

Exploring Secondary Data

Learning Objectives

After reading this chapter, you should understand . . .

- 1 **The purpose and process of exploratory research.**
- 2 **Two types and three levels of management decision-related secondary sources.**
- 3 **The five types of external information and the five critical factors for evaluating the value of a source and its content.**
- 4 **The process for conducting a productive literature search of external sources, including print and electronic sources.**
- 5 **The process for conducting a productive literature search with Web-based sources.**
- 6 **What is involved in internal data mining and how internal data mining techniques differ from literature searches.**

Bringing Research to Life

“Tell me what you know about big banks, Jason,” Myra asked.

“Well, for my master’s thesis I wrote an econometric model of the Federal Reserve System.”

“Swell,” she said, impatiently. She dropped into a chair and locked her eyes onto his, a gesture that convinced him she was reading his thoughts before his mouth could say them. “Academic economic research, yes. Very impressive, I’m sure.” This acerbic remark she sweetened with a wry smile and then prodded: “The political climate for banks. What do you know, Jason?”

“Well, Myra, it’s never good for banks when liberals are running Congress and the White House. Is it?”

“The big banks can feel the eyes of Uncle Sam on them, morning, noon, and night. Which brings us to this.” She produced his appointment planner from the top of his desk. “I stopped by today to drop off the draft of my final report for the economic council, the one you said you’d review and give me some feedback on. Thought I’d pencil myself into your planner for Wednesday. I could not help noticing you have an appointment this very noon with a Mr. Armand Croyand of Denver. Do you know who he is?”

“No, not really. He identified himself as the principal of Croyand Associates and wanted a few minutes of my time over lunch, as he was swinging through town. I was free, and I said OK.”

“Saying yes to a meeting was a smart move,” claimed Myra. “Leaving for the meeting without discovering in advance who Croyand is . . . well, that would be dumb.”

“Since your modem never sleeps, nor does the online database of the *Rocky Mountain News*, a few minutes ago I dialed in. Want to know what a few dollars of computer time discovered? Mr. Croyand and his associates are the long-time research organiza-

tion working for InterMountBanc, headquartered in Colorado, and he has a reputation for doing fine, deep, imaginative real estate research.

“Further rummaging in the News’s database reveals that several of the directors recently resigned rather than face scrutiny of their private dealings, and that the new CEO of InterMountBanc and the new directors are all Mr. Clean types who have publicly pledged to stand by the highest ethical standards.

“Now, the plot thickens, doesn’t it? I dipped next into the online database of the *Miami Herald*, AND WHAT DO I FIND? I find that InterMountBanc is heavily invested in that new Westridge development in West County, the one that Ed Byldor is throwing up, the one that is now stopped dead in its tracks while the county commission, in all its wisdom, awaits recommendations from the county planners.”

“You are telling me, then, that Mr. Armand Croyand is in town on business and I must treat this as a serious business lunch.”

“I am telling you to go to lunch loaded for bear, because this could lead to a very big contract. How are you going to handle Mr. Croyand?”

“Well, first,” considered Jason, “I will go into the *Herald’s* online database and see if the planning commission has taken a public position on Westridge.” Myra nodded. “I have to swing by the Chamber of Commerce in an hour to drop off a final report of the aviation study, and that will give me a chance to discover what they have picked up off the record about Westridge, the planning council, and the county commission. If they are sensitive to public opinion, that works in my favor, since I have a political polling background. If they are worried about real estate values, that is harder, but I suppose I can call the library and find out the address of some trade association that can give me a jump start.”

"And since the county government center is within walking distance, I'll check the minutes of the county commission for any reports it has received from the planning council."

"Good. What else will you do while you are in the governmental center?"

"Something else?"

"Think."

He thought. "I'll check the bulletin board in the lobby and see when the planning council is due to hold its next public meeting, and I'll see if Westridge is on the agenda of the county commission meeting

for this coming Tuesday."

"OK, so far. And suppose you run into one of the commissioners. What will you do?"

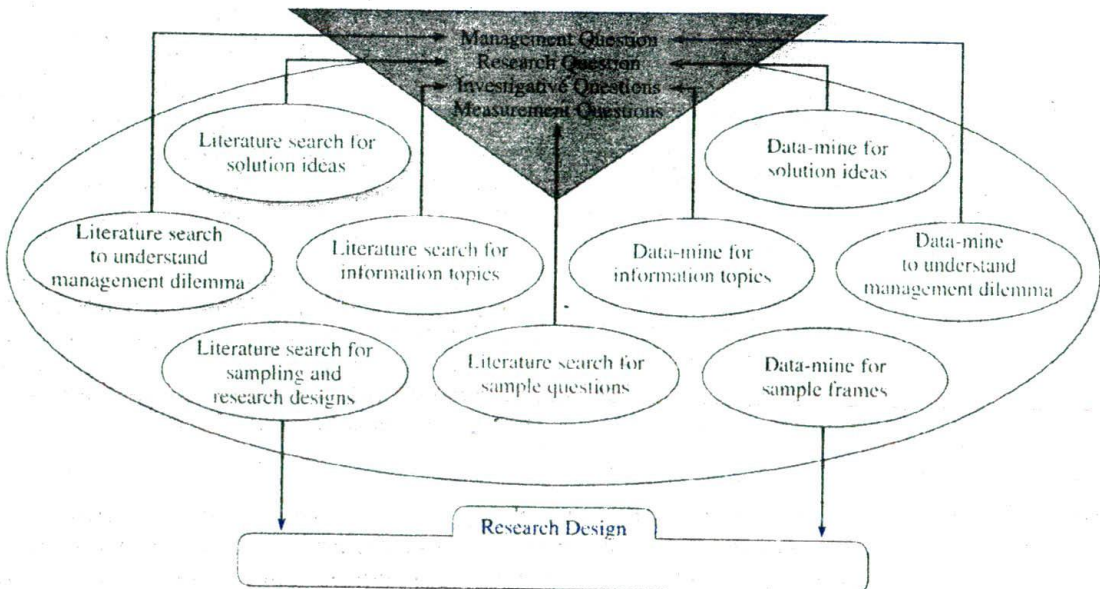
"Ask about his take on the Westridge project?"

"Now you're talking," smiled Myra. "When you meet Croyand, concentrate on what he is telling you, and let him discover almost by accident that you are on top of the real estate situation in West County. Make him believe his prowess as a researcher directed him to you, and that he found exactly the right person to help him here in Florida. Let your secondary research scream quietly of your competence."

The Exploratory Phase Search Strategy

As the exploration process modeled with the management-research question hierarchy suggests (see Exhibit 10-1), exploration of secondary sources may be useful at any stage of the hierarchy. But most researchers find a review of secondary sources critical to moving from management question to research question. In moving from management question to research question, the researcher uses both internal and external secondary sources. We address external sources first. Our discussion of data mining of internal sources completes the chapter.

EXHIBIT 10-1 Integration of Secondary Data into the Research Process



In this **exploratory research** phase of your project, your objective is to accomplish the following:

MANAGEMENT



- Expand your understanding of the management dilemma.
- Look for ways others have addressed and/or solved problems similar to your management dilemma or management question.
- Gather background information on your topic to refine the research question.
- Identify information that should be gathered to formulate investigative questions.
- Identify sources for and actual questions that might be used as measurement questions.
- Identify sources for and actual sample frames that might be used in sample design.

In most cases the exploration phase will begin with a literature search—a review of books as well as articles in journals or professional literature that relate to your management dilemma. A literature search requires the use of the library's online catalog and one or more bibliographic databases or indexes. For some topics, it may be useful to consult a handbook or specialized encyclopedia first to establish a list of key terms, people, or events that have influenced your topic and also to determine what the major publications are and who the foremost authors are. Other reference materials will be incorporated into your search strategy as needed. In general, this **literature search** has five steps:

MANAGEMENT



1. Define your management dilemma or management question.
2. Consult encyclopedias, dictionaries, handbooks, and textbooks to identify key terms, people, or events relevant to your management dilemma or management question.
3. Apply these key terms, people, or events in searching indexes, bibliographies, and the Web to identify specific secondary sources.
4. Locate and review specific secondary sources for relevance.
5. Evaluate the value of each source and its content.

The result of your literature search may be a solution to the management dilemma. In such a case, no further research is necessary. Often, however, the management question remains unresolved, so the decision to proceed generates a research proposal (see Chapter 4). The resulting proposal covers at minimum a statement of the research question and a brief description of proposed research methodology (see Chapter 4). The proposal summarizes the findings of the exploratory phase of the research, usually with a bibliography of secondary sources that have led to the decision to propose a formal research study.

In this chapter we will concentrate on the exploration phase of the project and focus on finding, selecting, and evaluating information in both printed and electronic formats. As you learned in Chapter 6, the first step in an exploratory study is a search of the secondary literature. We defined secondary literature as “studies made by others for their own purposes.” These studies, representing primary research to their authors, actually represent only a subset of all the information sources available.

Levels of Information

As you explore your problem or topic, you may consider many different types of information sources, some much more valuable than others. Information sources are generally categorized into three levels: (1) primary sources, (2) secondary sources, and (3) tertiary sources.

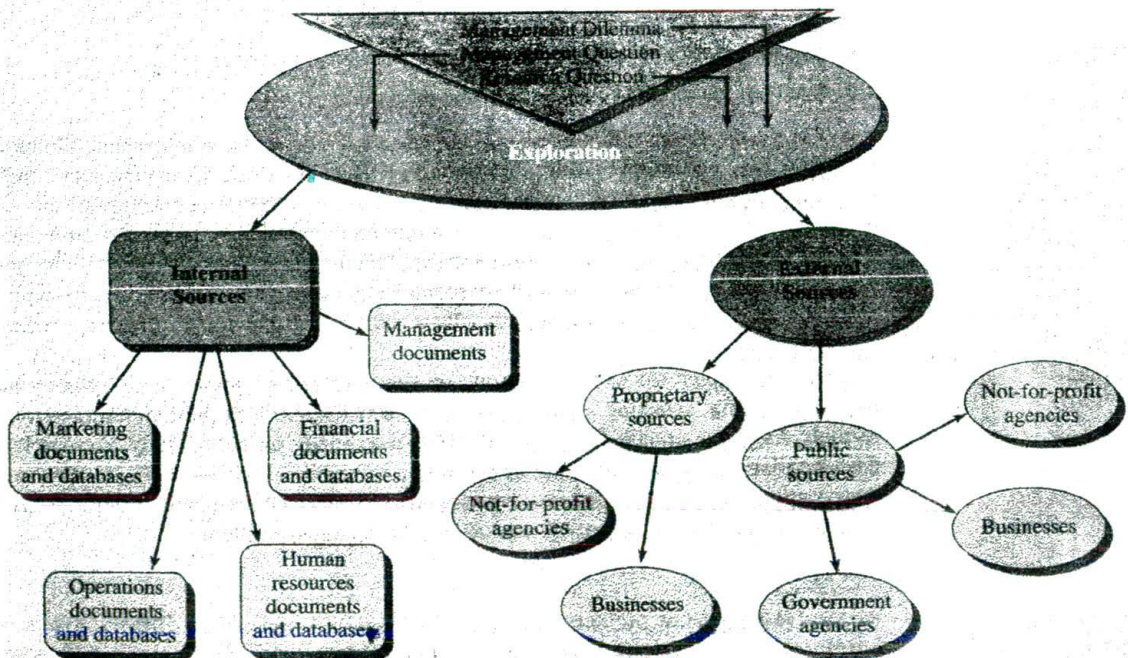
Primary sources are original works of research or raw data without interpretation or pronouncements that represent an official opinion or position. Included among the primary sources are memos, letters, complete interviews or speeches (in audio, video, or written transcript formats), laws, regulations, court decisions or standards, and most government data, including census, economic, and labor data. Primary sources are always the most authoritative because the information has not been filtered or interpreted by a second party. Information from all of the above will become your secondary literature supporting *your* original research. Internal sources of primary data would include inventory records, personnel records, purchasing requisition forms, statistical process control charts, and similar data.

