primary purpose is to support the performance of managerial analysis tasks. Analytical information is useful when making important decisions such as whether the organization should build a new manufacturing plant or hire additional sales personnel. Analytical information makes it possible to do many things that previously were difficult to accomplish, such as spot business trends, prevent diseases, and fight crime. For example, credit card companies crunch through billions of transactional purchase records to identify fraudulent activity. Indicators such as charges in a foreign country or consecutive purchases of gasoline send a red flag highlighting potential fraudulent activity.

Walmart was able to use its massive amount of analytical information to identify many unusual trends, such as a correlation between storms and Pop-Tarts. Yes, Walmart discovered an increase in the demand for Pop-Tarts during the storm season. Armed with that valuable

FIGURE 6.5

The Four Primary Traits of the Value of Information



information, the retail chain was able to stock up on Pop-Tarts that were ready for purchase when customers arrived. Figure 6.6 displays different types of transactional and analytical information.

Information Timeliness

Timeliness is an aspect of information that depends on the situation. In some firms or industries, information that is a few days or weeks old can be relevant, whereas in others information that is a few minutes old can be almost worthless. Some organizations, such as 911 response centers, stock traders, and banks, require up-to-the-second information. Other organizations, such as insurance and construction companies, require only daily or even weekly information.

Real-time information means immediate, up-to-date information. **Real-time systems** provide real-time information in response to requests. Many organizations use real-time systems to uncover key corporate transactional information. The growing demand for real-time information stems from organizations' need to make faster and more effective decisions, keep smaller inventories, operate more efficiently, and track performance more carefully. Information also needs to be timely in the sense that it meets employees' needs, but no more. If employees can absorb information only on an hourly or daily basis, there is no need to gather real-time information in smaller increments.

Most people request real-time information without understanding one of the biggest pitfalls associated with real-time information—continual change. Imagine the following scenario: Three managers meet at the end of the day to discuss a business problem. Each manager has gathered information at different times during the day to create a picture of the situation. Each manager's picture may be different because of the time differences. Their views on the business problem may not match because the information they are basing their analysis on is continually changing. This approach may not speed up decision making, and it may actually slow it down. Business decision makers must evaluate the timeliness of the information for every decision. Organizations do not want to find themselves using real-time information to make a bad decision faster.

Information Quality

Business decisions are only as good as the quality of the information used to make them. **Information inconsistency** occurs when the same data element has different values. Take for example the amount of work that needs to occur to update a customer who had changed her last name due to marriage. Changing this information in only a few organizational systems will lead to data inconsistencies causing customer 123456 to be associated with two last names. **Information integrity issues** occur when a system produces incorrect, inconsistent, or

