Chapter Vignette: Business Facts on a Grand Scale

A key problem that faces any business research manager is the need to constantly capture relevant data about customers, competitors, and/or market characteristics. The use of secondary data (i.e., data that has been collected previously for other purposes) has exploded with the advent of large-scale electronic information sources and the Web. One company that has taken full advantage of integrating various business-related information sources is Nielsen Claritas.

Prior to its merge with the Nielsen Company, Claritas (which in Latin means “brightness”) had a 40-year history of collecting and integrating business-related data from different sources. Its products include (1) PRIZM, which provides market segmentation information based upon consumer behavior and geographic location; (2) Consumer Point, a target marketing analysis solution for different industry spaces; and (3) Business-Facts, which provides accurate business data for market support and strategic planning.

Business-Facts holds great promise as a secondary data source for existing companies. Using Standard Industrial Classification (SIC) and North American Industry Classification (NAICS) codes developed through the Census Bureau, characteristics on business ownership, location, employment, and sales are available for 10 major industrial groupings. Data and employee counts within the Business-Facts system represent over 13 million businesses. Examples of these industry groups include construction, manufacturing, and retail sales establishments across the United States. Since business information can become quickly obsolete, Nielsen Claritas spends millions of dollars each year to verify business information on a quarterly basis.

The advantages of knowing broadly both the characteristics and location of major customer groups (or potential competitors) are very real. Using a sophisticated statistical modeling approach, the Nielsen Claritas company can link your customers to your existing or proposed locations, in a fashion such that the information is as timely and applicable as possible.

All of the information sources within Nielsen Claritas add value to business users by satisfying two very critical needs. First, Nielsen Claritas has expertise in linking different data streams into a cohesive system. This allows users to answer through secondary data sources critical existing business questions. Secondly, their information systems are geographically based, so that businesses can query data to a common point on the globe.

 Truly, the integration and utilization of secondary data sources by the Nielsen Claritas Company has put business research “on the map”—both literally and figuratively!
**Introduction**

Research projects often begin with secondary data, which are gathered and recorded by someone else prior to (and for purposes other than) the current project. Secondary data usually are historical and already assembled. They require no access to respondents or subjects.

**Advantages of Secondary Data**

The primary advantage of secondary data is their availability. Obtaining secondary data is almost always faster and less expensive than acquiring primary data. This is particularly true when researchers use electronic retrieval to access data stored digitally. In many situations, collecting secondary data is instantaneous.

Consider the money and time saved by researchers who obtained updated population estimates for a town during the interim between the 2000 and 2010 censuses. Instead of doing the fieldwork themselves, researchers could acquire estimates from a firm dealing in demographic information or from sources such as Claritas or PCensus. As in this example, the use of secondary data eliminates many of the activities normally associated with primary data collection, such as sampling and data processing.

Secondary data are essential in instances when data cannot be obtained using primary data collection procedures. For example, a manufacturer of farm implements could not duplicate the information in the Census of Agriculture because much of the information there (for example, amount of taxes paid) might not be accessible to a private firm.

**Disadvantages of Secondary Data**

An inherent disadvantage of secondary data is that they were not designed specifically to meet the researchers’ needs. Thus, researchers must ask how pertinent the data are to their particular project. To evaluate secondary data, researchers should ask questions such as these:

- Is the subject matter consistent with our problem definition?
- Do the data apply to the population of interest?
- Do the data apply to the time period of interest?
- Do the secondary data appear in the correct units of measurement?
- Do the data cover the subject of interest in adequate detail?

Even when secondary information is available, it can be inadequate. Consider the following typical situations:

- A researcher interested in forklift trucks finds that the secondary data on the subject are included in a broader, less pertinent category encompassing all industrial trucks and tractors. Furthermore, the data were collected five years earlier.
- An investigator who wishes to study individuals earning more than $100,000 per year finds the top category in a secondary study reported at $75,000 or more per year.
- A brewery that wishes to compare its per-barrel advertising expenditures with those of competitors finds that the units of measurement differ because some report point-of-purchase expenditures with advertising and others do not.
- Data from a previous warranty card study show where consumers prefer to purchase the product but provide no reasons why.

The most common reasons why secondary data do not adequately satisfy research needs are (1) outdated information, (2) variation in definition of terms, (3) different units of measurement, and (4) lack of information to verify the data’s accuracy. Furthermore, in our rapidly changing environment, information quickly becomes outdated. Because the purpose of most studies is to predict the future, secondary data must be timely to be useful.

Every primary researcher has the right to define the terms or concepts under investigation to satisfy the purpose of his or her primary investigation. This practice provides little solace, however,
to the investigator of the African-American market who finds secondary data reported as “percent nonwhite.” Variances in terms or variable classifications should be scrutinized to determine whether differences are important. The populations of interest must be described in comparable terms. Researchers frequently encounter secondary data that report on a population of interest that is similar but not directly comparable to their population of interest. For example, Arbitron reports its television audience estimates by geographical areas known as ADIs (Areas of Dominant Influence). An ADI is a geographic area consisting of all counties in which the home market commercial television stations receive a preponderance of total viewing hours. This unique population of interest is used exclusively to report television audiences. The geographic areas used in the census of population, such as Metropolitan Statistical Areas, are not comparable to ADIs.

Units of measurement may cause problems if they do not conform exactly to a researcher’s needs as well. For example, lumber shipments in millions of board feet are quite different from billions of ton miles of lumber shipped on freight cars. Head-of-household income is not the same unit of measure as total family income. Often the objective of the original primary study may dictate that the data be summarized, rounded, or reported. When that happens, even if the original units of measurement were comparable, aggregated or adjusted units of measurement are not suitable in the secondary study.

When secondary data are reported in a format that does not exactly meet the researcher’s needs, data conversion may be necessary. Data conversion (also called data transformation) is the process of changing the original form of data to a format more suitable for achieving a stated research objective. For example, sales for food products may be reported in pounds, cases, or dollars. An estimate of dollars per pound may be used to convert dollar volume data to pounds or another suitable format.