Secondary sources are interpretations of primary data. Encyclopedias, textbooks, handbooks, magazine and newspaper articles, and most newscasts are considered secondary information sources. Indeed, nearly all reference materials fall into this category. Internally, sales analysis summaries and investor annual reports would be examples of secondary sources as they are compiled from a variety of primary sources. To an outsider, however, the annual report is viewed as a primary source, as it represents the official position of the corporation. A firm searching for secondary sources can search either internally or externally, as Exhibit 10-2 depicts.

Tertiary sources may be an interpretation of a secondary source but generally are represented by indexes, bibliographies, and other finding aids (e.g., Internet search engines).

From the beginning, it is important to remember that all information is not of equal value. As the source levels indicate, primary sources have more value than secondary

EXHIBIT 10-2 Secondary Sources for Developing the Management-Research Question Hierarchy



sources, and secondary sources have more value than tertiary sources. In the opening vignette, Myra reads a *Rocky Mountain News* account (a secondary source) but suggests that Jason check the county council minutes (a primary source). While the *Rocky Mountain News* may have excellent staff writers who verify every detail in every story with two or more corroborating sources, Myra doesn't know this for sure. Since Jason's acceptance by a new potential client is likely to hinge on his appearing knowledgeable about the Westridge development, verifying the account is important. *If the information is essential to solving the management dilemma, it is wise to verify it in a primary source.*

Types of Information Sources

We provide a list of key business resources in Appendix A and on your CD.

There are dozens of types of information sources, each with a special function. In this section we describe five of the information types used most by business researchers. Later in this chapter we will provide a more in-depth examination of three information types: bibliographic databases, government information, and the World Wide Web.

Indexes and Bibliographies Indexes and bibliographies are the mainstay of any library because they help you identify and locate a single book or journal article from among the millions published. The single most important bibliography in any library is its online catalog. As with all other information types, there are many specialized indexes and bibliographies unique to business topics. These can be very useful in a literature search to find authors and titles of prior works on the topic of interest.

Dictionaries Dictionaries are so ubiquitous they probably need no explanation. We all use them to verify spelling or grammar usage or to define terms. In business, as in every field, there are many specialized dictionaries that define words, terms, or jargon unique to a discipline. Most of these specialized dictionaries include in their word lists information on people, events, or organizations that shape the discipline. They are also an excellent place to find acronyms. A growing number of dictionaries and glossaries (terms in a specialized field, area, or topic plus their definitions) are now available on the Web. One of these is the Federal Reserve Bank of Chicago's *Glossaries of Financial Terms* (www.chicagofed.org/publications/glossary/index.cfm). An example of a printed business dictionary is the *Dictionary of Business and Management*. Information from dictionaries and glossaries may be used to identify key terms for a search of an online or printed database.

Encyclopedias Use an **encyclopedia** to find background or historical information on a topic or to find names or terms that can enhance your search results in other sources. For example, you might use an encyclopedia to find when Microsoft introduced Windows, then use that date to draw more information from an index to the time period. Encyclopedias are also helpful in identifying the experts in a field and the key writings on any topic. The *Encyclopedia of Company Histories* and the *New Palgrave Dictionary of Economics and the Law* are two examples of specialized multivolume encyclopedias.

Handbooks A **handbook** is a collection of facts unique to a topic. Handbooks often include statistics, directory information, a glossary of terms, and other data such as laws and regulations essential to a field. *The best handbooks include source references for the facts they present.* The *Statistical Abstract of the United States* is probably the most valuable and frequently used handbook available. It contains an extensive variety of facts, an excellent and detailed index, and a gateway to even more in-depth data for every table included.

Directories Directories are used for finding names and addresses as well as other data. While many are available and useful in printed format, directories in digitized format that can be searched by certain characteristics or sorted and then downloaded are far more useful. Many are available free through the Web, but the most comprehensive directories are proprietary (that is, must be purchased). An especially useful directory available in most libraries either in print or electronic format is the *Encyclopedia of Associations* (called *Associations Unlimited* on the Web), which provides a list of public and professional organizations plus their locations and contact numbers.

Evaluating Information Sources

MANAGEMENT



As you begin to collect information about your topic, one thing you will certainly want to do is conduct a **source evaluation**. Librarians evaluate and select information sources based on five factors that can be applied to any type of source, whether printed or electronic. These are

- Purpose
- Scope
- Authority
- Audience
- Format

Exhibit 10-3 summarizes the critical questions a researcher asks when applying these factors to the evaluation of Web sources.

Purpose The **purpose** of the source is what the author (or in the case of many Internet sites, the collective authors in an institution) is trying to accomplish. In general, the purpose may be to enlighten or to entertain. Among purposes in the enlighten subset, authors may be attempting to establish credibility, broaden knowledge within a field or discipline, establish a company image, and, for websites, even manage inventory or sell merchandise. Often, for proprietary sources, the **purpose** is to make the search process easier for those who buy. For instance, the purpose of a source may be to consolidate Web advertising sources and make these available to students (www.advertising.utexas.edu/world/). Or the purpose may be to define basic computer terms so nontechnical managers can communicate with technical staff. Once you determine the **purpose** of the source, you will also want to determine whether or how it provides a bias to the information presented. **Bias** is the absence of a balanced presentation of information. Most researchers expect company websites to be biased in favor of the company. However, we expect sources offered by independent organizations to be more balanced, presenting both positive and negative information about relevant organizations without favoring one or the other.

Scope Tied closely to the purpose of the source is its **scope**. What is the date of publication? What time period does this source cover? How much of the topic is covered and to what depth? Is the material covered local, regional, national, or international? If the source is bibliographic, how comprehensive is it? If it is a biographical source or a directory or a bibliography, what are the criteria for inclusion? *If you do not know the scope of your information sources, you may miss essential information by relying on an incomplete source.*

This single factor led to tragic results in a university clinical research project in 2001. The medical researcher combed the literature online, found no problems, and proceeded to offer a drug to a health volunteer who then died of complications. The



EXHIBIT 10-3 Evaluating Websites as Information Sources

Evaluation Factor	Questions to Answer
Purpose	<ul style="list-style-type: none"> • Why does the site exist? • How evident is the purpose it is trying to convey? • Does it achieve its purpose? • How does its purpose affect the type and bias of information presented?
Authority	<ul style="list-style-type: none"> • What are the credentials of the author or institution or organization sponsoring the site? • Does the site give you a means of contacting anyone for further information? • Who links to this site? • If facts are supplied, where do they come from?
Scope	<ul style="list-style-type: none"> • How old is the information? • How often is it updated? • How much information is available? • Is it selective or comprehensive? • What are the criteria for inclusion? • If applicable, what geographic area or time period or language does it cover? • How does the information presented compare with similar sites? • Is it a series of links only (a metasite), or is there added value? • What is the nature of the added value? • What information did you expect to find that was missing? • Is the site self-contained or does it link to other websites?
Audience	<ul style="list-style-type: none"> • Whom does the site cater to? • What level of knowledge or experience is assumed? • How does this intended audience affect the type and bias of the information?
Format	<ul style="list-style-type: none"> • How quickly can you find needed information? • How easy is the site to use? Is it intuitive? • Does it load quickly? • Is the design appealing? • Are there navigation buttons? • Is there a site map or search button? • Is there an easily identifiable help button? • Is Help helpful? • Are pages in ASCII or graphic format? • Is the information downloadable into a spreadsheet or word-processing program, if desired?

salient reports on the ill effects of the drug had been published in the early 1960s and were too old to be included in the online database used by the researcher.¹

Authority Of major concern to any information user is the **authority** of the source. We have already noted that primary sources are the most authoritative. In any source, both the author and the publisher are indicators of the authority. The author and the author's credentials should be given both in printed and electronic sources. Footnotes should be provided when appropriate. If credentials are not given, then it is best to check a biographical source. Credentials may include the author's educational background, his/her position, or his/her other published and reviewed work.

Authority also applies to Web resources where anyone can post anything. In this environment it is always important to check the credentials of the site. For instance, data and statements about cancer are much more likely to be authoritative if they come from the National Cancer Institute or the American Cancer Society than from a personal page with no information about the author or producer. Any personal page on the Web is suspect unless the author's, or in some cases the institution's, credentials can be verified. Credentials alone are not always enough. Most scholarly articles are validated by a system called *peer review* in which colleagues from other institutions are asked to comment anonymously on the research presented. In several recent cases authors with very good credentials have bypassed the peer review process and published directly on the Web. In at least one instance the research was seriously flawed. Even the best scholars can make mistakes, so imagine what nonexperts might publish on the Web.

Audience Audience is also an important factor in evaluating an information source and it, too, is tied to the purpose of the source. The audience for this textbook is college students. More specifically, its audience is college students who are studying or majoring in business or public administration, some of whom are practicing managers. While others, for example educators, may benefit from the information, the authors take great care to select examples and to write in terms that management students will easily relate to.

It is often difficult to determine the intended audience for some Web resources, as many websites are designed to serve multiple audiences. The Web is available to all, so Web designers need to be especially creative both in capturing their audience(s) and in moving them to the appropriate section(s) of their website. While an organization may have several different print publications geared to different users, all the users on the Web go through one home page or portal. Universities, for example, have a very difficult time. Should they target their site to alumni or donors whom they count on for support? Or to prospective students? Or to current students? Or to faculty and staff? Luckily, technology is helping solve this problem, by helping the Web server to first identify characteristics of the user (through cookies) and then select the appropriate home page to deliver.

Brokerage firm Charles Schwab has no confusion on purpose or audience. In a recent ad, Charles Schwab is quoted as follows: "I see the Internet as the single most empowering force for the individual investor."² The Schwab.com website is designed to empower every single Charles Schwab customer, with rapid market summary updates and research on companies and funds. The numerous awards Schwab has won for its website indicate it is doing well at achieving its aims. It also uses an intranet to provide key information to its employees.³

When you are evaluating the plausible audience of a source, look for key indicators including vocabulary, types of information, and questions or directions that guide the search. CNNfn (www.cnnfn.com) clearly has a well-educated audience composed of people who are serious investors and people who are more than mildly interested in—

or who rely on information about—the global economy. Indicators of this are not only the site's coverage in general, but also the great detail offered and the emphasis on up-to-the-minute business news stories and data on U.S. and world markets. The site focuses on services that will facilitate investing. In addition, the site provides information about CNN-sponsored television programs, currency, and travel information—all of which relate to the U.S. and global economy.

Format **Format** factors may vary from source to source but in general relate to how the information is presented and how easy it is to find a specific piece of information. In a printed source, the arrangement of the information—alphabetical? hierarchical? chronological?—nearly always has an impact on the retrieval of information. Indexes are usually essential. Do cross-references link one term to related terms? How are acronyms handled? Is the reference to an item? Table number? Page? How do type fonts or color help you find information?

The format of an electronic source or website is generally related to design. Web designers contend with a variety of Internet browsers, individual computer “preferences,” and a wide variety of modem or Internet access speeds. On the one hand, the designer wants the page to look great and have the coolest bells and whistles. On the other hand, the designer has to plan for users who are not very patient. People with visual impairments, using text only or specialized software that “reads” the page, must be taken into consideration, too. A page that takes even 30 seconds to load may be abandoned while the user goes on to something else. Users can't flip through pages on the Web nearly so quickly or effectively as they can with a reference book. Navigation becomes an issue. If you are holding a book, you know how to get back to the title page. If you are using a website, you want to be able to go forward one screen, backward one screen, return to home page, search the site from any page and quickly find the home page. On every page within a site, you should be able to see the name of the site owner and the last time the page was updated and, preferably, you should be able to contact the author or site manager.⁴

Searching a Bibliographic Database

In a **bibliographic database**, each record is a bibliographic citation to a book or a journal article. In your university library, the online catalog is an example of a bibliographic database.

