2.6 Secondary Sources

While all data are, at some point, primary, the vast majority of information used by researchers is secondary data. The only real distinction between the two is that with secondary sources, you do not have control over the data-collection techniques.

In this section, you will learn basic research skills to ensure that you end up with quality data. You will also look at some issues related to using the Internet as a source of data.

The time and money required to conduct primary research makes it an impractical data-collection method, particularly in the context of a high school mathematics course. Fortunately, there are reputable secondary sources that can provide the type of reliable data you may require for your course project. One of these sources is Statistics Canada (Statscan). The federal government mandates and funds Statscan to conduct a census every five years, and the vast quantities of raw data that Statscan collects can be accessed by the public. However, as a researcher, it is up to you to make sense of the data you acquire.

RELIABILITY OF SOURCES

When collecting primary data, you must ensure that

• the sample size is reasonably large
• the random sampling technique is well-designed (simple, systematic, stratified, cluster, multi-stage)
• the questionnaires are designed to avoid bias
• the data is compiled accurately and experimental data is free of measurement bias

When you obtain secondary data, you must check to ensure they are reliable. Find out the author’s credentials, how up-to-date the information is, and what other researchers have cited the same data. Determine what bias might be inherent in the data-collection methods. Be especially wary of privately funded studies, which may put the interests of the organization funding the study ahead of public interest. Once you have established the credibility of the source, document key information: source name, applicable copyright information, when and where you accessed it (if it is a Web site), how the data were originally collected, and so on.

SURFING THE WEB

If you are unable to find suitable data in terms of quality or quantity using traditional sources, try searching the World Wide Web. Governments, charitable organizations, universities and colleges, corporations, sports teams, and individuals generate sites, which can be accessed via a computer and a Web-browsing program such as Netscape or Internet Explorer.
Data can be sought using either a Web directory or a search engine. In many cases, Web directories are preferable because the sites are pre-screened for quality. However, not all good sites are necessarily listed, and searching through those that are listed may take more time.

Whether you choose a Web directory or a search engine, be sure to structure your key words efficiently.

**Example 1 Finding Information on Fractals**
Use a Web directory to find information on fractals.

**Solution**
One of the most popular Web directories is Yahoo at [www.yahoo.com](http://www.yahoo.com). The following is a small portion of the initial categories and subcategories. (This information changes over time.)

- Reference
  - Libraries, Dictionaries, Quotations...
- Regional
  - Countries, Regions, US States...
- Science
  - Animals, Astronomy, Engineering...
  - Mathematics (1993)
  - Measurements and Units (242)
- Home->Science First subcategory:
  - Life Sciences (18)
  - Mathematics (1993)
  - Measurements and Units (242)
- Home->Science->Mathematics Second subcategory:
  - Chaos (27)
  - Combinatorics (17)
- Home->Science->Mathematics->Chaos Third subcategory (choose Fractals@ for a selection of sites with brief descriptions):
  - Contours of the Mind–Exhibition Page—celebration of fractal geometry, feedback, and chaos.
  - Fractal Explorer—simple tutorial on Mandelbrot and Julia sets with a nice image gallery.
**Search Engines**

Most people are familiar with using key word or phrase searches on the Internet. It is useful, however, to better understand what search engines do and then find out about how to use more powerful search techniques.

Search engine providers use Web “spiders” to reach out into the Web to find and collect data from:

- titles
- content found in the initial paragraph
- meta tags (lines of text hidden within a Web page’s HTML code)
- content from the entire document.

When a search is conducted, the “hits” are listed in order from the most relevant to the least relevant. Some search providers (e.g., [www.google.com](http://www.google.com)) use site-rating systems that are based on popularity (i.e., the ranking is determined by the number of Web sites that include it as a link), positioning the most popular site first. There are also metasearch programs (e.g., [www.metacrawler.com](http://www.metacrawler.com)) that will allow the user to search several providers simultaneously.

**Some Advice Regarding Searches**

- Words that are broad in scope may not be helpful (e.g., the word *Canada* will generate tens of thousands of hits).
- As you narrow your search, be on the lookout for words that may provide better results.
- Most search engines will allow you to limit your search to English sites only.
- Check through your search results thoroughly before trying again.
- Consider using some advanced search techniques to improve the quality of your results.

**Advanced Search Techniques**

It is important to understand how a search engine interprets the *query* you provide. Two words separated by a space will be treated inclusively. That means that the computer will return sites that use one word, the other word, or both. Sometimes, this is not what you want. You may need to use restrictions or *Boolean operators* to find what you are looking for. Both are useful in refining a search to include only pages that relate to your topic.

**Restrictions**

- If you want a specific word to be part of your search result, a plus sign in front of the word (with no space after the sign) will restrict the search results to Web sites that include this word.
- A minus sign immediately before a word is equivalent to using the word *not*. It instructs the computer to disqualify sites that include this word.
- Using quotation marks to surround a phrase will return sites that include the complete phrase as opposed to sites that contain each individual word.
Boolean Operators

AND (&) finds sites containing both words.