Another disadvantage of secondary data is that the user has no control over their accuracy. Although timely and pertinent secondary data may fit the researcher’s requirements, the data could be inaccurate. Research conducted by other persons may be biased to support the vested interest of the source. For example, media often publish data from surveys to identify the characteristics of their subscribers or viewers, but they will most likely exclude derogatory data from their reports. If the possibility of bias exists, the secondary data should not be used.

Investigators are naturally more prone to accept data from reliable sources such as the U.S. government. Nevertheless, the researcher must assess the reputation of the organization that gathers the data and critically assess the research design to determine whether the research was correctly implemented. Unfortunately, such evaluation may be impossible without full information that explains how the original research was conducted.
Researchers should verify the accuracy of the data whenever possible. **Cross-checks** of data from multiple sources, similar to what Nielsen Claritas does with its Business-Facts database, should be made to determine the similarity of independent projects. When the data are not consistent, researchers should attempt to identify reasons for the differences or to determine which data are most likely to be correct. If the accuracy of the data cannot be established, the researcher must determine whether using the data is worth the risk. Exhibit 8.1 illustrates a series of questions that should be asked to evaluate secondary data before they are used.

**Cross-checks**
The comparison of data from one source with data from another source to determine the similarity of independent projects.

**Exhibit 8.1 Evaluating Secondary Data**

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do the data help to answer questions set out in the problem definition?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>Do the data apply to the time period of interest?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>Do the data apply to the population of interest?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>Do other terms and variable classifications presented apply to the current project?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>Are the units of measurement comparable?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>Is it possible to go to the original source of the data?</td>
<td>Yes</td>
<td>Stop</td>
</tr>
<tr>
<td>Is the cost of data acquisition worth it?</td>
<td>Yes</td>
<td>Stop</td>
</tr>
<tr>
<td>Is there a possibility of bias?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>Can the accuracy of data collection be verified?</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>(inaccurate or unsure)</td>
<td>No</td>
<td>Stop</td>
</tr>
<tr>
<td>(accurate)</td>
<td>Yes</td>
<td>Stop</td>
</tr>
</tbody>
</table>

Typical Objectives for Secondary-Data Research Designs

It would be impossible to identify all the purposes of research using secondary data. However, some common business and marketing problems that can be addressed with secondary research designs are useful. Exhibit 8.2 shows three general categories of research objectives: fact-finding, model building, and database marketing.

<table>
<thead>
<tr>
<th>Broad Objective</th>
<th>Specific Research Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact-finding</td>
<td>Identifying consumption patterns</td>
</tr>
<tr>
<td></td>
<td>Tracking trends</td>
</tr>
<tr>
<td>Model building</td>
<td>Estimating market potential</td>
</tr>
<tr>
<td></td>
<td>Forecasting sales</td>
</tr>
<tr>
<td></td>
<td>Selecting trade areas and sites</td>
</tr>
<tr>
<td>Database marketing</td>
<td>Enhancing customer databases</td>
</tr>
<tr>
<td></td>
<td>Developing prospect lists</td>
</tr>
</tbody>
</table>

Fact-Finding

The simplest form of secondary-data research is fact-finding. A restaurant serving breakfast might be interested in knowing what new products are likely to entice consumers. Secondary data available from National Eating Trends, a service of the NPD Group, show that the most potential may be in menu items customers can eat on the go. According to data from the survey of eating trends, take-out breakfasts have doubled over the past few years, and they have continued to surpass dine-in breakfast sales for over a decade. These trends make smoothies and breakfast sandwiches sound like a good bet for a breakfast menu. Also, NPD found that 41 percent of breakfast sandwiches are consumed by people in their cars and 24 percent of people polled take them to work. These findings suggest that the sandwiches should be easy to handle. But what to put on the biscuit or bun? Another research firm, Market Facts, says almost half of consumers say they would pay extra for cheese. These simple facts would interest a researcher who was investigating the market for take-out breakfasts. Fact-finding can serve more complex purposes as well. In the digital age we live in, the use of music as a means to notify users of a call is commonplace. The Research Snapshot on the next page gives some of the amazing growth facts predicted in this industry.
Environmental Scanning

In many instances, the purpose of fact-finding is simply to study the environment to identify trends. Environmental scanning entails information gathering and fact-finding designed to detect indications of environmental changes in their initial stages of development. The Internet can be used for environmental scanning; however, there are other means, such as periodic review of contemporary publications...
and reports. For example, environmental scanning has shown many researchers that consumer demand in China is skyrocketing. In the case of beauty products such as cosmetics, Chinese authorities in the early 1990s stopped discouraging the use of makeup, and sales of these products took off—hitting $524 million in 2005—and were expected to grow by over one-third, reaching $705 million by 2009. Companies including Procter & Gamble, L’Oréal, and Shiseido have captured a sizable share of this market by realizing the potential and developing products to get into the Chinese market early.4

A number of online information services, such as Factiva and LexisNexis, routinely collect news stories about industries, product lines, and other topics of interest that have been specified by the researcher. In addition, push technology is an Internet information technology that automatically delivers content to the researcher’s or manager’s desktop. Push technology uses “electronic smart agents,” custom software that filters, sorts, prioritizes, and stores information for later viewing.5 This service frees the researcher from doing the searching. The true value of push technology is that the researcher who is scanning the environment can specify the kinds of news and information he or she wants, have it delivered to his or her computer quickly, and view it at leisure.

Model Building

The second general objective for secondary research, model building, is more complicated than simple fact-finding. Model building involves specifying relationships between two or more variables, perhaps extending to the development of descriptive or predictive equations, a technique that is used by the Nielsen Claritas Company routinely to add value to their secondary data. Models need not include complicated mathematics, though. In fact, decision makers often prefer simple models that everyone can readily understand over complex models that are difficult to comprehend. For example, market share is company sales divided by industry sales. Although some may not think of this simple calculation as a model, it represents a mathematical model of a basic relationship.

We will illustrate model building by discussing three common objectives that can be satisfied with secondary research: estimating market potential, forecasting sales, and selecting potential facility or expansion sites.