There are several bibliographic databases available to business researchers. (See Appendix A and your CD.) Some of the more popular and comprehensive business bibliographic databases are:

- ABI/Inform (from Proquest Information and Learning).
- Business and Industry (from Gale Group).
- Business Source (from EBSCO).
- Dow Jones Interactive.
- Lexis-Nexis Universe (from a division of Reed Elsevier).

Most of the above databases offer numerous purchase options both in the amount and type of coverage. Some include abstracts, short summaries of the articles cited. Nearly all of the above databases include the contents of around two-thirds of the indexed journals in full text, although the amount and the specific titles may vary widely from database to database. Full-text options vary from an exact image of the page to ASCII-text-only or text-plus-graphics. Search options also vary considerably from database to database. It is for these reasons that most libraries supporting business programs offer more than one business periodical database.

MANAGEMENT



The process of searching bibliographic databases and retrieving results is basic to all databases.

MANAGEMENT

Tip

1. Select a database appropriate to your topic.
2. Construct a **search query** (also called a *search statement*).
 - Review and evaluate search results.
 - Modify the search query, if necessary.
3. Save those valuable results of your search.
4. Retrieve articles not available in the database.
5. Supplement your results with information from Web sources.

Select a Database Most of us select the most convenient database without regard to its scope; but considering the database contents and its limitations and criteria for inclusion at the beginning of your search will probably save you time in the long run. Remember that a library's online catalog is a bibliographic database that will help identify books and perhaps other media on a topic. While journal or periodical *titles* are listed in a library's online catalog, periodical or journal *articles* are rarely included. Use books for older, more comprehensive information. Use periodical articles for more current information or for information on very specific topics. A librarian can suggest one or more appropriate databases for the topic you are researching.

The Close-Up provides direction for advanced database searches, including how initial searches are modified to obtain more relevant results.



Close-Up

Advanced Searching

In advanced searches, you use your knowledge of the database to make the search more productive.

Construct the Search Query

Use the keywords from your management question to prepare a query for the database. Bibliographic databases, including the library's online catalog, all have similar search options, such as a basic keyword search, an advanced search, and a way to choose a subject from a browse list. Like all databases, bibliographic databases consist of several standard fields.

In most bibliographic databases, all searches are keyword searches, but it is possible to search for a specific author or title or series (a known-item search) by limiting your results to a specific field of the bibliographic record. This is especially important if you are researching a prolific author such as Peter Drucker, who may have many works both by and about him. If you do not limit or narrow your search to a specific field, then you will conduct a general keyword search of all the records in the database. Because of the size of most databases, single word searches generally yield results that are not very useful

unless the single word is very unique. Instead, examine your management question for all relevant keywords and variations and establish a more precise search query using the connectors described on the next page:

The most important thing to remember about search engines for the Web or for databases is that they do not all work alike. In fact, they have widely varying search protocols. What you do not know can act against you. So, if finding good information is important to you, take a couple of minutes to determine what special features and search options are used. For instance, if you enter a multiword term, what happens? Does the database search your term as a phrase? Or does it insert a connector such as *and* or *or* between each word? How does it handle stopwords (the, in, and other similar small words)? The results will vary considerably in these three scenarios.

Standard Search Fields for Monographs

Author	Publisher
Title	Series
Subject headings	

Limiters in Book Catalogs

Language	Type of publication
Date of publication	Format (book, video, etc.)

Standard Search Fields for Periodical Databases

Author	Abstract
Title of article	Company name
Subject headings	NAICS/SIC code
Periodical title	

Limiters in Periodical Databases

Date	Periodical title
Full text	Peer review (scholarly journals)

Search Strategy Options

Basic Searching If you have a unique term, try a basic search using that term. Most bibliographic databases will present the results list in reverse data order, that is, the most recently published items will appear first. Review the list of items your search has retrieved. Are there too many?

Not enough? Are they very relevant or not very relevant? If they meet the Goldlocks test of "just right," then you can move on to the next step (saving results).

Advanced Searching If you have retrieved too few or no relevant items, or if you have retrieved hundreds of items, you should consider modifying your search query. Start with the most relevant items in the results list. Then do one of the following:

- Search for the cited works (the bibliography) of the full-text articles.
- Search for other works by the author or authors of the relevant citations.
- Check the subject headings assigned to the articles. Are there any more precise terms or synonyms that would improve your search results? More importantly, are there pairs of terms that appear in all of the most relevant items? Is there a thesaurus with the database that defines or expands the terminology used in the subject headings?

As a result of your examination of the relevant citations and any background preparation you have done in other sources such as encyclopedias, you should now have one or more keywords and synonyms for each concept. You can now use Boolean operators or connectors (see Exhibit 10-4)

EXHIBIT 10-4 Review of Advanced Search Options

Narrowing Your Search		
OR	AND	Phrases
Use OR to search for plurals, synonyms, or spelling variations. Either or both terms will be present in results. <ul style="list-style-type: none"> • woman OR women • business OR corporation • international OR foreign 	Use AND to require that all terms you specify be present in the results. <ul style="list-style-type: none"> • child AND advertising 	Use a term consisting of two or more words. Some phrases require double quotes to enclose the phrase, while others do not. <ul style="list-style-type: none"> • human resource management • "human resource management"
Truncation	NOT	ADJ
Symbols (?, *, !) that replace one or more characters or letters in a word or at the end of a word. <ul style="list-style-type: none"> • electr* (retrieves electricity, electric, electrical) <ul style="list-style-type: none"> • child? (retrieves children, childish, child's)	Use NOT to eliminate terms from your search. But use NOT with care. It is easy to eliminate the good with the unwanted. <ul style="list-style-type: none"> • medicine NOT nursing • Caribbean NOT Cuba 	ADJ requires the first term specified to immediately precede the last term specified. <ul style="list-style-type: none"> • six ADJ sigma
Limiters		
Conditions (date, publication type, language) for limiting your search. Most databases also offer <i>field limiting</i> , limiting the occurrences of your search to a specific database field, such as the author field, title, etc. Some bibliographic databases offer the convenience of limiting the search results to peer-reviewed articles or to articles only available in full text. Use the latter with care as some significant articles may be overlooked even though they are available in the library.		

EXHIBIT 10-5 Advanced Searching Process

Concept A	Operator	Concept B	Operator	Concept C
training	AND	sex* harassment	AND	lawsuit*
awareness		wom* n		law
behavior		female		courts
professional		gender		legal
development		men		

Step 2: Create and search with a concept group by combining each term in a column with OR. Put each concept group in parentheses. Then combine each concept group with AND.

(training OR awareness OR behavior OR professional development) AND (sex* harassment OR wom* n OR men OR female OR gender) AND (lawsuit OR legal OR law OR courts)

to combine terms or sets of terms to expand or narrow your search. There are four basic Boolean operators or connectors: OR, AND, NOT, and AND.

Think of your management question as a series of keywords. For example, your management question might be, "How can I design an appropriate training or awareness program to prevent sexual harassment lawsuits in my company?" In this example, concept A would be training; concept B would be harassment; concept C would be lawsuits. In the most basic of keyword searches, you

could use a keyword search with the operator AND to combine them:

training AND harassment AND lawsuits

If your search results are inadequate, you might need to expand your search statement with synonyms connected with the operator OR. For our sample management question, your search would look like Exhibit 10-5. If your search results are too numerous, you will need to limit your search.

Save Results of Search While the temptation to print may be overwhelming, remember that if you download your results, you can cut and paste quotations, tables, and other information into your proposal without rekeying. In either case, make sure you keep the bibliographic information for your footnotes and bibliography. Most databases offer the choice of marking the records and printing or downloading them all at once or printing them one by one.

Retrieve Articles For articles not available in full text online, retrieval will normally require the additional step of searching the library's online catalog (unless there is a link from the database to the catalog) to determine if the desired issue is available and where it is located. Many libraries offer a document delivery service for articles not available. Some current articles may be available on the Web or via a fee-based service.

Searching the World Wide Web for Information

The World Wide Web is a vast information, business, and entertainment resource that would be difficult, if not foolish, to overlook. Millions of pages of data are publicly available, and the size of the Web doubles every few months.⁵ But searching and retrieving reliable information on the Web is a great deal more problematic than searching a bibliographic database. There are no standard database fields, no carefully defined

subject hierarchies (called *controlled vocabulary*), no cross-references, no synonyms, no selection criteria, and, in general, no rules. There are dozens of search engines and they all work differently, but how they work is not always easy to determine. Nonetheless, the convenience of the Web and the extraordinary amount of information to be found on it are compelling reasons for using it as an information source.

As you can see in Exhibit 10-6, the basic steps to searching the Web are similar to those outlined for searching a bibliographic database. As you approach the Web, you start at the same point: focusing on your management question. Are you looking for a known item (for example, IBM's website or that of the American Marketing Association)? Are you looking for information on a specific topic? If you are looking for a specific topic, what are its parameters? For example, if your topic is managed health care, are you hoping to find statistics, marketing ideas, public policy issues, accounting standards, or evidence of its impact on small business?

There are perfectly legitimate reasons to browse for information, and with its hypertext linking system, the Web is the ultimate resource for browsing. The trick is to browse and still stay focused on the topic at hand. In the browse mode you do not have any particular target. You follow hypertext links from site to site for the sheer joy of discovery. This is somewhat analogous to window shopping at the mall or browsing the bookshelves in the library. It may or may not be fruitful. And browsing is not likely to be efficient. Researchers often work on tight deadlines, as managers often cannot delay critical decisions. Therefore, researchers rarely have the luxury of undirected browsing.

Below we detail those steps in the Web search process that pose altered behavior for Web searches.

Select a Search Engine or Directory A search for specific information or for a specific site that will help you solve your management question requires a great deal more skill and knowledge than browsing. Start by selecting one or more Internet search engines. Web search engines vary considerably in the following ways:

- The types of Internet sources they cover (http, telnet, Usenet, ftp, etc.).
- The way they search Web pages (every word? titles or headers only?).
- The number of pages they include in their indexes.

EXHIBIT 10-6 Web Search Process Compared to Bibliographic Search Process

Bibliographic Search Process	Web Search Process
1. Select a database appropriate to your topic.	1. Select a search engine or directory.*
2. Construct a search query. <ul style="list-style-type: none"> • Review and evaluate search results. • Modify the search query, if necessary. 	2. Determine your search options.*
3. Save those valuable results of your search.	3. Construct a search query.* <ul style="list-style-type: none"> • Review and evaluate search results. • Modify the search query, if necessary.
4. Retrieve articles not available in the database.	4. Save those valuable results of your search.*
5. Supplement your results with information from Web sources.	5. Supplement your results with information from non-Web sources.*

*Denotes a variation.

- The search and presentation options they offer.
- The frequency with which they are updated.

Furthermore, some information publicly indexable via the Web is not retrievable at all using current Web search engines. Among the material open to the public, but not indexed by search engines, is the following:⁶

MANAGEMENT



- Pages that are proprietary (that is, fee-based) and/or password-protected, including the contents of bibliographic and other databases. Some password-protected databases, such as the Thomas Register, are actually free and available to the public after initial registration.
- Pages accessible only through a search form (databases), including such highly popular Web resources as library catalogs, e-commerce catalogs (such as Lands'End[®], Amazon.com[®], and similar offerings), and the Security and Exchange Commission's EDGAR catalog of SEC filings. If you want to find the price of a book at Amazon.com, you first will have to find the Amazon.com page, then search the database for the title.
- Poorly designed framed pages.
- Some non-HTML or nonplain-text pages, especially PDF graphics files, for which no text alternative is offered. These pages cannot be retrieved using any current search engine.
- Pages excluded by the Robots Exclusion Standard (usually implemented with a robots.txt file). This standard is used by Web administrators to tell indexing robots that certain pages are off-limits. An outstanding example of this is the U.S. government's extensive information resource called GPO Access (described below).