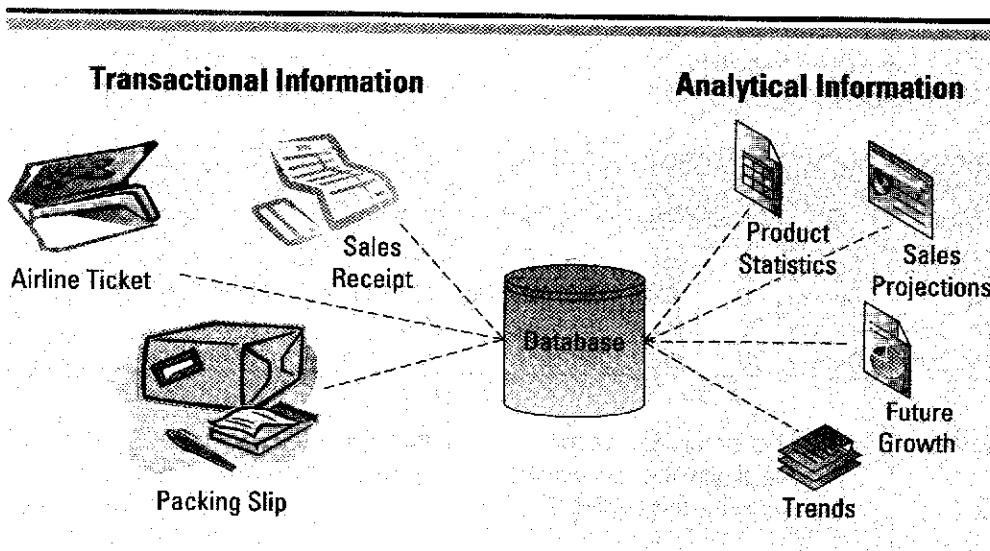
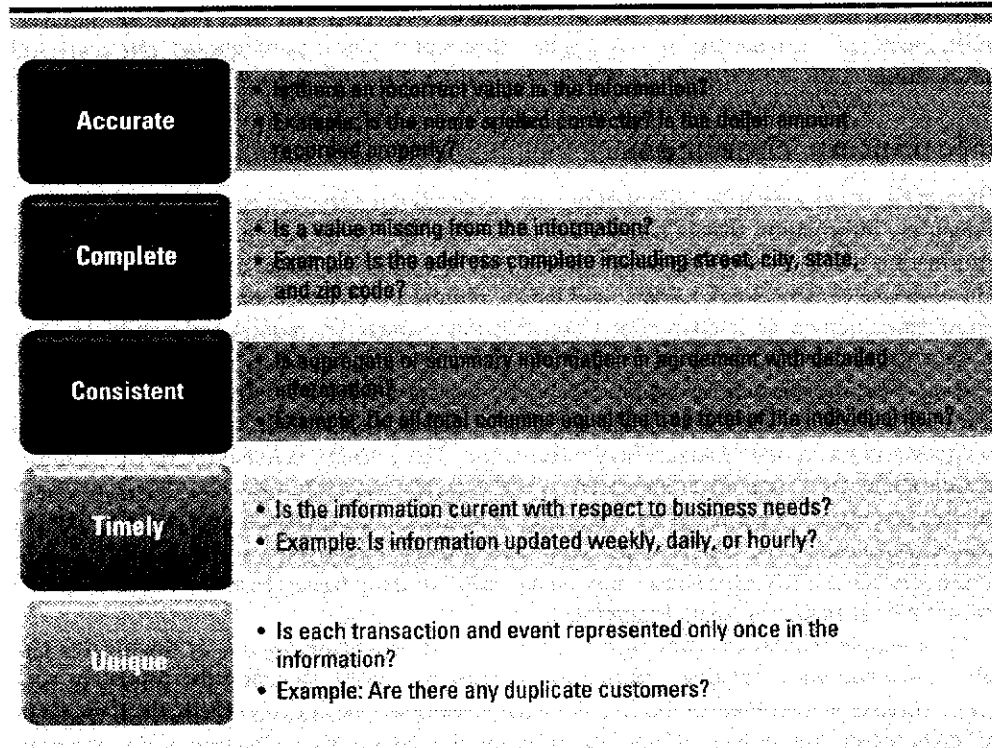


FIGURE 6.6

Transactional versus Analytical Information

FIGURE 6.7

Five Common Characteristics of High-Quality Information



duplicate data. Data integrity issues can cause managers to consider the system reports invalid and will make decisions based on other sources.

To ensure that your systems do not suffer from data integrity issues, review Figure 6.7 for the five characteristics common to high-quality information: accuracy, completeness, consistency, timeliness, and uniqueness. Figure 6.8 provides an example of several problems associated with using low-quality information, including:

1. *Completeness.* The customer's first name is missing.
2. Another issue with *completeness.* The street address contains only a number and not a street name.
3. *Consistency.* There may be a duplication of information since there is a slight difference between the two customers in the spelling of the last name. Similar street addresses and phone numbers make this likely.