ESTIMATING MARKET POTENTIAL FOR GEOGRAPHIC AREAS

Business researchers often estimate their company’s market potential using secondary data. In many cases exact figures may be published by a trade association or another source. However, when the desired information is unavailable, the researcher may estimate market potential by transforming secondary data from two or more sources. For example, managers may find secondary data about market potential for a country or other large geographic area, but this information may not be broken down into smaller geographical areas, such as by metropolitan area, or in terms unique to the company, such as sales territory. In this type of situation, researchers often need to make projections for the geographic area of interest.
An extended example will help explain how secondary data can be used to calculate market potential. Suppose a brewing company is looking for opportunities to expand sales by exporting or investing in other countries. Managers decide to begin by estimating market potential for the Czech Republic, Germany, Japan, and Spain. Secondary research uncovered data for per capita beer consumption and population projections for the year 2010. The data for the four countries appear in Exhibit 8.4.

To calculate market potential for the Czech Republic in 2010, multiply that country’s population in the year 2010 by its per capita beer consumption:

\[
10,175,000 \text{ people} \times 157 \text{ liters/person} = 1,597,475,000 \text{ liters}
\]

In the Czech Republic, the market potential for beer is 1,597,475,000 liters. To get a sense of the expected sales volume, the researcher would have to multiply this amount by the price per liter at which beer typically sells in the Czech Republic. As Exhibit 8.4 reveals, Japan’s population is much higher, so its market potential is greater, even though the average Czech drinks much more beer.

Of course, the calculated market potential for each country in Exhibit 8.4 is a rough estimate. One obvious problem is that not everyone in a country will be of beer-drinking age. If the researcher can get statistics for each country’s projected adult population, the estimate will be closer. Also, you might want to consider whether each country is experiencing growth or decline in the demand for beer to estimate whether consumption habits are likely to be different in 2010. For example, beer consumption is barely growing in Europe and Japan, but it is expanding in Latin America (at about 4 percent a year) and even faster in China (by at least 6 percent a year). Perhaps this information will cause you to investigate market potential in additional countries where more growth is expected.

■ FORECASTING SALES

For any project, such as forecasting sales, you need information about the future. You will need to know what company sales will be next year and in future time periods. Sales forecasting is the process of predicting sales totals over a specific time period.

Accurate sales forecasts, especially for products in mature, stable markets, frequently come from secondary-data research that identifies trends and extrapolates past performance into the future. Researchers often use internal company sales records to project sales. A rudimentary model would multiply past sales volume by an expected growth rate. A researcher might investigate a secondary source and find that industry sales are expected to grow by 10 percent; multiplying company sales volume by 10 percent would give a basic sales forecast.

Exhibit 8.5 on the next page illustrates trend projection using a moving average projection of growth rates. Average ticket prices for a major-league baseball game are secondary data from
Team Marketing Report for each year of interest (http://www.teammarketing.com/fancost/mlb/). The moving average is the sum of growth rates for the past three years divided by 3 (number of years). The resulting number is a forecast of the percentage increase in ticket price for the coming year. Using the three-year average growth rate of 6.4 percent for the 2008, 2007, and 2006 sales periods, we can forecast the average ticket price for 2009 as follows:

\[ \text{Average Ticket Price} = \text{Average Ticket Price} + (\text{Average Ticket Price} \times .064) \]

\[ = 25.43 + (25.43 \times .064) = 27.05 \]

Using the same information, the projected price of a beer at a ballgame in 2009 is $6.43. This lets the fan know how much to take out to the old ballgame.

Moving average forecasting is best suited to a static competitive environment. More dynamic situations make other sales forecasting techniques more appropriate.

Statistical trend analysis using secondary data can be much more advanced than this simple example. Many statistical techniques build forecasting models using secondary data. This chapter emphasizes secondary-data research rather than statistical analysis, which is covered in later chapters.

### ANALYSIS OF TRADE AREAS AND SITES

Managers routinely examine trade areas and use **site analysis techniques** to select the best locations for retail or wholesale operations. Secondary-data research helps managers make these site selection decisions. Some organizations, especially franchisers, have developed special computer software based on analytical models to select sites for retail outlets. The researcher must obtain the appropriate secondary data for analysis with the computer software.

The **index of retail saturation** offers one way to investigate retail sites and to describe the relationship between retail demand and supply.\(^7\) It is easy to calculate once the appropriate secondary data are obtained:

\[ \text{Index of retail saturation} = \frac{\text{Local market potential (demand)}}{\text{Local market retailing space}} \]

For example, Exhibit 8.6 shows the relevant secondary data for shoe store sales in a five-mile radius surrounding a Florida shopping center. These types of data can be purchased from vendors...
of market information such as Urban Decision Systems. First, to estimate local market potential (demand), we multiply population by annual per capita shoe sales. This estimate, line 3 in Exhibit 8.6, goes in the numerator to calculate the index of retail saturation:

\[
\text{Index of retail saturation} = \frac{\$14,249,000}{94,000} = 152
\]

The retailer can compare this index figure with those of other areas to determine which sites have the greatest market potential with the least amount of retail competition. An index value above 200 is considered to indicate exceptional opportunities.

**Data Mining**

Large corporations’ decision support systems often contain millions or even hundreds of millions of records of data. These complex data volumes are too large to be understood by managers. Consider, for example, Capital One, a consumer lending company with nearly 50 million customer accounts, including credit cards and auto loans. Suppose the company collects data on customer purchases, and each customer makes five transactions in a month, or 60 per year. With 50 million customers and decades of data (the company was founded in 1988), it’s easy to see how record counts quickly grow beyond the comfort zone for most humans.

Two points about data volume are important to keep in mind. First, relevant data are often in independent and unrelated files. Second, the number of distinct pieces of information each data record contains is often large. When the number of distinct pieces of information contained in each data record and data volume grows too large, end users don’t have the capacity to make sense of it all. Data mining helps clarify the underlying meaning of the data.

The term **data mining** refers to the use of powerful computers to dig through volumes of data to discover patterns about an organization’s customers and products. As seen in the Research Snapshot on the next page, this can even apply to Internet content from blogs. It is a broad term that applies to many different forms of analysis. For example, **neural networks** are a form of artificial intelligence in which a computer is programmed to mimic the way that human brains process information. One computer expert put it this way:

> A neural network learns pretty much the way a human being does. Suppose you say “big” and show a child an elephant, and then you say “small” and show her a poodle. You repeat this process with a house and a giraffe as examples of “big” and then a grain of sand and an ant as examples of “small.” Pretty soon she will figure it out and tell you that a truck is “big” and a needle is “small.” Neural networks can similarly generalize by looking at examples.