The search engine, portal, or directory you select may well be determined by how comprehensive you want your results to be.⁷ If you want to use some major sites only, then start with a directory such as Yahoo![®]. If, however, you are interested in gathering comments and opinions that are the focus of usenet groups, then use a more inclusive search engine such as Northern Light[®]. At least within the publicly indexable pages, one approach emphasizes selectivity, and the other, comprehensiveness. *If you are interested in comprehensiveness, use more than one search engine. You are likely to yield very different and perhaps better results using additional search engines.*

What is the difference between a search engine, a portal, and a directory? Directories rely on human intervention to select, index, and categorize Web contents. Subject directories build an index based on Web pages or websites, but not on words within a page. Presenting a series of subject categories that are then further subdivided, Yahoo![®] (<http://www.yahoo.com/>) was the first Web subject directory and is still one of the most popular choices for finding information on the Web. This is because most users are satisfied with a few good sites rather than a long list of possibilities.

A search engine's different software components allow it to search and retrieve Web pages. These include:

- Software that automatically sends robots, sometimes called spiders, out to comb the Web going from server to server to build an index of the words, pages, and files that are publicly indexable.
- Algorithms that determine how those pages will be selected and prioritized for display.
- User interface software that determines the search options available to the user.

Robots may be sent to roam the Web on a daily basis or on a six-week basis, so it is possible that some newer pages may not be included in the index developed by a particular search engine. Most search engines try for at least a monthly update of their indexes. Some robots may only search the upper-level pages and totally ignore valuable pages buried within the site.

The algorithm used by the search engine can have an enormous impact on the type and quantity of information retrieved. Search algorithms determine whether every word is to be included or only the top 50 or so words; whether more weight will be given to words in metadata or titles or in words used frequently. The possibilities are limitless and are the major reason that results from one search may vary considerably from those of another search.

A **portal** is, as the name suggests, a gateway to the Web. A portal often includes a directory, a search engine, and other user features such as news and weather. Most Internet service providers (ISP) are portals to the Web. The AOL[®] homepage is an example. This portal uses information based on past user search behavior to determine what to offer on the opening screen. Therefore, some valuable search engines, indexes, directories, and more may be relegated to an "other search aids" category. If as a researcher your behavior differs from the majority pattern, you have to be more knowledgeable about search strategies to bypass those front-line strategies offered by the portal. Several ISP portals now offer subscribers the option of customizing the portal with the user-chosen search engines and secondary sources. Most of the major search engines are now actually portals to the Web that include their search engine. Specialized portals are increasingly popular. An industry portal, one type of specialized portal, lists many different resources about a specific industry. Competia Express, the competitive intelligence site (<http://www.competia.com/express/index.html>), offers industry portals for many different industries.

Determine Your Search Options Nearly all search services have a Help button that will lead you to information about the search protocols and options of that particular search engine. How does the search engine work? Can you combine terms using Boolean operators (AND, OR, NOT) or other connectors? How do you enter phrases? Truncate terms? Determine output display? Limit by date or other characteristic? Some search engines provide a basic and an advanced search option. How do they differ?

Construct a Search Query and Enter Your Search Term(s) The Web is not a database nor does it have a controlled vocabulary. Therefore, you must be as specific as possible, using the keywords in your management question and any variations you can think of. It is up to you to determine synonyms, variant spellings, and broader or narrower terms that will help you retrieve the information you need. This may involve some trial and error. For instance, a general term (such as *business*) would be useless in a search engine that purports to index every word in every document.

Save Results of Your Search If you have found good information, you will want to keep it for future reference so that you can cite it in your proposal or refer to it later in the development of your investigative questions. If you do not keep documents, you may have to reconstruct your search. At a future time, given that some portion of the Web is revised and updated daily, those same documents may no longer be available.

Supplement Your Results with Information from Non-Web Sources There is still a great deal of information in books, journals, and other print sources that is not available on the Web. While many novice researchers start and end here, the more sophisticated researcher knows a Web search is just one of many important options.

Searching for Specific Types of Information on the Web

MANAGEMENT



Once you have defined your topic and established your search terms, you need to determine whether you are looking for a specific site (known-item) or an address of a person or institution (who), a geographic place name and location (where), or a topic (what).

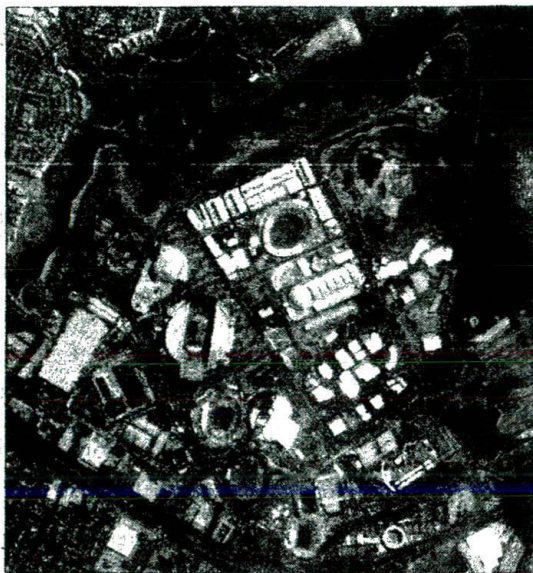
Known-Item Searches In the same way that search protocols for the library's online catalog vary between a known-item (author or exact title), and a more general keyword search, the way you query the Web for an exact item also varies from a more general query. A general trend among search engines is to establish algorithms that will yield more precise results. One of the first to follow this trend was the search engine Google® (<http://www.google.com/>), which debuted in 1998. Google® and others like it help you retrieve the most precise results from known-item searches by creating an algorithm that interprets a link to a site as a vote for that site. The sites that receive the most votes (links) rise to the top of the results list in a known-item search. The Google® system also emphasizes the importance of the linking page in its algorithm.

Who Searches In the "who" searches, you are looking for an e-mail address, a phone number, a street address, or a Web address of a person or institution. For this type of information you will first need to identify a database containing the information you need and then search that database according to its search protocols. At this writing, almost all Web search engines and portal sites partner with infoUSA® (<http://www.infoUSA.com/>) to supply the phone number databases for their white and yellow page services.⁸ See "Finding People on the Net" (<http://alabanza.com/kabacoff/Inter-Links/phone.html>) for a site listing many "who" options.

Where Searches A "where" option comprises the mapping services that help you locate an address on a map or discover the route from one place to another. Mapping services are databases tied to geographical information systems (GIS). Popular sites are Lycos® Roadmaps (<http://www.lycos.com/roadmap.html>) and MapQuest® (<http://www.mapquest.com>).

Today's sophisticated databases include more than numbers representing data. Many contain detailed geographic images, like the one here provided by Space Imaging, a leading supplier of visual information products, geographic information products, and related services. Space Imaging collects and/or distributes earth imagery from the Indian Remote Sensing satellites, the U.S. Landsat satellites, and radar imaging systems from Canada, Japan, and Europe. The company also delivers high-resolution aerial imagery collected by its own Digital Airborne Imaging System. Among a long list of activities, these digital information databases are used in insurance and risk management, mapping, telecommunications and utilities planning, and real estate planning and development, as well as security enhancement. This image of Sidney was used in planning the last Summer Olympics. Visit Space Imaging's Internet gallery to see a different perspective of news events.

www.spaceimaging.com



What Searches As we have already noted, search engines vary considerably in the way they work and in their size. If you are searching for a very unusual term, select one of the more comprehensive search engines, such as Northern Light® (<http://www.northernlight.com/>) or AltaVista® (<http://altavista.digital.com/>). Generally, it is more efficient to start with a directory such as Yahoo!® or one of the more specialized directories on the Web. Some especially good specialized sites are the Argus Clearinghouse (<http://www.clearinghouse.net/>), featuring subject guides on dozens of topics prepared by librarians and other specialists, and INFOMINE® Scholarly Internet Resource Collections from the University of California housed at University of California-Riverside (<http://infomine.ucr.edu/>). See Exhibit 10-7 or Appendix A for a selection of business-related websites.

Since the Web was introduced to the world in 1992, Web technology has been seeking ways to make the contents more accessible. The dynamic nature of the Web, its lightning growth rate, the ephemeral nature of some Web pages, the different skill levels and interests of users, and the lack of standards make this an enormous challenge. Trends indicate that the Web will continue to grow and that we will continue to apply new technologies to tapping the information available on this vast and unique resource.⁹ Already some search engines are better able to identify key resources. Efforts are under way to adopt standardized metatags included in the coding to describe the contents of Web pages. Some search engines are using expert systems to learn more about the information requestor. This is already being used extensively to target advertisements and is being used more frequently to "select" from among several options the information source that will be delivered to the requestor. More efforts are being made to index a larger portion of the Web content but, at the same time, efforts are also under way to improve the relevancy of the results delivered for any one search.

Government Information

The U.S. government is the world's largest source of data and information used by managers in all types of organizations.

Government publications, especially those of the U.S. government, are mandatory resources for many business research projects. The agencies of the U.S. government, considered as a whole, are the largest publishing body in the world. The government

It could be a best seller.



But it's free.

You can't buy the Consumer Information Catalog anywhere. But you can send for it, free! It's your guide to more than 200 free or low-cost government publications about getting federal benefits, saving and investing, staying healthy, and more. Send today for your latest free Consumer Information Catalog. The Catalog is free. The information is priceless. Send your name and address to:

**Consumer Information Center
Department BEST Pueblo, Colorado 81009**



EXHIBIT 10-7 Selected Websites for General Business Research

Site ID	Site URL	Sponsor
Advertising World	http://advertising.utexas.edu/world/	Department of Advertising, University of Texas, Austin
Annual Reports Online	http://www.zpub.com/sf/arl/aarl-www.html	Annual Reports Library, San Francisco
BizLink, Your On-line Business Resource	http://www.bizlink.org/	Public Library of Charlotte and Mecklenburg County
Business and Economics Resources Online	http://lcweb.loc.gov/tr/business/beonline/beonline.html	Library of Congress
CNNfn-the Financial Network	http://cnnfn.com	CNN America, Inc.
Electronic Commerce	http://ecommerce.about.com	About.com Inc.
ENTERWeb (Enterprise Development Website)	http://www.enterweb.org/	ENTERWeb
Guide to Labor-Oriented Internet Resources	http://www.lib.berkeley.edu/IIRL/iirlnet.html	Institute of Industrial Relations Library, University of California, Berkeley
Hoover's Online	http://www.hoovers.com	Hoover's, Inc.
Industry Research Desk	http://www.virtualpet.com/industry/rdindex2.htm	Polson Enterprises Research Services
MSU-CIBER International Business Resources on the WWW	http://ciber.bus.msu.edu/busres.htm	Michigan State University Center for International Business Education and Research
Rutgers Accounting Web	http://accounting.rutgers.edu/	Accounting Research Center, Rutgers University
Tax and Accounting Sites Directory	http://www.taxsites.com/	Schmidt Enterprises
Thomas Register of American Manufacturers	http://www.thomasregister.com/index.html	Thomas Publishing Co.
VIBES: Virtual International Business and Economics Resources	http://libweb.uncc.edu/ref-bus/vibehome.htm	J. Murrey Atkins Library, University of North Carolina, Charlotte
WorkIndex ("gateway to human resource solutions")	http://workindex.com	Human Resource Executive Magazine, in cooperation with the Cornell School of Industrial Relations

collects and provides access to a wide variety of social, economic, and demographic data. In addition, government laws and regulations, court decisions, policy papers, and studies all have a potential impact on business. Additionally, the government provides directories, maps, and other information sources. Specialists are available throughout the government to provide individual assistance.

Searching for government information is a complicated task that usually requires some knowledge of how government functions. In recent years, the U.S. government has been working aggressively to make its information available, not only through the Depository Library Program, but also through the development of electronic resources. As a result, information that used to be tucked into the farthest corners of the library is now readily available and searchable on the Web. In many libraries, the entire government documents collection is included in the library's catalog with links to Internet resources. In this section of the chapter, we examine three of the most useful government information types. A list of selected government resources on the Web is included in Exhibit 10-8.