FIGURE 6.8

Example of Low-Quality Information

1. Missing information (no first name)			2. Incomplete information (no street)				5. Inaccurate information (invalid email)		
ID	Last Name	First Name	Street	City	State	Zip	Phone	Fax	Email
113	Smith		123 S. Main	Denver	CO	80210	(303) 777-1258	(303) 777-5544	ssmith@aol.com
114	Jones	Jeff	12A	Denver	CO	80224	(303) 666-6868	(303) 666-6868	(303) 666-6868
115	Roberts	Jenny	1244 Colfax	Denver	CO	85231	759-5654	853-6584	jr@msn.com
116	Robert	Jenny	1244 Colfax	Denver	CO	85231	759-5654	853-6584	jr@msn.com
3. Probable duplicate information (similar names, same address, phone number)			4. Potential wrong information (are the phone and fax numbers the same or is this an error?)				6. Incomplete information (missing area codes)		

BUSINESS DRIVEN MIS

Real People magazine is geared toward working individuals and provides articles and advice on everything from car maintenance to family planning. The magazine is currently experiencing problems with its distribution list. More than 30 percent of the magazines mailed are returned because of incorrect address information, and each month it receives numerous calls from angry customers complaining that they have not yet received their magazines. Below is a sample of *Real People's* customer information. Create a report detailing all the issues with the information, potential causes of the information issues, and solutions the company can follow to correct the situation.

Determining Information Quality Issues

ID	First Name	Middle Initial	Last Name	Street	City	State	Zip Code
433	M	J	Jones	13 Denver	Denver	CO	87654
434	Margaret	J	Jones	13 First Ave.	Denver	CO	87654
434	Brian	F	Hoover	Lake Ave.	Columbus	OH	87654
435	Nick	H	Schweitzer	65 Apple Lane	San Francisco	OH	65664
436	Richard	A		567 55th St.	New York	CA	98763
437	Alana	B	Smith	121 Tenny Dr.	Buffalo	NY	142234
438	Trevor	D	Darrian	90 Fresdestil	Dallas	TX	74532

4. **Accuracy.** This may be inaccurate information because the customer's phone and fax numbers are the same. Some customers might have the same number for phone and fax, but the fact that the customer also has this number in the email address field is suspicious.
5. Another issue with **accuracy.** There is inaccurate information because a phone number is located in the email address field.
6. Another issue with **completeness.** The information is incomplete because there is not a valid area code for the phone and fax numbers.

Nestlé uses 550,000 suppliers to sell more than 100,000 products in 200 countries. However, due to poor information, the company was unable to evaluate its business effectively. After some analysis, it found that it had 9 million records of vendors, customers, and materials, half of which were duplicated, obsolete, inaccurate, or incomplete. The analysis discovered that some records abbreviated vendor names, and other records spelled out the vendor names. This created multiple accounts for the same customer, making it impossible to determine the true value of Nestlé's customers. Without being able to identify customer profitability, a company runs the risk of alienating its best customers.²

Knowing how low-quality information issues typically occur can help a company correct them. Addressing these errors will significantly improve the quality of company information and the value to be extracted from it. The four primary reasons for low-quality information are:

1. Online customers intentionally enter inaccurate information to protect their privacy.
2. Different systems have different information entry standards and formats.
3. Data-entry personnel enter abbreviated information to save time or erroneous information by accident.
4. Third-party and external information contains inconsistencies, inaccuracies, and errors.

Understanding the Costs of Using Low-Quality Information Using the wrong information can lead managers to make erroneous decisions. Erroneous decisions in turn can cost time, money, reputations, and even jobs. Some of the serious business consequences that occur due to using low-quality information to make decisions are:

- ✱ Inability to track customers accurately.
- ✱ Difficulty identifying the organization's most valuable customers.
- ✱ Inability to identify selling opportunities.
- ✱ Lost revenue opportunities from marketing to nonexistent customers.
- ✱ The cost of sending undeliverable mail.
- ✱ Difficulty tracking revenue because of inaccurate invoices.
- ✱ Inability to build strong relationships with customers.

Understanding the Benefits of Using High-Quality Information High-quality information can significantly improve the chances of making a good decision and directly increase an organization's bottom line. One company discovered that even with its large number of golf courses, Phoenix, Arizona, is not a good place to sell golf clubs. An analysis revealed that typical golfers in Phoenix are tourists and conventioners who usually bring their clubs with them. The analysis further revealed that two of the best places to sell golf clubs in the United States are Rochester, New York, and Detroit, Michigan. Equipped with this valuable information, the company was able to place its stores strategically and launch its marketing campaigns.