**Market-basket analysis** is a form of data mining that analyzes anonymous point-of-sale transaction databases to identify coinciding purchases or relationships between products purchased and other retail shopping information. Consider this example about patterns in customer purchases: Osco Drugs mined its databases provided by checkout scanners and found that when men go to its drugstores to buy diapers in the evening between 6:00 p.m. and 8:00 p.m., they sometimes walk out with a six-pack of beer as well. Knowing this behavioral pattern, supermarket managers may consider laying out their stores so that these items are closer together.
A data-mining application of interest to some researchers is known as customer discovery, which involves mining data to look for patterns identifying who is likely to be a valuable customer. For example, a larger provider of business services wanted to sell a new product to its existing customers, but it knew that only some of them would be interested. The company had to adapt each product offering to each customer's individual needs, so it wanted to save money by identifying the best prospects. It contracted with a research provider called DataMind to mine its data on sales, responses to marketing, and customer service to look for the customers most likely to be interested in the new product. DataMind assigned each of the company's customers an index number indicating their expected interest level, and the selling effort was much more efficient as a result.12

When a company knows the identity of the customer who makes repeated purchases from the same organization, an analysis can be made of sequences of purchases. The use of data mining to detect sequence patterns is a popular application among direct marketers, such as catalog retailers. A catalog merchant has information for each customer, revealing the sets of products that the customer buys in every purchase order. A sequence detection function can then be used to discover the set of purchases that frequently precedes the purchase of, say, a microwave oven. As another example, a sequence of insurance claims could lead to the identification of frequently occurring medical procedures performed on patients, which in turn could be used to detect cases of medical fraud.

Data mining requires sophisticated computer resources, and it is expensive. That's why companies like DataMind, IBM, Oracle, Information Builders, and Acxiom Corporation offer data-mining services. Customers send the databases they want analyzed and let the data-mining company do the "number crunching."
a database of previous customers, including what purchases they made during the Christmas holidays. Each year the company sends last year’s gift list to customers to help them send the same gifts to their friends and relatives.

Because database marketing requires vast amounts of CRM data compiled from numerous sources, secondary data are often acquired for the exclusive purpose of developing or enhancing databases. The transaction record, which often lists the item purchased, its value, customer name, address, and zip code, is the building block for many databases. This may be supplemented with data customers provide directly, such as data on a warranty card, and by secondary data purchased from third parties. For example, credit services may sell databases about applications for loans, credit card payment history, and other financial data. Several companies, such as Donnelley Marketing (with its BusinessContentFile and ConsumerContentFile services) and Claritas (with PRIZM), collect primary data and then sell demographic data that can be related to small geographic areas, such as those with a certain zip code. (Remember that when the vendor collects the data, they are primary data, but when the database marketer incorporates the data into his or her database, they are secondary data.)

Now that some of the purposes of secondary-data analysis have been addressed, we turn to a discussion of the sources of secondary data.

**Sources of Secondary Data**

Secondary data can be classified as either internal to the organization or external. Modern information technology makes this distinction seem somewhat simplistic. Some accounting documents are indisputably internal records of the organization. Researchers in another organization cannot have access to them. Clearly, a book published by the federal government and located at a public library is external to the company. However, in today’s world of electronic data interchange, the data that appear in a book published by the federal government may also be purchased from an online information vendor for instantaneous access and subsequently stored in a company’s decision support system.

Internal data should be defined as data that originated in the organization, or data created, recorded, or generated by the organization. **Internal and proprietary data** is perhaps a more descriptive term.

**Sources of Internal and Proprietary Data**

Most organizations routinely gather, record, and store internal data to help them solve future problems. An organization’s accounting system can usually provide a wealth of information. Routine documents such as sales invoices allow external financial reporting, which in turn can be a source of data for further analysis. If the data are properly coded into a modular database in the accounting system, the researcher may be able to conduct more detailed analysis using the decision support system. Sales information can be broken down by account or by product and region; information related to orders received, back orders, and unfilled orders can be identified; sales can be forecast on the basis of past data. Other useful sources of internal data include salespeople’s call reports, customer complaints, service records, warranty card returns, and other records.

Researchers frequently aggregate or disaggregate internal data. For example, a computer service firm used internal secondary data to analyze sales over the previous three years, categorizing business by industry, product, purchase level, and so on. The company discovered that 60 percent of its customers represented only 2 percent of its business and that nearly all of these customers came through telephone directory advertising. This simple investigation of internal records showed that, in effect, the firm was paying to attract customers it did not want.

Internet technology is making it easier to research internal and proprietary data. Often companies set up intranets so that employees can use Web tools to store and share data within the organization. And just as Google’s search software lets people search the entire World Wide Web, Google is offering the enterprise search, which is essentially the same technology in a version that searches a corporate intranet. The enterprise search considers not only how often a particular
document has been viewed but also the history of the user’s past search patterns, such as how
often that user has looked at particular documents and for how long. In addition, other companies
have purchased specialized software, such as Autonomy, which searches internal sources plus such
external sources as news government Web sites.¹³

**External Data: The Distribution System**

External data are generated or recorded by an entity other than the researcher’s organization. The
government, newspapers and journals, trade associations, and other organizations create or pro-
duce information. Traditionally, this information has been in published form, perhaps available
from a public library, trade association, or government agency. Today, however, computerized
data archives and electronic data interchange make external data as accessible as internal data.
Exhibit 8.7 illustrates some traditional and some modern ways of distributing information.

**Information as a Product and Its Distribution Channels**

Because secondary data have value, they can be bought and sold like other products. And just
as bottles of perfume or plumbers’ wrenches may be distributed in many ways, secondary data
also flow through various channels of distribution. Many users, such as the Fortune 500 corpora-
tions, purchase documents and computerized census data directly from the government. How-
ever, many small companies get census data from a library or another intermediary or vendor of
secondary information.