Government Organization Two of the most useful resources regarding government organization are:

- *U.S. Government Manual* (published annually);
<http://www.access.gpo.gov/nara/nara001.html> (updated annually).
- *Congressional Directory* (published annually);
<http://www.access.gpo.gov/congress/cong016.html> (updated regularly).

EXHIBIT 10-8 A List of Selected Government Sources

Source	Internet
CBDNet (Commerce Business Daily) (a listing of government procurement, sales, and contract awards)	http://cbdnet.access.gpo.gov
Economic Indicators	http://www.access.gpo.gov/congress/cong002.html
Economic Statistics Briefing Room (current economic statistics)	http://www.whitehouse.gov/lsbr/esbr.html
EDGAR Database of Corporate Information (SEC filings)	http://www.sec.gov/edgar.shtml
Federal Bulletin Board	http://fedbbs.access.gpo.gov
FedStats	http://www.fedstats.gov
FirstGov: Business	http://www.firstgov.gov/topics/business.html
GPO Access	http://www.access.gpo.gov/su_docs/
Stat-USA (includes National Trade Data Bank)	http://www.stat-usa.gov/
Thomas (U.S. Congress on the Internet)	http://thomas.loc.gov
U.S. Bureau of Economic Analysis	http://www.bea.doc.gov/
U.S. Bureau of Labor Statistics	http://stats.bls.gov/
U.S. Bureau of the Census	http://www.census.gov/
U.S. Department of Commerce	http://www.doc.gov/
U.S. Department of Labor	http://www.dol.gov/
U.S. Patent and Trademark Office	http://www.uspto.gov/
U.S. Small Business Administration	http://www.sbaonline.sba.gov/

The *U.S. Government Manual* describes the functions of every government agency and is particularly useful for identifying key personnel and agency contacts, including those at the local and regional levels. Generally very knowledgeable and helpful, these people are invaluable in any research project for their ability to cut through red tape and to answer questions pertinent to their agencies.

The *Congressional Directory* lists members of Congress and Congressional committees, as well as key personnel throughout the U.S. government.

Laws, Regulations, and Court Decisions Government information regarding these key areas and other legal information can be obtained by consulting GPO Access (http://www.access.gpo.gov/su_docs/). GPO Access is the government's official and real-time site for finding government information. Included at this site is the complete *Monthly Catalog of U.S. Government Publications*. Use it to identify the full range of government publications, printed and electronic, issued in or after 1994.

Especially valuable on GPO Access are the databases covering laws, regulations, and congressional debates and publications. The key databases for laws and regulations are:

- **Congressional Bills.** Provides texts of varying versions of bills. Only a small portion of this collection ever becomes law, but the topics can reveal trends of interest to business researchers.
- **Public Laws.** Provides the texts of laws as they are passed; covers 1994 to the present. The printed version of this source is called the *U.S. Statutes at Large*.
- **U.S. Code.** A codification of laws currently in effect and as revised over time. Also available in libraries as a printed document.
- **Federal Register.** "The official daily publication for rules, proposed rules, and notices of federal agencies and organizations, as well as executive orders and other presidential documents;"¹⁰ covers 1995 to the present. Also available in libraries as a printed or microfiche document.
- **Code of Federal Regulations (CFR).** "A codification of the general and permanent rules originally published in the Federal Register by the executive departments and agencies of the federal government."¹¹ A printed version is available in libraries.

GPO Access includes other valuable databases, including the Supreme Court decisions from 1937 to the present, the *Economic Report of the President*, the *U.S. Budget, Commerce Business Daily* (CBDNet), and GAO reports. New databases are added regularly.

Many libraries have created local gateways to GPO Access that help speed information retrieval (<http://www.access.gpo.gov/congress/cong002.html>). GPO Access offers dozens of fields to search and these can be searched independently or together. Searching each database independently provides more flexibility and more precise searching options because there are some fields that are unique to a particular database. In general, search options are similar to other databases. Use the search hints for each file for more details and special search possibilities.

Government Statistics Information regarding government statistics may be obtained by consulting the following sources:

- *Statistical Abstract of the United States* (<http://www.census.gov/prod/www/statistical-abstract-us.html>).
- FedStats (<http://www.fedstats.gov>).
- U.S. Bureau of the Census (<http://www.census.gov/>).

MANAGEMENT



The government collects statistics on just about every topic imaginable—from crimes to hospital beds, from teachers to tax revenue, from steel production to flower imports. For any statistical inquiry, start with the *Statistical Abstract of the United States*. This annual compendium compiles statistics issued by nearly every government agency as well as additional data from selected nongovernment organizations. Many are time series tables covering several years, or even decades. All tables indicate the source of the statistics. These sources can then be consulted if desired for even more comprehensive data. Check the library's catalog or ask the librarian to find these more specialized resources. Some may be available via FedStats.

FedStats is an online compilation of statistics provided by more than 70 U.S. agencies, including the Census Bureau, the Bureau of Labor Statistics, and the Federal Bureau of Investigation. Use the search option or the directory option to find the needed statistical tables. An especially useful feature of FedStats is the state and regional statistical data option.

No discussion of government statistical information would be complete without an examination of the U.S. Bureau of the Census. The Census Bureau is probably most well known for the Decennial Census of Population and Housing. The first such census was taken in 1790 to meet the Constitutional mandate for apportioning seats in Congress. It has been taken in every year ending in zero since that date. Now it is used not only to apportion seats in the House of Representatives, but also for allocation of federal aid to states and for a myriad of other purposes. The decennial census asks a certain core of questions of everyone. These are known as the 100%-questions. While these may vary slightly from census to census, age, race, gender, relationship, and Hispanic origin are fairly constantly collected. A longer questionnaire is sent to a sample of the population. These additional questions used in conjunction with the data from 100%-questions are used by government agencies at all levels, local planners, business and industry, schools, and social service agencies among others for planning, grant writing, economic development, and many other purposes.

To make census information easier to understand, the Census Bureau, in cooperation with local planning agencies, has created a multilevel mapping system. The entire country is mapped into small units called blocks. Data from 100%-questions is available for all blocks, but sample data is not. Both 100%-question data and sample data are available for census tracts (groups of blocks) and for larger mapped units such as cities, metropolitan statistical areas, counties, and states. Tracts are especially valuable to local level researchers because their boundaries remain mostly constant from census to census, thus allowing comparison. In cases where there is population growth, tracts may be split from one census to another and therefore may need to be added together to achieve comparable statistics. Metropolitan statistical areas, defined by the Office of Management and Budget, consist of a large population nucleus, together with adjacent communities having a high degree of social and economic integration with that core. Metropolitan statistical areas comprise one or more entire counties, except in New England, where cities and towns are the basic geographic units.

In addition to the decennial census, the Census Bureau conducts the economic census in years ending in two and seven, covering all areas of the economy from the national to the local level. Both the decennial census and the economic census are supplemented by numerous survey reports, including the new American Community Survey, initiated to provide more up-to-date information on American communities. In fact, the Census Bureau has proposed using the American Community Survey—instead of the long (sample) questionnaire used through 2000—in the next decennial census. For an overview of the many report topics available, see the Subjects A–Z listing on the Census Bureau website (<http://www.census.gov/>).

Mining Internal Sources

An organization's own internal historical data is often an underutilized source of information in the exploratory phase. Due to employee turnover, the researcher may lack knowledge that such historical data exist or, based on time or budget constraints and the lack of an organized archive, the researcher may choose to ignore such data. While digging through data archives can be as simplistic as sorting through a file containing past patient records or inventory shipping manifests, or rereading company reports and management-authored memos that have grown dusty with age, we will concentrate the remainder of our discussion on more sophisticated structures and techniques.

A **data warehouse** is an electronic repository for databases that organizes large volumes of data into categories to facilitate retrieval, interpretation, and sorting by end-users. The data warehouse provides an accessible archive to support dynamic organizational intelligence applications. The key words here are dynamically accessible. Data warehouses that offer archaic methods for data retrieval are seldom used. Data in a data warehouse must be continually updated to ensure that managers have access to data appropriate for real-time decisions. In a data warehouse, the contents of departmental computers are duplicated in a central repository where standard architecture and consistent data definitions are applied. These data are available to departments or cross-functional teams for direct analysis, or through intermediate storage facilities or **data marts** that compile locally required information. The entire system must be constructed for integration and compatibility among the different data marts.

SNAPSHOT

Will XP Release a Hailstorm of Abuse with Passport?

The Windows XP operating system is bundled with two software programs that the Electronic Privacy Information Center (EPIC), an advocacy group concerned with protecting personal privacy, feels oversteps a business's need to know. Passport is Microsoft's authentication technology that stores and reveals passwords as Web users access different sites. "Hailstorm will enable a person to store [personal information in a] vast database at Microsoft. [Hailstorm shares] personal information according to a user's instructions and [automates] making doctor's appointments, buying music online, booking airline tickets and handling other tasks." According to Microsoft, Hailstorm services "are oriented around people. They give users control of their own data and information, protecting personal information and requiring the consent of the individual with respect to who can access the information, what they can do with it, and how long they have that permission to do so." The executive director of EPIC, Marc Rotenberg, in his intention to file a complaint with the Federal Trade Commission, expressed concern that Microsoft "has not established itself as the Fort Knox of Internet privacy." Microsoft indicates the

changes in Windows XP are heralded by expectations that the desktop PC will less and less be the device chosen for accessing the Web as hand-held devices and cell phones become more commonplace. EPIC fears Windows XP's marriage with Passport and Hailstorm is bad news for those of us who want to protect our privacy. According to Bob Muglia, Microsoft group vice president of the .NET Services Group, "Hailstorm turns the industry debate over online privacy on its head. It starts with the fundamental assumption that the user owns and controls their personal information and is empowered to decide who gets to decide with whom they share any of their information and under what terms." Given that most large corporations and many smaller ones are developing extensive customer databases, what are the possible abuses that should concern researchers?

www.epic.org

www.microsoft.com

www.ftc.gov

The term **data mining** describes the process of discovering knowledge from databases stored in data marts or data warehouses. The purpose of data mining is to identify valid, novel, useful, and ultimately understandable patterns in data.¹² Similar to traditional mining, where we search beneath the surface for valuable ore, data mining searches large databases for indispensable information for managing an organization. Both require sifting a large amount of material to discover a profitable vein. Data mining is a useful tool, an approach that combines exploration and discovery with confirmatory analysis.

The more accessible the databases that comprise the data warehouse, the more likely a researcher will use such databases to reveal patterns. Thus researchers are more likely to mine electronic databases than paper ones. It will be useful to remember that data in a data warehouse were once primary data, collected for a specific purpose. When researchers data-mine a company's data warehouse, all the data contained within that database have become secondary data. The patterns revealed will be used for purposes other than those originally intended. For example, in an archive of sales invoices, we have a wealth of data about what was sold, how much of each item or service, at what price level, to whom, and where and when and how the products were shipped. Initially the company generated the sales invoice to facilitate the process of getting paid for the items shipped. When a researcher mines that sales invoice archive, the search is for patterns of sales, by product, category, region of the country or world, price level, shipping methods, etc. Therefore, data mining forms a bridge between primary and secondary data.

Traditional database queries are unidimensional and historical—for example, "How much beer was sold during December 1999 in the Sacramento area?" In contrast, data mining attempts to discover patterns and trends in the data and to infer rules from these patterns. For example, an analysis of retail sales by Sacramento FastShop identified products that are often purchased together—like beer and diapers—although they may appear to be unrelated. With the rules discovered from the data mining, a manager is able to support, review, and/or examine alternative courses of action for solving a management dilemma, alternatives which may later be studied further in the collection of new primary data.