High-quality information does not automatically guarantee that every decision made is going to be a good one, because people ultimately make decisions and no one is perfect. However, such information ensures that the basis of the decisions is accurate. The success of the organization depends on appreciating and leveraging the true value of timely and high-quality information.

Information Governance

Information is a vital resource, and users need to be educated on what they can and cannot do with it. To ensure that a firm manages its information correctly, it will need special policies and procedures establishing rules on how the information is organized, updated, maintained, and accessed. Every firm, large and small, should create an information policy concerning data governance. **Data governance** refers to the overall management of the availability, usability, integrity, and security of company data. **Master data management (MDM)** is the practice of gathering data and ensuring that it is uniform, accurate, consistent, and complete, including such entities as customers, suppliers, products, sales, employees, and other critical entities that are commonly integrated across organizational systems. MDM is commonly included in data governance. A company that supports a data governance program has a defined a policy that specifies who is accountable for various portions or aspects of the data, including its accuracy, accessibility, consistency, timeliness, and completeness. The policy should clearly define the processes concerning how to store, archive, back up, and secure the data. In addition, the company should create a set of procedures identifying accessibility levels for employees. Then, the firm should deploy controls and procedures that enforce government regulations and compliance with mandates such as Sarbanes-Oxley.

LO 6.2: Describe a database, a database management system, and the relational database model.

STORING INFORMATION USING A RELATIONAL DATABASE MANAGEMENT SYSTEM

The core component of any system, regardless of size, is a database and a database management system. Broadly defined, a **database** maintains information about various types of objects (inventory), events (transactions), people (employees), and places (warehouses). A **database management system (DBMS)** creates, reads, updates, and deletes data in a database while controlling access and security. Managers send requests to the DBMS, and the DBMS performs the actual manipulation of the data in the database. Companies store their information in databases, and managers access these systems to answer operational questions such

BUSINESS DRIVEN DEBATE

Excel is a great tool with which to perform business analytics. Your friend, John Cross, owns a successful publishing company specializing in Do It Yourself books. John started the business 10 years ago and has slowly grown to 50 employees and \$1 million in sales. John has been using Excel to run the majority of his business, tracking book orders, production orders, shipping orders, and billing. John even uses Excel to track employee payroll and vacation dates. To date, Excel has done the job, but as the company continues to grow, the tool is becoming inadequate.

You believe John could benefit from moving from Excel to Access. John is skeptical of the change because Excel has done the job up to now, and his employees are comfortable with the current processes and technology. John has asked you to prepare a presentation explaining the limitations of Excel and the benefits of Access. In a group, prepare the presentation that will help convince John to make the switch.

Excel or Access?

as how many customers purchased Product A in December or what the average sales were by region. Two primary tools are available for retrieving information from a DBMS. First is a **query-by-example (QBE) tool** that helps users graphically design the answer to a question against a database. Second is a **structured query language (SQL)** that asks users to write lines of code to answer questions against a database. Managers typically interact with QBE tools, and MIS professionals have the skills required to code SQL. Figure 6.9 displays the relationship between a database, a DBMS, and a user. Some of the more popular examples of DBMS include MySQL, Microsoft Access, SQL Server, FileMaker, Oracle, and FoxPro.