### LIBRARIES

Traditionally, libraries’ vast storehouses of information have served as a bridge between users and
producers of secondary data. The library staff deals directly with the creators of information, such
as the federal government, and intermediate distributors of information, such as abstracting and
indexing services. The user need only locate the appropriate secondary data on the library shelves.
Libraries provide collections of books, journals, newspapers, and so on for reading and reference.
They also stock many bibliographies, abstracts, guides, directories, and indexes, as well as offer
access to basic databases.

The word *library* typically connotes a public or university facility. However, many major cor-
porations and government agencies also have libraries. A corporate librarian’s advice on sources of
industry information or the United Nations librarian’s help in finding statistics about international
markets can be invaluable.

### THE INTERNET

Today, of course, much secondary data is conveniently available over the Internet. Its creation has
added an international dimension to the acquisition of secondary data. For example, Library Spot,
at [http://www.libraryspot.com](http://www.libraryspot.com), provides links to online libraries, including law libraries, medical librar-
ies, and music libraries. Its reference desk features links to calendars, dictionaries, encyclopedias,
maps, and other sources typically found at a traditional library’s reference desk.

Exhibit 8.8 on page 174 lists some of the more popular Internet addresses where secondary
data may be found.

### VENDORS

The information age offers many channels besides libraries through which to access data. Many
external producers make secondary data available directly from the organizations that produce the
data or through intermediaries, which are often called *vendors*. Vendors such as Factiva now allow
managers to access thousands of external databases via desktop computers and telecommunications
systems. Hoovers ([http://www.hoovers.com](http://www.hoovers.com)) specializes in providing information about thousands of
companies’ financial situations and operations.
PRODUCERS

Classifying external secondary data by the nature of the producer of information yields five basic sources: publishers of books and periodicals, government sources, media sources, trade association sources, and commercial sources. The following section discusses each type of secondary data source.
Books and Periodicals

Some researchers consider books and periodicals found in a library to be the quintessential secondary data source. A researcher who finds books on a topic of interest obviously is off to a good start.

Professional journals, such as the Journal of Marketing, Journal of Management, Journal of the Academy of Marketing Science, The Journal of Business Research, Journal of Advertising Research, American Demographics, and The Public Opinion Quarterly, as well as commercial business periodicals such as the Wall Street Journal, Fortune, and BusinessWeek, contain much useful material. Sales and Marketing Management’s Survey of Buying Power is a particularly useful source of information about markets. To locate data in periodicals, indexing services such as the ABI/INFORM and Business Periodicals

---

TO THE POINT

The man who does not read good books has no advantage over the man who cannot read them.

—Mark Twain
Index and the Wall Street Journal Index are very useful. Guides to data sources also are helpful. For example, American Statistical Index and Business Information Sources is a very valuable source. Most university libraries provide access to at least some of these databases.

**Government Sources**

Government agencies produce data prolifically. Most of the data published by the federal government can be counted on for accuracy and quality of investigation. Most students are familiar with the U.S. Census of Population, which provides a wealth of data.

The Census of Population is only one of many resources that the government provides. Banks and savings and loan companies rely heavily on the Federal Reserve Bulletin and the Economic Report of the President for data relating to research on financial and economic conditions. Builders and contractors use the information in the Current Housing Reports and American Housing Survey for their research. The Statistical Abstract of the United States is an extremely valuable source of information about the social, political, and economic organizations of the United States. It abstracts data available in hundreds of other government publications and serves as a convenient reference to more specific statistical data.

The federal government is a leader in making secondary data available on the Internet. Visit FedWorld (http://www.fedworld.gov) for a central access point and links to many of these important documents. STAT-USA/Internet is another authoritative and comprehensive source of U.S. government information that focuses on economic, financial, and trade data. It contains the following types of information:

- More than 18,000 market research reports on individual countries and markets compiled by foreign experts at U.S. embassies
- Economic data series, current and historical, such as gross domestic product, balance of payment, and merchandise trade
- Standard reference works, such as the Economic Report of the President, the Budget of the United States Federal Government, and the World Factbook
- Worldwide listings of businesses interested in buying U.S. products

The STAT-USA/Internet Web address is http://www.stat-usa.gov. However, only subscribers who pay a fee have access to this service.

State, county, and local government agencies can also be useful sources of information. Many state governments publish state economic models and forecasts, and many cities have metropolitan planning agencies that provide data about the population, economy, transportation system, and so on. These are similar to federal government data but are more current and are structured to suit local needs.

Many cities and states publish information on the Internet. Many search engines have directory entries that allow easy navigation to a particular state’s Web site. A researcher using Yahoo!, for example, needs only to click Regional Information to find numerous paths to information about states.

**Media Sources**

Information on a broad range of subjects is available from broadcast and print media. CNN Financial News and BusinessWeek are valuable sources for information on the economy and many industries. Media frequently commission research studies about various aspects of Americans’ lives, such as financial affairs, and make reports of survey findings available to potential advertisers free of charge. Data about the readers of magazines and the audiences for broadcast media typically are profiled in media kits and advertisements.

Information about special-interest topics may also be available. Hispanic Business reports that the number of Hispanic-owned companies in the United States is expected to grow at a rate of 35 percent between 2004 and 2010, reaching 3.2 million firms, with revenue growth for the period of 70 percent. According to the magazine, most of these firms are located in 20 states, with over half in California and Florida. For researchers willing to pay a modest $85, Hispanic Business offers a more detailed report about Hispanic-owned businesses.14

Data such as these are plentiful because the media like to show that their vehicles are viewed or heard by advertisers’ target markets. These types of data should be evaluated carefully, however,
because often they cover only limited aspects of a topic. Nevertheless, they can be quite valuable for research, and they are generally available free of charge.

**Trade Association Sources**

Trade associations, such as the Food Marketing Institute or the American Petroleum Institute, serve the information needs of a particular industry. The trade association collects data on a number of topics of specific interest to firms, especially data on market size and market trends. Association members have a source of information that is particularly germane to their industry questions. For example, the Newspaper Advertising Bureau (NAB) has catalogued and listed in its computer the specialized sections that are currently popular in newspapers. The NAB has surveyed all daily, Sunday, and weekend newspapers in the United States and Canada on their editorial content and has stored this information, along with data on rates, circulation, and mechanical requirements, in its computer for advertisers’ use. As seen in the Research Snapshot above, trade associations are valuable sources of interesting data.