Evolution of Data Mining

The complex algorithms used in data mining have existed for more than two decades. The U.S. government has employed customized data-mining software using neural networks, fuzzy logic, and pattern recognition to spot tax fraud, eavesdrop on foreign

Numerous companies build large consumer purchase behavior databases by collecting purchases made via store-owned credit programs or frequent purchase cards not linked directly with payment plans. Studying such data can reveal the likely success of a new product introduction or the sales lift effect of a coupon drop.



EXHIBIT 10-9 The Evolution of Data Mining

Evolutionary Step	Investigative Question	Enabling Technologies	Characteristics
Data collection (1960s)	"What was my average total revenue over the last five years?"	Computers, tapes, disks	Retrospective, static data delivery
Data access (1980s)	"What were unit sales in California last December?"	Relational databases (RDBMS), structured query language (SQL), ODBC	Retrospective, dynamic data delivery at record level
Data navigation (1990s)	"What were unit sales in California last December? Drill down to Sacramento."	Online analytic processing (OLAP), multidimensional databases, data warehouses	Retrospective, dynamic data delivery at multiple levels
Data mining (2000)	"What's likely to happen to Sacramento unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Prospective, proactive information delivery

communications, and process satellite imagery.¹³ Until recently, these tools have been available only to very large corporations or agencies due to their high costs. However, this is rapidly changing.

In the evolution from *business data to information*, each new step has built on previous ones. For example, large database storage is crucial to the success of data mining. The four stages listed in Exhibit 10-9 were revolutionary because each allowed new management questions to be answered accurately and quickly.¹⁴

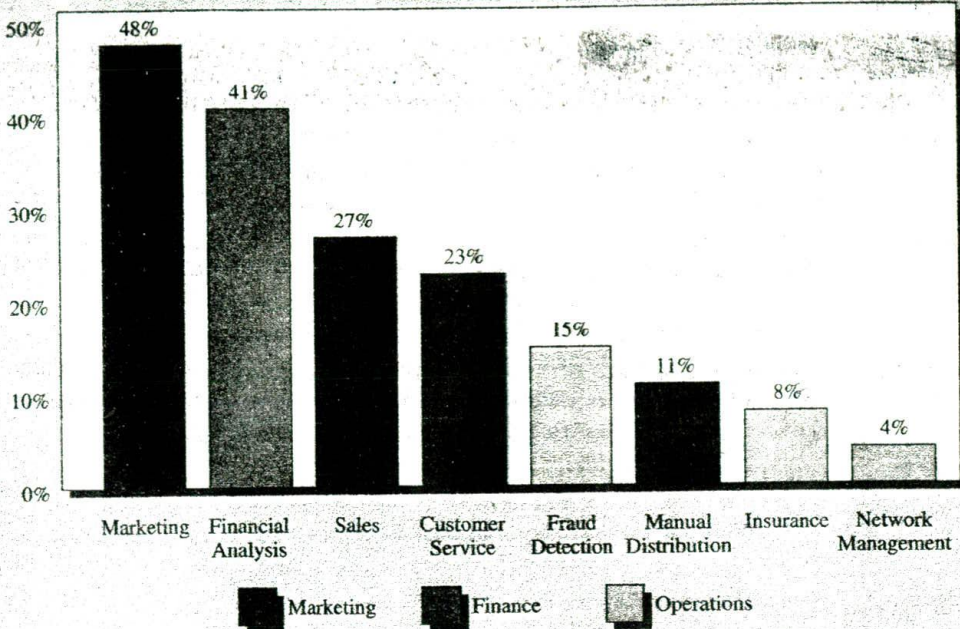
The process of extracting information from data has been done in some industries for years. Insurance companies often compete by finding small market segments where the premiums paid greatly outweigh the risks. They then issue specially priced policies to a particular segment with profitable results. However, two problems have limited the effectiveness of this process: Getting the data has been both difficult and expensive, and processing it into information has taken time—making it historical rather than predictive. Now, instead of incurring high data collection costs to resolve management questions, secondary data are available to assist the manager's decision making. It was State Farm Insurance's ability to mine its extensive nationwide database of accident locations and conditions at intersections that allowed it to identify high risk intersections and then plan a primary data study to determine alternatives to modify such intersections.

Functional areas of management and select industries are currently driving data-mining projects: marketing, customer service, administrative/financial analysis, sales, manual distribution, insurance, fraud detection, and network management (see Exhibit 10-10).¹⁵ Data-mining technology provides two unique capabilities to the researcher or manager: pattern discovery and prediction.

Pattern Discovery Data-mining tools can be programmed to sweep regularly through databases and identify previously hidden patterns. An example of pattern discovery is the detection of stolen credit cards based on analysis of credit card transaction

See the State Farm: Dangerous Intersections case in the case section of this text.

EXHIBIT 10-10 Data Mining in Business



records. MasterCard processes 12 million transactions daily and uses data mining to detect fraud.¹⁶ Other uses include finding retail purchase patterns (used for inventory management), identifying call center volume fluctuations (used for staffing), and locating anomalous data that could represent data entry errors (used to evaluate the need for training, employee evaluation, or security).

Predicting Trends and Behaviors A typical example of a predictive problem is targeted marketing. Using data from past promotional mailings to identify the targets most likely to maximize return on investment, future mailings can be more effective. Bank of America and Mellon Bank both use data-mining software to pinpoint marketing programs that attract high-margin, low-risk customers. Bank of America focuses on credit lines and retail lending; Mellon Bank has used data mining to optimize its home equity line of credit marketing to existing customers.¹⁷ Other predictive problems include forecasting bankruptcy and loan default and finding population segments with similar responses to a given stimulus. Data-mining tools also can be used to build risk models for a specific market, such as discovering the top 10 most significant buying trends each week (see Exhibit 10-11).

Data-Mining Techniques¹⁸

An understanding of statistics is essential to the data-mining process. Data-mining tools perform exploratory and confirmatory statistical analysis to discover and validate relationships. Data-mining tools even extend confirmatory statistical approaches by allowing the automated examination of large numbers of hypotheses. Suppose there are 12 variables in the MindWriter survey and we have a process whose outcome can be predicted when three variables are in a particular range. But we are unfamiliar with the repair process and don't know which variables are relevant. With this small problem

EXHIBIT 10-11 Sample Data-Mining Applications

Company	Use of Data Mining
Blue Cross Blue Shield	Mines data on member and provider questions about insurance claims and patient history, as well as inquiry/call history to improve call center staffing and training. ¹
BMW	Uses data mining in the execution of its "stochastic crash-simulation technology," enabling "design improvements to the crashworthiness and safety of automobiles" while reducing product design cycle time. ⁵
Coca-Cola FEMSA	This Mexican bottler and distributor to Mexico, Argentina, and Latin America uses data mining to decrease stock-outs, reduce inventory, minimize variability in operations, and manage promotions. ²
Colgate-Palmolive Company	Operating in more than 200 countries, it uses data mining to improve forecasts, production plans, and customer order fulfillment, as well as reduce overall order cycle times and costs. ²
Dow Chemical	Its data warehouse stores news articles about clients, as well as a complete purchase history; data mining is used to improve sales visit effectiveness. ¹
Hewlett-Packard	Used data mining to identify every World Trade Center customer and its systems and components at the time of the September 11, 2001, collapse. These components were matched with on-hand inventory in warehouses, in preparation to replace destroyed systems. ¹
IBM	Helps biologists make sense of the drastically increased amount of data that can be brought to bear on any biological or medical question. Its <i>GeneMine</i> was "designed to help scientists rapidly infer, validate, and propose experimental tests for the likely functions of unknown genes." ⁴
Meineke Discount Muffler Shops	This 900-plus franchised nationwide retail service provider adds more than 5,000 customer and service records each week to an extensive data warehouse it then mines using MapInfo's MapMarker to make more effective advertising decisions and identify new franchise locations. ⁷
The Gillette Company	Partnered with PricewaterhouseCoopers, it uses data mining to plan on-location promotional store displays for its German retailers. ³
Wal-Mart Stores	Uses data mining to design and modify store layouts to reduce shoplifting and theft. ⁶

¹Extracted from "Software to Track Customers' Needs Helped Firms React," *New York Times on the Web*, October 1, 2001 (<http://www.nytimes.com/2001/10/01/technology/01CRM.html>).

²Extracted from supply-chain management special advertising section, *Business Week*, October 8, 2001.

³Extracted from SAP Customer Success Stories (http://www.mysap.com/solutions/crm/customersuccesses.asp/Gillette_50050088.pdf).

⁴Extracted from "The GeneMine System for Genome/Proteome Annotation and Collaborative Data Mining," *IBM Systems Journal* 40, no. 2, 2001 (<http://www.research.ibm.com/journal/sj/402/lee.html>).

⁵"BMW Pioneers Stochastic Crash Simulation for Improved Vehicle Safety," Silicon Graphics Inc. (<http://www.sgi.com/manufacturing/success/bmw.html>).

⁶"NCR More than Doubles Data Warehouse for World's Leading Retailer to Over 100 Terabytes" (<http://hpcwire.com/dsstart/99/0824/100966.html>).

⁷S. Reese, "Bad Mufflers Make Good Data," *American Demographics*, November 1998, pp. 42-44.

there are $12 \times 11 \times 10 = 1,320$ combinations. If you spent a minute examining a plot of each pair of variables, you could easily spend 22 hours on the problem.

Numerous techniques are used in data mining; often they are used together. The type of data available and the nature of information sought determine the technique used. Here we explore the first five techniques listed below and mention several others (to be covered in later chapters):

- Data visualization
- Clustering
- Neural networks
- Tree models
- Classification
- Estimation
- Association
- Market-basket analysis
- Sequence-based analysis
- Fuzzy logic
- Genetic algorithms
- Fractal-based transformation

Data Visualization By viewing aggregated data on multiple dimensions (e.g., product, brand, date of sale, and region), both the analyst and the end-user gain a deeper, more intuitive understanding of the data in picture form. A multidimensional database typically contains three axes: (1) *dimensions*, like the fields in a table; (2) *measurements*, aggregate computations to be viewed; and (3) *hierarchies*, which impose structure on the dimensions. The set {month, quarter, year} is a time-based hierarchy.¹⁹ Using this approach, the researcher views the data at various levels (“drill down/drill up”). Starting with a total of sales of breakfast cereal by region, the researcher observes that one region is more profitable than others. Next, she drills down to sales by store and discovers that one store is outperforming all others. Looking deeper yet reveals that this store spends the most on training warehousing personnel.

Data visualization uses two types of interactive queries. The first type is usually available to the end-user via a query language. The second, for complex queries that go beyond traditional two-dimensional row and column data analysis, uses online analytical processing (OLAP) tools. Multidimensional analysis, another phrase for OLAP, provides fast, flexible data summarization, analyses, and reporting capabilities with the ability to view trends over time. So, complex queries such as, “How much coffee did we sell last month in Sacramento, compared to the same month last year?” can be answered quickly.

Clustering **Clustering** enables the researcher to segment a population. This approach assigns each data record to a group or segment. The assignment process is performed automatically by clustering algorithms that identify the similar characteristics in the data set and then partition them into groups often referred to as the “nearest neighbors.”

Clustering is often used as the first step in data mining. For example, clustering may be used to segment a customer database for further analysis of customers’ buying habits to decide which segments to target for a new sales campaign.

You may want to review data-mining software and download the demo for SPSS (<http://www.spss.com/clementine/newshow/sld003.htm>) or explore the SAS Enterprise Miner (<http://www.sas.com/service/consult/usconsult/miningcompaq.html>). Also visit Compaq’s Advanced Data Mining Center (<http://www.compaq.com>, keyword search: data mining, or <http://www.sas.com/partners/directory/compaq/index.html>).

We discuss cluster analysis in more detail in Chapter 19.

Neural Networks Neural networks, or *artificial neural networks (ANN)*, are collections of simple processing nodes that are connected. Each node operates only on its local data and on the inputs it receives via the connections. The result is a nonlinear predictive model that resembles biological neural networks and learns through training.