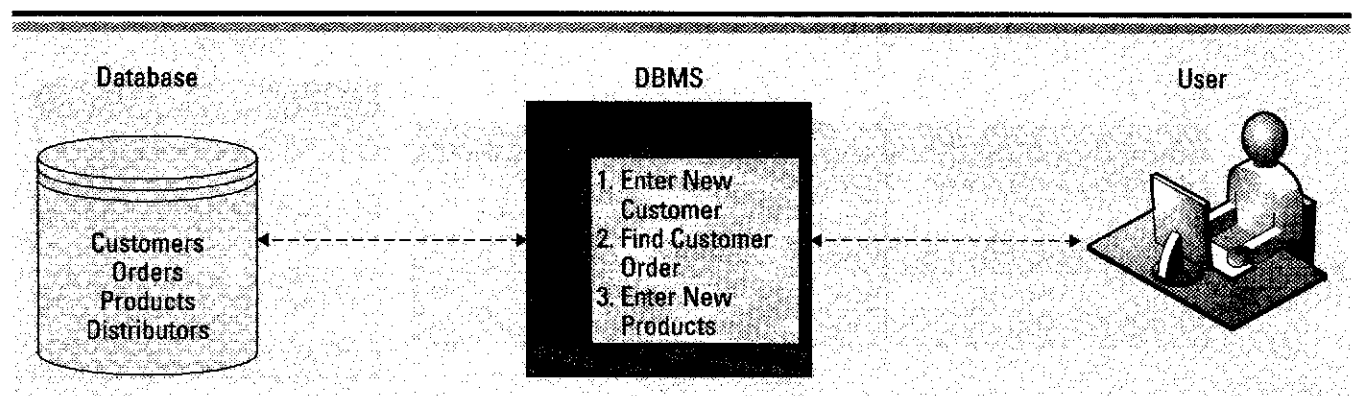
A **data element (or data field)** is the smallest or basic unit of information. Data elements can include a customer's name, address, email, discount rate, preferred shipping method, product name, quantity ordered, and so on. **Data models** are logical data structures that detail the relationships among data elements by using graphics or pictures.

Metadata provides details about data. For example, metadata for an image could include its size, resolution, and date created. Metadata about a text document could contain document length, data created, author's name, and summary. Each data element is given a description, such as Customer Name; metadata is provided for the type of data (text, numeric, alphanumeric, date, image, binary value) and descriptions of potential predefined values such as a certain area code; and finally the relationship is defined. A **data dictionary** compiles all of the metadata about the data elements in the data model. Looking at a data model along with reviewing the data dictionary provides tremendous insight into the database's functions, purpose, and business rules.

DBMS use three primary data models for organizing information—hierarchical, network, and the relational database, the most prevalent. A **relational database model** stores

FIGURE 6.9

Relationship of Database, DBMS, and User



information in the form of logically related two-dimensional tables. A **relational database management system** allows users to create, read, update, and delete data in a relational database. Although the hierarchical and network models are important, this text focuses only on the relational database model.

Storing Data Elements in Entities and Attributes

For flexibility in supporting business operations, managers need to query or search for the answers to business questions such as which artist sold the most albums during a certain month. The relationships in the relational database model help managers extract this information. Figure 6.10 illustrates the primary concepts of the relational database model—entities, attributes, keys, and relationships. An **entity** (also referred to as a table) stores information about a person, place, thing, transaction, or event. The entities, or tables, of interest in Figure 6.10 are *TRACKS*, *RECORDINGS*, *MUSICIANS*, and *CATEGORIES*. Notice that each entity is stored in a different two-dimensional table (with rows and columns).

Attributes (also called columns or fields) are the data elements associated with an entity. In Figure 6.10, the attributes for the entity *TRACKS* are *TrackNumber*, *TrackTitle*, *TrackLength*, and *RecordingID*. Attributes for the entity *MUSICIANS* are *MusicianID*, *MusicianName*, *MusicianPhoto*, and *MusicianNotes*. A **record** is a collection of related data elements (in the *MUSICIANS* table, these include “3, Lady Gaga, gaga.tiff, Do not bring young kids to live shows”). Each record in an entity occupies one row in its respective table.

Creating Relationships Through Keys

To manage and organize various entities within the relational database model, you use primary keys and foreign keys to create logical relationships. A **primary key** is a field (or group of fields) that uniquely identifies a given record in a table. In the table *RECORDINGS*, the primary key is the field *RecordingID* that uniquely identifies each record in the table. Primary keys are a critical piece of a relational database because they provide a way of distinguishing each record in a table; for instance, imagine you need to find information on a customer

FIGURE 6.10
Primary Concepts of the
Relational Database Model