**Commercial Sources**

Numerous firms specialize in selling and/or publishing information. For example, the Polk Company publishes information on the automotive field, such as average car values and new-car purchase rates by zip code. Many of these organizations offer information in published formats and as CD-ROM or Internet databases. The following discussion of several of these firms provides a sampling of the diverse data that are available.

**Market-Share Data.** A number of syndicated services supply either wholesale or retail sales volume data based on product movement. Information Resources, Inc., collects market-share data using Universal Product Codes (UPC) and optical scanning at retail store checkouts. INFOSCAN is a syndicated store tracking service that collects scanner data weekly from more than 32,000 supermarket, drug, and mass merchandiser outlets across the United States. Sales in France, Germany, Greece, Italy, the Netherlands, Spain, and the United Kingdom also are tracked by INFOSCAN.

Although it is best known for its television rating operations, ACNielsen also has a scanner-based marketing and sales information service called ScanTrack. This service gathers sales and marketing data from a sample of more than 4,800 stores representing more than 800 retailers in 50 major U.S. markets. As part of Nielsen’s Retail Measurement Service, auditors visit the stores...
at regular intervals to track promotions to customers, retail inventories, displays, brand distribution, out-of-stock conditions, and other retail marketing activity. Scanner data allow researchers to monitor sales data before, during, and after changes in advertising frequency, price changes, distribution of free samples, and similar marketing tactics.

Wal-Mart operates its own in-store scanner system called RetailLink. Key suppliers can have online access to relevant data free of charge. The *Market Share Reporter* is produced each year, made available for sale, and provides market share data for most industries.

Many primary data investigations use scanner data to measure the results of experimental manipulations such as altering advertising copy. For example, scanning systems combined with consumer panels are used to create electronic test-markets. Systems based on UPCs (bar codes) and similar technology have been implemented in factories, warehouses, and transportation companies to research inventory levels, shipments, and the like.

**Demographic and Census Updates.** A number of firms, such as CACI Marketing Systems and Urban Information Systems, offer computerized U.S. census files and updates of these data broken down by small geographic areas, such as zip codes. Many of these research suppliers provide in-depth information on minority customers and other market segments.

**Consumer Attitude and Public Opinion Research.** Many research firms offer specialized syndicated services that report findings from attitude research and opinion polls. For example, Yankelovich provides custom research, tailored for specific projects, and several syndicated services. Yankelovich’s public opinion research studies, such as the voter and public attitude surveys that appear in *Time* and other news magazines, are a source of secondary data. One of the firm’s services is the *Yankelovich Monitor*, a syndicated annual census of changing social values and an analysis of how they can affect consumer marketing. The *Monitor* charts the growth and spread of new social values, characterizes the types of customers who support the new values and those who continue to support traditional values, and outlines the ways in which people’s values affect purchasing behavior.

Harris/Interactive is another public opinion research firm that provides syndicated and custom research for business. One of its services is its ABC News/Harris survey. This survey, released three times per week, monitors the pulse of the American public on topics such as inflation, unemployment, energy, attitudes toward the president, elections, and so on.

**Consumption and Purchase Behavior Data.** NPD’s *National Eating Trends* (NET) is the most detailed database available on consumption patterns and trends for more than 4,000 food and beverage products. This is a syndicated source of data about the types of meals people eat and when and how they eat them. The data, called *diary panel data*, are based on records of meals and diaries kept by a group of households that have agreed to record their consumption behavior over an extended period of time.

National Family Opinion (NFO), Marketing Research Corporation of America (MRCA), and many other syndicated sources sell diary panel data about consumption and purchase behavior. Since the advent of scanner data, diary panels are more commonly used to record purchases of apparel, hardware, home furnishings, jewelry, and other durable goods, rather than purchases of non-durable consumer packaged goods. More recently, services have been tracking consumer behavior online, collecting data about sites visited and purchases made over the Internet.

**Advertising Research.** Advertisers can purchase readership and audience data from a number of firms. W. R. Simmons and Associates measures magazine audiences; Arbitron measures radio audiences; ACNielsen Media Measurement estimates television audience ratings. By specializing in collecting and selling audience information on a continuing basis, these commercial sources provide a valuable service to their subscribers.

Assistance in measuring advertising effectiveness is another syndicated service. For example, Roper Starch Worldwide measures the impact of advertising in magazines. Readership information can be obtained for competitors’ ads or the client’s own ads. Respondents are classified as noted readers, associated readers, or read-most readers.

Burke Marketing Research provides a service that measures the extent to which respondents recall television commercials aired the night before. It provides product category norms, or average DAR (Day-After Recall) scores, and DAR scores for other products.
An individual advertiser would be unable to monitor every minute of every television program before deciding on the appropriate ones in which to place advertising. However, numerous clients, agencies, television networks, and advertisers can purchase the Nielsen television ratings service.

**Single-Source Data-Integrated Information**

ACNielsen Company offers data from both its television meters and scanner operations. The integration of these two types of data helps marketers investigate the impact of television advertising on retail sales. In other ways as well, users of data find that merging two or more diverse types of data into a single database offers many advantages.

PRIZM by Nielsen Claritas, CACI, ClusterPlus by SMI, Mediamark Research Inc., and many other syndicated databases report product purchase behavior, media usage, demographic characteristics, lifestyle variables, and business activity by geographic area such as zip code. Although such data are often called *geodemographic*, they cover such a broad range of phenomena that no one name is a good description. These data use small geographic areas as the unit of analysis.

The data and information industry uses the term *single-source data* for diverse types of data offered by a single company. Exhibit 8.9 identifies three major marketers of single-source data.