The neural model has to train its network on a training data set. One drawback is that no explanation of the results is available. Neural networks are best used where a predictive model is more useful than an explanatory model. For database marketing, a neural network can be constructed that predicts whether a specific person is likely to purchase a particular product. This enables the marketing organization to be very specific in its target marketing, reducing costs and dramatically improving sales “hits.”

Tree Models This technique segregates data by using a hierarchy of if-then statements based on the values of the variables and creates a tree-based structure that represents the segregation decisions. ~~Decision~~ **tree models** are faster and easier to understand than neural networks, but the data must be interval or categorical. Specific ~~decision tree methods include~~ classification and regression trees (CART) and chi-square automatic interaction detection (CHAID), a type of Automatic Interaction Detection model (see Chapter 16, Exhibit 16-20).

Classification Classification uses a set of preclassified examples to develop a model that can classify the population of records at large. Fraud detection and risk assessments of credit applications are particularly well suited to this type of analysis. Classification frequently employs decision trees or neural network-based classification algorithms (see descriptions, below).

Classification begins with training the software with a set of preclassified sample transactions. For a fraud detection application, this would include complete records of both fraudulent and valid activities. The algorithm uses these cases as criteria to set the parameters for proper discrimination. Once developed, the model can correctly classify new records into the same predefined classes. For example, a model capable of classifying loan applicants may generate a rule stating, “If applicant earns \$45,000, is between 35 and 45 years old, and lives in a specific zip code, then the applicant is a good credit risk.” **Estimation** is a variation of classification. Instead of using a binary classifier (e.g., a loan applicant is a good risk or a bad risk), estimation generates a score (e.g., of creditworthiness), based on a prescored training set.

Other Mining Techniques **Association** is the process used to recognize and understand patterns in the data. The goal is to find, across large numbers of small transactions, trends that can be used to understand and exploit natural buying patterns.

The most common form of association is **market-basket analysis**. A classic example of market-basket analysis was the discovery that beer and diapers are often purchased during the evening in the same transaction. Presumably this is not because babies like beer before bed, but because the baby’s father buys the diapers and chooses to purchase beer at the same time. Information from market-basket analysis can lead stores to change their layout, adjust their inventories, or introduce a targeted promotional campaign. In the financial sector, association can be used to analyze customers’ accounts. The patterns identified may be used to create a “bundle” of service offerings.

Results from association analyses are often expressed in terms of confidence, such as, “Seventy-five percent of all transactions in which soft drinks were purchased also included potato chips.” The results of the association analysis often trigger additional analysis to help the manager understand why the association exists.

A variant of traditional market-basket analysis, **sequence-based analysis** ties together a series of activities or purchases (for example: using an account number, a credit card, or a frequent-flyer number). The association algorithm takes into account not only the combination of items but also the order of the items. Rules derived from these relationships can be used to predict a specific purchase based on previous purchases. In health care, such methods can be used to predict the course of a disease and order preventive care.

Fuzzy logic, genetic algorithms, and fractal transforms are used in more complex mining operations. **Fuzzy logic**, an extension of conventional (Boolean) logic, handles the concept of partial truth—truth values between “completely true” and “completely false.” **Genetic algorithms** are optimization techniques that use processes analogous to mutation, natural selection, and genetics for search and identification of meaningful relationships. Fractal geometry was originally applied to the compression of topographic images. It is a mathematical means of compressing data with no data. Models using **fractal-based transformation** can work on many gigabytes of data, offering the possibility of identifying tiny subsets of data that have common characteristics. For example, fractal-based transformations would enable a researcher to discover the 5 customers out of 5 million who responded to a Neiman Marcus catalog offering.

Data-Mining Process

Data mining, as depicted in Exhibit 10–12, involves a five-step process:²⁰

- **Sample:** Decide between census and sample data.
- **Explore:** Identify relationships within the data.
- **Modify:** Modify or transform data.
- **Model:** Develop a model that explains the data relationships.
- **Assess:** Test the model’s accuracy.

To better visualize the connections between the techniques just described and the process steps listed in this section, you may want to download a demonstration version of data-mining software from the Internet.

For an appropriate sampling technique, choose from among those described in Chapter 7.

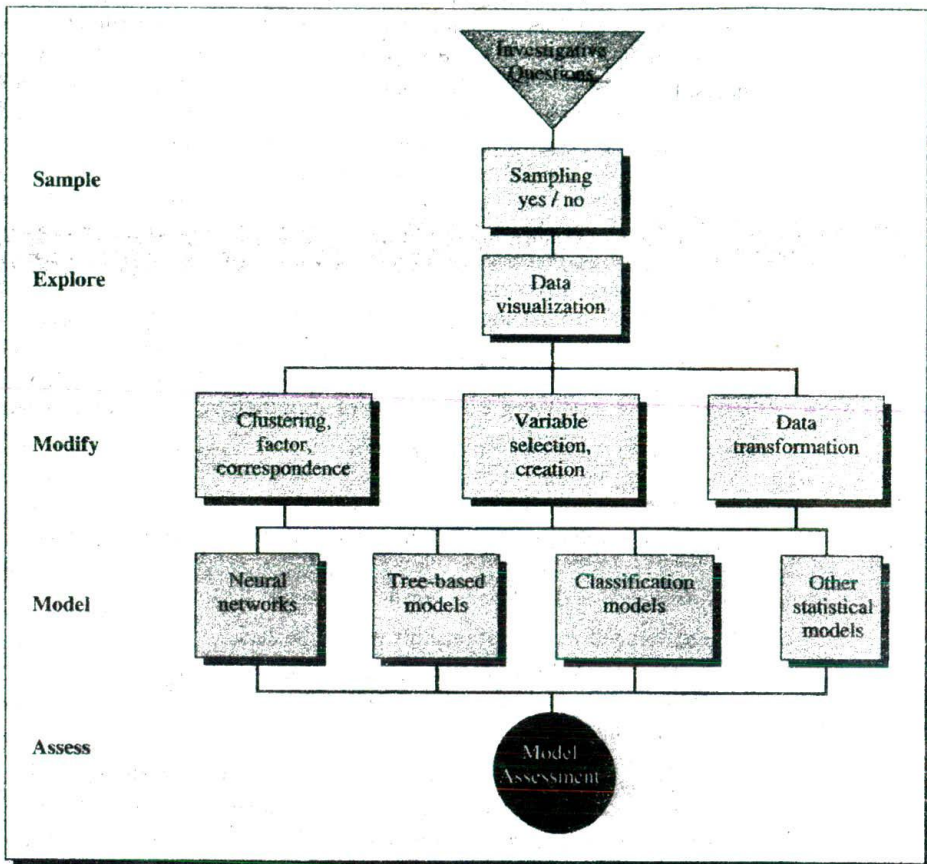
Sample Exhibit 10–12 suggests that the researcher must decide whether to use the entire data set or a sample of the data.²¹ If the data set in question is not large, if the processing power is high, or if it is important to understand patterns for every record in the database, sampling should not be done. However, if the data warehouse is very large (terabytes of data), the processing power is limited, or speed is more important than complete analysis, it is wise to draw a sample. In some instances, researchers may use a data mart for their sample—with local data that are appropriate for their geography. Alternatively, the researcher may select an appropriate sampling technique. Since fast turnaround for decisions is often more important than absolute accuracy, sampling is appropriate.

We explore EDA techniques in Chapter 16.

If general patterns exist in the data as a whole, these patterns will be found in the sample. If a niche is so tiny that it is not represented in a sample yet is so important that it influences the big picture, it will be found using exploratory data analysis (EDA).

Explore After the data are sampled, the next step is to explore them visually or numerically for trends or groups. Both visual and statistical exploration (data visualization) can be used to identify trends. The researcher also looks for outliers to see if the data need to be cleaned, cases need to be dropped, or a larger sample needs to be drawn.

EXHIBIT 10-12 Data-Mining Process



Modify Based on the discoveries in the exploration phase, the data may require modification. Clustering, fractal-based transformation, and the application of fuzzy logic are completed during this phase as appropriate. A data reduction program, such as factor analysis, correspondence analysis, or clustering, may be used (see Chapter 19). If important constructs are discovered, new factors may be introduced to categorize the data into these groups. In addition, variables based on combinations of existing variables may be added, recoded, transformed, or dropped.

At times, descriptive segmentation of the data is all that is required to answer the investigative question. However, if a complex predictive model is needed, the researcher will move to the next step of the process.

Model Once the data are prepared, construction of a model begins. Modeling techniques in data mining include neural networks as well as decision tree, sequence-based, classification and estimation, and genetic-based models.

Assess The final step in data mining is to assess the model to estimate how well it performs. A common method of assessment involves applying a portion of data that was not used during the sampling stage. If the model is valid, it will work for this “holdout” sample. Another way to test a model is to run the model against known data. For example, if you know which customers in a file have high loyalty and your model predicts loyalty, you can check to see whether the model has selected these customers accurately.

SUMMARY

1 The exploratory phase of the research process uses secondary data to expand understanding of the management dilemma, look for ways others have addressed and/or solved problems similar to the management dilemma or management question, and gather background information on the topic to refine the research question.

2 Researching secondary sources is complex and challenging. There are two categories of sources available (external and internal) and three types of sources (primary, secondary, and tertiary). Primary sources are original works of research or raw data without interpretation. Secondary sources are interpretations of primary data. Tertiary sources may be interpretations of secondary sources or, more commonly, finding aids such as indexes, bibliographies, and Internet search engines.

3 There are generally five types of information sources used in most literature searches, including indexes and bibliographies, dictionaries, encyclopedias, handbooks, and directories. Each is useful to a literature search in a variety of ways. One of the harder tasks associated with using secondary sources is evaluating the quality of the information. Five factors to consider when evaluating the quality of the source are purpose, scope, authority, audience, and format.

4 The process for searching bibliographic databases applies to both print and online sources.

1. Select a database appropriate to your topic.
2. Construct a search query (or search statement).
 - Review and evaluate search results.
 - Modify the search query, if necessary.
3. Save those valuable results of your search.
4. Retrieve articles not available in the database.
5. Supplement your results with information from Web sources.

Many online and Web-based sources use Boolean logic for constructing search queries, but protocols do differ. One purpose for reviewing the results of your original search is to modify it with newly discovered information. The researcher should check the bibliographies of cited works, check the subject headings assigned to the extracted articles, and search for works by referenced authors.

5 The basic steps for searching Web-based sources include a critical last step that novice researchers often skip.

1. Select a search engine or directory.
2. Determine your search options.

3. Construct a search query and enter your search term(s).
 - Review and evaluate search results.
 - Modify your search query, if necessary.
4. Save those valuable results of your search.
5. Supplement your results with information from non-Web sources.

When doing a Web-based search, several options are available: known-item searches, “who” searches, “where” searches, and “what” searches. Due to the special characteristics of each type of search, each starts with a different search strategy. Several special sites have evolved to offer the researcher assistance for each type of search. These can be found in Appendix A.

U.S. government data are exceptionally detailed and readily available both in print and on the Web. All government agencies provide detailed studies and often make these results available in full text or in summary tables. The U.S. Census Bureau generates the most complete databases on the U.S. population, being most well known for the decennial census. The Census Bureau has partnered with local agencies to develop a multilevel mapping system for its data, making the bureau an even more valuable source of data.

6 Managers faced with current decisions requiring immediate attention often overlook internal data in a company’s data warehouse. Data mining refers to the process of discovering knowledge from databases. Data-mining technology provides two unique capabilities to the researcher or manager: pattern discovery and the prediction of trends and behaviors. Data-mining tools perform exploratory and confirmatory statistical analyses to discover and validate relationships. These tools even extend confirmatory statistical approaches by allowing the automated examination of large numbers of hypotheses. The type of data available and the nature of information sought determine which of the numerous data-mining techniques to select. Data mining involves a five-step process: sample, explore, modify, model, and assess.

