

## MANAGING DATA RESOURCES<sup>i</sup>

Setting up a database is only a start. In order to make sure that the data for your business remain accurate, reliable, and readily available to those who need it, your business will need special policies and procedures for data management.

### ESTABLISHING AN INFORMATION POLICY

Every business, large and small, needs an information policy. Your firm's data are an important resource, and you don't want people doing whatever they want with them. You need to have rules on how the data are to be organized and maintained, and who is allowed to view the data or change them.

An **information policy** specifies the organization's rules for sharing, disseminating, acquiring, standardizing, classifying, and inventorying information. Information policy lays out specific procedures and accountabilities, identifying which users and organizational units can share information, where information can be distributed, and who is responsible for updating and maintaining the information. For example, a typical information policy would specify that only selected members of the payroll and human resources department would have the right to change and view sensitive employee data, such as an employee's salary or social security number, and that these departments are responsible for making sure that such employee data are accurate. If you are in a small business, the information policy would be established and implemented by the owners or managers. In a large organization, managing and planning for information as a corporate resource often requires a formal data administration function.

**Data administration** is responsible for the specific policies and procedures through which data can be managed as an organizational resource. These responsibilities include developing information policy, planning for data, overseeing logical database design and data dictionary development, and monitoring how information systems specialists and end-user groups use data.

You may hear the term **data governance** used to describe many of these activities. Promoted by IBM, data governance deals with the policies and processes for managing the availability, usability, integrity, and security of the data employed in an enterprise, with special emphasis on promoting privacy, security, data quality, and compliance with government regulations. A large organization will also have a database design and management group within the corporate information systems division that is responsible for defining and organizing the structure and content of the database, and maintaining the database. In close cooperation with users, the design group establishes the physical database, the logical relations among elements, and the access rules and security procedures. The functions it performs are called **database administration**.

### ENSURING DATA QUALITY

A well-designed database and information policy will go a long way toward ensuring that the business has the information it needs. However, additional steps must be taken to ensure that the data in organizational databases are accurate and remain reliable.

What would happen if a customer's telephone number or account balance were incorrect? What would be the impact if the database had the wrong price for the product you sold or your sales system and inventory system showed different prices for the same product? Data that are inaccurate, untimely, or inconsistent with other sources of information lead to incorrect decisions, product recalls, and financial losses. Gartner Inc. reported that more than 25 percent of the critical data in large Fortune 1000 companies' databases is inaccurate or incomplete, including bad product codes and product descriptions, faulty inventory descriptions, erroneous financial data, incorrect supplier information, and incorrect employee data. A Sirius Decisions study on "The Impact of Bad Data on Demand Creation" found that 10 to 25 percent of customer and prospect records contain critical data errors. Correcting these errors at their source and following best practices for promoting data quality increased the productivity of the sales process and generated a 66 percent increase in revenue.

Some of these data quality problems are caused by redundant and inconsistent data produced by multiple systems feeding a data warehouse. For example, the sales ordering system and the inventory management system might both maintain data on the organization's products. However, the sales ordering system might use the term *Item Number* and the inventory system might call the same attribute *Product Number*. The sales, inventory, or manufacturing systems of a clothing retailer might use different codes to represent values for an attribute. One system might represent clothing size as "extra large," whereas the other system might use the code "XL" for the same purpose. During the design process for the warehouse database, data describing entities, such as a customer, product, or order, should be named and defined consistently for all business areas using the database. Think of all the times you've received several pieces of the same direct mail advertising on the same day. This is very likely the result of having your name maintained multiple times in a database. Your name may have been misspelled or you used your middle initial on one occasion and not on

another or the information was initially entered onto a paper form and not scanned properly into the system. Because of these inconsistencies, the database would treat you as different people! We often receive redundant mail addressed to Laudon, Lavdon, Lauden, or Landon.

If a database is properly designed and enterprise-wide data standards established, duplicate or inconsistent data elements should be minimal. Most data quality problems, however, such as misspelled names, transposed numbers, or incorrect or missing codes, stem from errors during data input. The incidence of such errors is rising as companies move their businesses to the Web and allow customers and suppliers to enter data into their Web sites that directly update internal systems. Before a new database is in place, organizations need to identify and correct their faulty data and establish better routines for editing data once their database is in operation. Analysis of data quality often begins with a **data quality audit**, which is a structured survey of the accuracy and level of completeness of the data in an information system. Data quality audits can be performed by surveying entire data files, surveying samples from data files, or surveying end users for their perceptions of data quality.

**Data cleansing**, also known as *data scrubbing*, consists of activities for detecting and correcting data in a database that are incorrect, incomplete, improperly formatted, or redundant. Data cleansing not only corrects errors but also enforces consistency among different sets of data that originated in separate information systems. Specialized data-cleansing software is available to automatically survey data files, correct errors in the data, and integrate the data in a consistent company-wide format.

Data quality problems are not just business problems. They also pose serious problems for individuals, affecting their financial condition and even their jobs. For example, inaccurate or outdated data about consumers' credit histories maintained by credit bureaus can prevent creditworthy individuals from obtaining loans or lower their chances of finding or keeping a job.

---

<sup>i</sup> Laudon, K. C., & Laudon, J. P. Management information systems: Managing the digital firm (13th ed.). Upper Saddle River, NJ: Prentice Hall. (2014).