---

**EXHIBIT 8.9**

Examples of Single-Source Databases

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Description</th>
</tr>
</thead>
</table>
| CACI Marketing Systems  
http://www.caci.com | Provides industry-specific marketing services, such as customer profiling and segmentation, custom target analysis, demographic data reports and maps, and site evaluation and selection. CACI offers demographics and data on businesses, lifestyles, consumer spending, purchase potential, shopping centers, traffic volumes, and other statistics. |
| PRIZM by Claritas Corporation  
http://www.claritas.com | PRIZM, which stands for Potential Rating Index for Zip Markets, is based on the “birds-of-a-feather” assumption that people live near others who are like themselves. PRIZM combines census data, consumer surveys about shopping and lifestyle, and purchase data to identify market segments. Colorful names such as Young Suburbia, Shot Guns, and Pickups describe 40 segments that can be identified by zip code. Claritas also has a lifestyle census in the United Kingdom (http://www.claritas.co.uk). |
| MRI Cable Report—Mediamark Research Inc.  
http://www.mediamark.com | Integrates information on cable television viewing with demographic and product usage information. |

---

**Sources for Global Research**

As business has become more global, so has the secondary data industry. The Japan Management Association Research Institute, Japan’s largest provider of secondary research data to government and industry, maintains an office in San Diego. The Institute’s goal is to help U.S. firms access its enormous store of data about Japan to develop and plan their business there. The office in San Diego provides translators and acts as an intermediary between Japanese researchers and U.S. clients.

Secondary data compiled outside the United States have the same limitations as domestic secondary data. However, international researchers should watch for certain pitfalls that frequently are associated with foreign data and cross-cultural research. First, data may simply be unavailable in certain countries. Second, the accuracy of some data may be called into question. This is especially likely with official statistics that may be adjusted for the political purposes of foreign...
governments. Finally, although economic terminology may be standardized, various countries use different definitions and accounting and recording practices for many economic concepts. For example, different countries may measure disposable personal income in radically different ways. International researchers should take extra care to investigate the comparability of data among countries. The Research Snapshot above provides some of the many Web site locations for data from around the world.

The U.S. government and other organizations compile databases that may aid international secondary data needs. For example, The European Union in the U.S. (http://www.eurunion.org) reports on historical and current activity in the European Union providing a comprehensive reference guide to information about laws and regulations. The European Union in the U.S. profiles in detail each European Union member state, investment opportunities, sources of grants and other funding, and other information about business resources.

The U.S. government offers a wealth of data about foreign countries. The CIA’s World Factbook and the National Trade Data Bank are especially useful. Both can be accessed using the Internet. The National Trade Data Bank (NTDB), the U.S. government’s most comprehensive source of world trade data, illustrates what is available.

The National Trade Data Bank was established by the Omnibus Trade and Competitiveness Act of 1988. Its purpose was to provide “reasonable public access, including electronic access” to an export promotion data system that was centralized, inexpensive, and easy to use.

The U.S. Department of Commerce has the responsibility for operating and maintaining the NTDB and works with federal agencies that collect and distribute trade information to keep the NTDB up-to-date. The NTDB has been published monthly on CD-ROM since 1990. Over one thousand public and university libraries offer access to the NTDB through the Federal Depository Library system.

The National Trade Data Bank consists of 133 separate trade- and business-related programs (databases). By using it, small- and medium-sized companies get immediate access to information that until now only Fortune 500 companies could afford.

Topics in the NTDB include export opportunities by industry, country, and product; foreign companies or importers looking for specific products; how-to market guides; demographic, political, and socioeconomic conditions in hundreds of countries; and much more. NTDB offers one-stop shopping for trade information from more than 20 federal sources. You do not need to know which federal agency produces the information: All you need to do is consult NTDB.

Some of the specific information that can be obtained from the NTDB is listed in Exhibit 8.10 on the next page.
EXHIBIT 8.10
Examples of Information Contained in the NTDB

- Agricultural commodity production and trade
- Basic export information
- Calendars of trade fairs and exhibitions
- Capital markets and export financing
- Country reports on economic and social policies and trade practices
- Energy production, supply, and inventories
- Exchange rates
- Export licensing information
- Guides to doing business in foreign countries
- International trade terms directory
- How-to guides
- International trade regulations/agreements
- Labor, employment, and productivity
- Maritime and shipping information
- Market research reports
- Overseas contacts
- Overseas and domestic industry information
- Price indexes
- Small business information
- State exports
- State trade contacts
- Trade opportunities
- U.S. export regulations
- U.S. import and export statistics by country and commodity
- U.S. international transactions
- World Factbook
- World minerals production

Always consider the possibility that secondary data may exist which can address the research question at hand.
Only rely on secondary data that are reliable and valid. Generally, the reliability and validity are established by details the data source provides about how the data were collected and processed.
Only rely on secondary data for which the units of measure are clear.
Secondary data are particularly useful for trend analysis, environmental scanning, and estimating market potential for geographic areas.

Government sites such as the Census Bureau (www.census.gov), the CIA Factbook (www.cia.gov), and STAT-USA (www.stat-usa.gov) are great sources for geodemographic data about locations and peoples around the world.
Summary

1. Discuss the advantages and disadvantages of secondary data. Secondary data are data that have been gathered and recorded previously by someone else for purposes other than those of the current researcher. The chief advantage of secondary data is that they are almost always less expensive to obtain than primary data. Generally they can be obtained rapidly and may provide information not otherwise available to the researcher. The disadvantage of secondary data is that they were not intended specifically to meet the researcher’s needs. The researcher must examine secondary data for accuracy, bias, and soundness. One way to do this is to cross-check various available sources.

2. Define types of secondary data analysis conducted by business research managers. Secondary research designs address many common business research problems. There are three general categories of secondary research objectives: fact-finding, model building, and database marketing. A typical fact-finding study might seek to uncover all available information about consumption patterns for a particular product category or to identify business trends that affect an industry. Model building is more complicated; it involves specifying relationships between two or more variables. The practice of database marketing, which involves maintaining customer databases with customers’ names, addresses, phone numbers, past purchases, responses to past promotional offers, and other relevant data such as demographic and financial data, is increasingly being supported by business research efforts.

3. Identify various internal and proprietary sources of secondary data. Managers often get data from internal proprietary sources such as accounting records. Data mining is the use of powerful computers to dig through volumes of data to discover patterns about an organization’s customers and products. It is a broad term that applies to many different forms of analysis.