KEY TERMS

- | | | |
|----------------------------|----------------------------------|-------------------------------------|
| association 306 | encyclopedias 283 | primary sources 282 |
| bibliographies 283 | estimation 306 | search query (search statement) 288 |
| bibliographic database 287 | exploratory research 281 | secondary sources 282 |
| classification 306 | fractal-based transformation 307 | sequenced-based analysis 307 |
| clustering 305 | fuzzy logic 307 | source evaluation 284 |
| data marts 300 | genetic algorithms 307 | purpose 284 |
| data mining 301 | handbook 283 | scope 284 |
| data visualization 305 | indexes 283 | authority 286 |
| data warehouse 300 | literature search 281 | audience 286 |
| decision tree models 306 | market-basket analysis 306 | format 287 |
| dictionaries 283 | neural networks 306 | tertiary sources 282 |
| directories 284 | portal 293 | |

EXAMPLES

Company	Scenario	Page
Advertising World (University of Texas)	Developing a portal to advertising websites of interest to students and practitioners.	296
Bank of America	Using data-mining software to pinpoint marketing programs that attract high-margin, low-risk customers.	303
Blue Cross Blue Shield	Using data mining to staff and train customer service call centers.	304
BMW	Using data mining to design improvements to the crashworthiness and safety of automobiles.	304
Charles Schwab	Describes the Internet's importance to investor research.	286
CNNfn	Design of a website for a special-interest audience.	286
Coca-Cola FEMSA	Using data mining for inventory, production, and promotion management in Mexico, Argentina, and Latin America.	304
Colgate-Palmolive Company	Using data mining to improve forecasting, production, and customer order fulfillment.	304
Croy and Associates*	Conducting secondary research to prepare for an interview with a potential research client.	BRTL
Dow Chemical	Using data mining to increase sales call effectiveness.	304
Hewlett-Packard	Using data mining to respond to the September 11, 2001, World Trade Center collapse.	304
IBM	Using its <i>GeneMine</i> technology to help biologists develop better cures and treatments for disease.	304
MasterCard	Detection of stolen credit cards based on data mining of credit card transactions.	303
Meineke Discount Muffler Shops	Using data mining to plan more effective advertising, as well as select new locations.	304
Mellon Bank	Using data mining to optimize home equity line of credit marketing to existing customers.	303
MindWriter*	Finding variables of association to predict repair success.	303
Neiman Marcus	Fractal-based transformations to discover likely customers for catalog merchandise.	307
Sacramento FastShop	Unidimensional and historical query in data mining; finding and explaining combination purchases.	301
Space Imaging	Using multiple sources of geospatial data to create map-accurate and information-rich products for government and industry.	294
State Farm Insurance	Using data mining to reveal the most accident-prone intersections within the United States and Canada.	302
The Gillette Company (with Pricewaterhouse- Coopers LLP)	Using data mining to plan on-location promotional store displays for German retailers.	304

U.S. government	Using data mining to detect tax fraud, eavesdrop on foreign communications, and process satellite imagery.	295
Wal-Mart Stores	Using data mining to design and modify store layouts to reduce shoplifting and theft.	304

*Due to the confidential and proprietary nature of most research, the names of some companies have been changed.

DISCUSSION QUESTIONS

Terms in Review

1. Explain how each of the five evaluation factors for a secondary source influences its management decision-making value.
 - a. Purpose
 - b. Scope
 - c. Authority
 - d. Audience
 - e. Format
2. Define the distinctions between primary, secondary, and tertiary sources in a secondary search.
3. What is data mining?
4. Explain how internal data-mining techniques differ from a literature search.
5. Some researchers find that their sole sources are secondary data. Why might this be? Name some management questions for which secondary data sources are probably the only ones feasible.

Making Research Decisions

6. Assume you are asked to investigate the use of mathematical programming in accounting applications. You decide to depend on secondary data sources. What search tools might you use? Which do you think would be the most fruitful? Sketch a flow diagram of your search sequence.
7. What problems of secondary data quality must researchers face? How can they deal with them?
8. Below are a number of requests that a staff assistant might receive. What specific tools or services would you expect to use to find the requisite information? (Hint: Use Appendix A and the CD that accompanies this text.)
 - a. The president wants a list of six of the best references on executive compensation that have appeared during the last year.
 - b. Has the FTC published any recent statements (within the last year) concerning its position on quality stabilization?
 - c. I need a list of the major companies located in Greensboro, North Carolina.
 - d. Please get me a list of the directors of General Motors, Microsoft, and Morgan Stanley & Co.
 - e. Is there a trade magazine that specializes in the flooring industry?
 - f. I would like to track down a study of small-scale service franchising that was recently published by a bureau of business research at one of the southern universities. Can you help me?

Bringing Research to Life

9. Using Exhibits 10-4, 10-5, and 10-6, state the research question and then the search plan that Jason should have conducted before his meeting with Armand Croyard.
10. What government sources should be included in Jason's search?

From Concept to Practice

11. Using Exhibits 10-4, 10-5, and 10-6, state a research question and then plan a bibliographic and Web search.

WWW Exercises

Visit our website for Internet exercises related to this chapter at www.mhhe.com/business/cooper8

CASES

A GEM OF A STUDY

KNSD SAN DIEGO

MCDONALD'S TESTS CATFISH SANDWICH

OUTBOARD MARINE CORPORATION

PEBBLE BEACH CO.

RAMADA DEMONSTRATES ITS PERSONAL BEST

STATE FARM: DANGEROUS INTERSECTIONS

THE CATALYST FOR WOMEN IN FINANCIAL SERVICES

*All cases indicating a video icon are located on the Instructor's Videotape Supplement. All nonvideo cases are in the case section of the textbook. All cases indicating a CD icon offer a data set, which is located on the accompanying CD.

REFERENCE NOTES

1. Faith McLellan, "1966 and All That—When Is a Literature Search Done?" *The Lancet* 358 (August 25, 2001), p. 646.
2. Ad in *Business Week*, June 28, 1999, p. 3.
3. "Lots of online brokers say they are number one. Here's what some of the industry sources have to say," Schwab.com, January 1, 2001 (<http://www.schwab.com/SchwabNOW/SNLibrary/SNLib132/SN132mainAwardsAndRatingsHome/>).
4. For general information on evaluating reference sources, see William A. Katz, *Introduction to Reference Work*, 7th ed., McGraw-Hill, 1997. For more information about evaluating Web resources, see "Checklist for a Business/Marketing Web Page" (Widener University, Wolfgram Memorial Library. Compiled by Jan Alexander and Marsh Tate, 1999; <http://www2.widener.edu/Wolfgram-Memorial-Library/busmark.htm>); "Evaluating World Wide Web Resources" (Indiana University Libraries, 1999; http://www.indiana.edu/~libinstr/worksheets/evaluating_web.html); and the metasite "Evaluating Information Sources" (Alastair Smith, 2000; <http://www.vuw.ac.nz/~agsmith/evaln/evaln.htm>).
5. Good sources for Web size estimates are the studies by Steve R. Lawrence and C. Lee Giles, "Searching the World Wide Web," *Science* 280 (April 1998), pp. 98–100, and "Accessibility of Information on the Web," *Nature* 400 (July 8, 1999), pp. 107–9, with updated summary data at <http://www.wwwmetrics.com>.
6. Michael Dahm, "Counting Angels on a Pinhead: Critically Interpreting Web Size Estimates," *Online* 24 (January/February 2000), pp. 35–44. This article further interprets the pioneering research by authors Steve R. Lawrence and C. Lee Giles (op. cit.).
7. The May/June 1999 issue of *Online* focuses on search engine technology. See, for example, Danny Sullivan, "Crawling under the Hood: An Update on Search Engine Technology," *Online* 23 (May/June 1999), pp. 30–38. See also Danny Sullivan's "Search Engine Watch" (<http://searchenginewatch.com/>) and Greg Nottess's "Search Engine Showdown" (<http://www.nottess.com/search/>) for current information about search engines and their features.
8. Greg R. Nottess, "Duplicative Databases: Yellow Pages from infoUSA," *Database* 22 (February/March 1999), pp. 73–76.
9. See, for example, the June 22, 1999, issue of *PC Magazine* for a special report detailing "Ten Trends That Are Defining the Future" (Introduction, p. 100).
10. "Federal Register Online Via GPO Access" (http://www.access.gpo.gov/su_docs/aces/aces140.html).
11. "About the Code of Federal Regulations" (<http://www.access.gpo.gov/nara/about-ctr.html#page1>).
12. R. Srikanth and R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance Improvements," *Proceedings, 5th International Conference on Extending Database Technology*, Paris, France, March 1996.
13. B. DePompe, "There's Gold in Databases," *CMP Publications*, January 8, 1996; website: <http://techweb.cmp.com/iwk>.
14. Table adapted from DIG White Paper 95/01, "An Overview of Data Mining at Dun & Bradstreet," Data Intelligence Group, Pilot Software, Cambridge, MA (September 1995).
15. "Data Mining: Plumbing the Depths of Corporate Databases," *Computerworld Customer Publication*, insert to *Computerworld*, April 21, 1997, p. 12.
16. DePompe, *CMP Publications*; website: <http://techweb.cmp.com/iwk>.
17. "Data Mining: Plumbing the Depths of Corporate Databases," *ComputerWorld*, pp. 6, 18.
18. The section on data-mining techniques was adapted from Bruce Moxon, "Defining Data Mining," DBMS Data Warehouse Supplement, *DBMS Online* (August 1996), <http://www.dbmsmag.com>; Mark Kantrowitz, Erik Horstkotte, and Cliff Joslyn, "Fuzzy Logic FAQ," <http://comp.ai.fuzzynews-group>; DIG White Paper 95/01; and Michael Bell, "A Data Mining FAQ," Quintillion Corporation, Web page: <http://www.qwhy.com/dmfaq.htm>.
19. Information Discovery Inc., website: <http://www.idi.com>.
20. SAS Institute Inc., "Data Mining," website: <http://www.sas.com>.
21. Exhibit 10–12 was adapted from SAS Institute Inc., "Data Mining," website: <http://www.sas.com>.

REFERENCES FOR SNAPSHOTS AND CAPTIONS

Microsoft

"Microsoft Announces Hailstorm, a New Set of XML Web Services Designed to Give Users Greater Control." Microsoft press release, March 19, 2001 (<http://www.microsoft.com/presspass/features/2001/mar01/03-19hailstorm.asp>).

Steve Lohr, "Privacy Group Is Taking Issue with Microsoft," *The New York Times on the Web*, July 25, 2001 (<http://www.nytimes.com/2001/07/25/technology/ebusiness/25COMP.html?todayshdlines>).

nytimes.com/2001/07/25/technology/ebusiness/25COMP.html?todayshdlines).

Space Imaging

"The Company," Spaceimaging.com (January 1, 2002) (http://www.spaceimaging.com/newsroom/press_kits/factsheet.htm).

CLASSIC AND CONTEMPORARY READINGS

Berry, Michael J. A., and Gordon Linoff. *Mastering Data Mining: The Art and Science of Customer Relationship Management*. New York: John Wiley & Sons, 1999.

Fayyad, U. M., and G. Piatetsky-Shapiro. *Advances in Knowledge Discovery and Data Mining*. Cambridge, MA: AAAI Press/MIT Press, 1996. An excellent text that provides an overview

of knowledge discovery and data mining using statistical methods.

Katz, William A. *Introduction to Reference Work*. 7th ed. New York: McGraw-Hill, 1997.

Woy, James (ed.). *Encyclopedia of Business Information Sources*. 14th ed. Gale Group, 2000.

THE UNIVERSITY OF CHICAGO
LIBRARY

1950

1951

1952

1953

Survey Methods: Communicating with Participants

Learning Objectives

After reading this chapter, you should understand . . .

- 1 The process for selecting the appropriate and optimal communication approach.**
- 2 What factors affect participation in communication studies.**
- 3 The major sources of error in communication studies and how to minimize them.**
- 4 The major advantages and disadvantages of the three communication approaches.**
- 5 Why an organization might outsource a communication study.**