4. Give examples of various external sources of secondary data. External data are generated or recorded by another entity. The government, newspaper and journal publishers, trade associations, and other organizations create or produce information. Traditionally this information has been distributed in published form, either directly from producer to researcher, or indirectly through intermediaries such as public libraries. Modern computerized data archives, electronic data interchange, and the Internet have changed the distribution of external data, making them almost as accessible as internal data. Push technology is a term referring to an Internet information technology that automatically delivers content to the researcher’s or manager’s desktop. This service helps in environmental scanning.

5. Describe the impact of single-source data and globalization on secondary data research. The marketing of multiple types of related data by single-source suppliers has radically changed the nature of secondary-data research. Businesses can measure promotional efforts and related buyer behavior by detailed customer characteristics. As business has become more global, so has the secondary-data industry. International researchers should watch for pitfalls that can be associated with foreign data and cross-cultural research, such as problems with the availability and reliability of data.

Key Terms and Concepts

cross-checks, 163
customer discovery, 170
data conversion, 162
data mining, 169
database marketing, 170
external data, 172
index of retail saturation, 168
internal and proprietary data, 171
market tracking, 165
market-basket analysis, 169
model building, 166
neural networks, 169
secondary data, 161
single-source data, 178
site analysis techniques, 168
Questions for Review and Critical Thinking

1. Secondary data have been called the first line of attack for business researchers. Discuss this description.
2. Suppose you wish to learn about the size of the soft-drink market, particularly root beer sales, growth patterns, and market shares. Indicate probable sources for these secondary data.
3. What is push technology?
4. Identify some typical research objectives for secondary-data studies.
5. How might a researcher doing a job for a company such as Pulte Homes (http://www.pultehomes.com) or David Weekley Homes (http://www.davidweekley.com) use secondary data and data mining?
6. What would be a source for the following data?
   a. Population, average income, and employment rates for Oregon
   b. Maps of U.S. counties and cities
   c. Trends in automobile ownership
   d. Divorce trends in the United States
   e. Median weekly earnings of full-time, salaried workers for the previous five years
   f. Annual sales of the top ten fast-food companies
   g. Top ten Web sites ranked by number of unique visitors
   h. Attendance at professional sports events
7. Suppose you are a business research consultant and a client comes to your office and says, “I must have the latest information on the supply of and demand for Maine potatoes within the next 24 hours.” What would you do?
8. Find the following data in the Survey of Current Business:
   a. U.S. gross domestic product for the first quarter of 2006
   b. Exports of goods and services for the fourth quarter of 2006
   c. Imports of goods and services for the fourth quarter of 2006
9. ETHICS A newspaper reporter finds data in a study that surveyed children that reports a high percentage of children can match cartoon characters with the products they represent. For instance, they can match cereal with Captain Crunch and Ronald McDonald with a Big Mac. The reporter used this to write a story about the need to place limits on the use of cartoon characters. However, the study also provided data suggesting that matching the cartoon character and the product did not lead to significantly higher consumption. Would this be a proper use of secondary data?

Research Activities

1. Use secondary data to learn the size of the U.S. golf market and to profile the typical golfer.
2. *NET Where could a researcher working for the U.S. Marine Corps (http://www.marines.com) find information that would identify the most productive areas of the United States in which to recruit? What would you recommend?
3. *NET PopClocks estimate the U.S. and world populations. Go to the Census Bureau home page (http://www.census.gov), navigate to the population section, and find today’s estimate of the U.S. and world populations.
4. *NET Try to find the U.S. market share for the following companies within 30 minutes:
   a. Home Depot
   b. Burger King
   c. Marlboro
   d. Was this a difficult task? If so, why do you think it is this difficult?
5. *NET Use the Internet to learn what you can about Indonesia.
   b. What additional kinds of information are available from the following sources?
   • Go to http://freetheworld.com/member.html and view info for Indonesia.
   • Go to Google, Yahoo! Search, or another search engine, and use “Indonesia” as a search word.
6. *NET Go to Statistics Norway at http://www.ssb.no. What data, if any, can you obtain in English? What languages can be used to search this Web site? What databases might be of interest to the business researcher?
7. *NET Go to Statistics Canada at http://www.statcan.gc.ca. What languages can be used to search this Web site? What databases might be of interest to the business researcher?
8. *NET Suppose you were working for a company that wanted to start a business selling handmade acoustic guitars that are reproductions of classic vintage guitars. Pricing is a big part of the decision. Secondary information is available via the Internet. Use eBay (http://ebay.com) to identify four key brands of acoustic guitars by studying the vintage acoustic guitars listed for sale. Since the company wishes to charge premium prices, they will model after the most expensive brand. What brand seems to be associated with the highest prices?
In fall 2005, Hurricanes Katrina and Rita churning in the Gulf of Mexico damaged oil rigs and refineries, contributing to a spike in oil prices. Many observers expressed confidence that those events were the long-awaited trigger that would kill off demand for SUVs and other gas-guzzling vehicles.¹⁷ They were only partly right.

In the months leading up to the hurricanes, sales of SUVs had already been falling, according to data from Automotive News. Automakers had been shifting ad dollars away from these products. CNW Market Research said that in August 2005, consumers had for the first time placed fuel economy ahead of performance when ranking factors for choosing a new vehicle. When gas prices approached three dollars a gallon in September 2005, marketers felt sure that fuel economy would remain a top concern. Advertisers began creating more ads featuring vehicles’ gas mileage.

But by the end of the year, attitudes were shifting again. The National Automobile Dealers Association surveyed consumers visiting its Web site for information about car purchases, and it learned they ranked price as most important, followed by make and model, then performance. Fuel economy ranked last, with 3 percent considering it most important and 11 percent considering it least important. What’s a carmaker to do? General Motors gathers data from the shoppers who visit Web sites such as www.kbb.com to look up information, and it is analyzing the data to identify the price of fuel at which car buyers adjust their priorities.

Questions
1. From the standpoint of an automobile company, what sources of information in this article offer secondary data?
2. Suggest two or three other sources of data that might be of interest to auto companies interested in forecasting demand.
3. Online or at your library, look for information about recent trends in SUV purchases. Report what you learned, and forecast whether SUV sales are likely to recover or continue their decline. What role do gas prices play in your forecast?