OR ( | ) finds sites containing at least one of the words.

() are used for grouping complex Boolean phrases.

Domain Restrictions

If you know the Web site, or family of Web sites to which you want to limit your search, use a domain restriction. For example, to search for the phrase national parks among all the Web sites that end in .ca, key in “national parks” +domain:ca. You can also restrict a search to a group of Web sites using the site: command. Searching bruins +site:yahoo.com will search for the word bruins among all the Web sites that end in yahoo.com.

Wildcards

By typing an asterisk at the end of a word, you can search for multiple forms of the word. For example, a query that includes big* will return sites that contain words such as big, bigger, biggest, bigwig, and so on.

Example 2 Advanced Internet Searches

Write the query for the following search requests.

(a) Find recipes for chocolate cookies without chocolate chips.

(b) Find sites that include the phrase The History of Algebra in the .org domain.

(c) Find information on Lions, but not about the CFL’s B.C. Lions.

(d) Find more information about peanut butter and jelly sandwiches, but also include jam as a possible synonym for jelly.

(e) Search for more information on bass (the fish, not the musical instrument, tone, or vocal register).

(f) Find biographical information on Gauss from somewhere on the Nelson Web site.

Solution

(a) recipe cookie +chocolate –chips

(b) “The History of Algebra” +domain:org

(c) Lions –(B.C. OR CFL)

(d) (Peanut AND Butter) AND (Jelly OR Jam)

(e) bass –music* +fish

(f) “Carl Friedrich Gauss” +site:nelson.com
2.6 Exercises

1. Describe the kind of results each of the following search strings will generate.
   (a) Thomas Engine +trains
   (b) Gandalf +site:members.tripod.com
   (c) Sharks –hockey
   (d) Canad* +domain:ca
   (e) (Hot & Dog) | (Ketchup | Mustard)
   (f) “Slim Whitman” OR “Slim Witman”

2. Knowledge and Understanding
   (a) Explain the difference between a Web directory and a search engine.
   (b) For each of the following scenarios, indicate which service would provide better results: a Web directory or a search engine.
      (i) researching the history of Trinidad
      (ii) finding discussion forums on novels written by J.R.R. Tolkien
      (iii) conducting research on a brand of cell phone
      (iv) researching cell phones in general
3. Write a query for each of the following search requests.
   (a) sites that sell CD burners
   (b) sites with information about statistical software
   (c) Canadian sites featuring literature
   (d) sites with information about your high school
   (e) the Ontario Ministry of Education Web site
   (f) sites where you can download work from independent Canadian recording artists

4. Improve the following queries so that they will return more useful results.
   (a) sites that contain quotes from Ralph Waldo Emerson: {Ralph Waldo Emerson +quotes}
   (b) sites with information about the altitude of Mount Logan: {what is the height of Mt. Logan?}
   (c) sites with historical information about the RCMP: {RCMP HISTORY}
   (d) sites within the Canoe Web site (www.canoe.ca) with biographical information about Emily Carr: {Emily Carr site:canoe.ca}

5. **Application** For each of the following topics, provide what you would type into a search engine (e.g., using the symbols +, –, &, |, and “ ”). Try each one and record what the top two hits are.
   (a) theatre in New York City
   (b) refugees in Kosovo
   (c) music, but not classical music
   (d) player statistics for the Toronto Blue Jays and the Montreal Expos
   (e) only Canadian mathematicians
   (f) illicit drugs, not including heroin or cocaine

6. You are seeking information about municipal libraries in Ontario. In detail, describe your strategies to find these data on the Web using both a search engine and a Web directory. Write down each step in the process and a short description of the top Web site that is found.

7. Using secondary data, find three sources of background information and numerical data about each of the following.
   (a) first-year university enrolment information (i.e., age of entrant, province of origin, country of origin, and so on)
   (b) Canadian youth purchasing habits in the late 1990s
   (c) temperature data dating back to 1900 from a variety of large Canadian cities
   (d) television-viewing habits of Canadian children from 1990 to the present
   (e) results from a local field hockey team
   Be sure to record the Web address (URL), the date you accessed the data, and your evaluation of the usefulness of each source.
8. Communication  Given the type of data and sources listed in each of the examples that follow, do you think the information you would obtain would be reliable or unreliable. Explain. Retrieve an example of the document (Web or library) to establish your case, if necessary.

(a) socio-economic data found on the Statscan Web site
(b) medical research found in the British Medical Journal
(c) crime data found in the National Enquirer
(d) archaeological information from the Royal Ontario Museum
(e) export data from the University of Western Ontario business school
(f) forest fire data from Lakehead University’s department of forestry
(g) environment information from NASA
(h) used car sales data from http://www.autobuyersadvice.com/
(i) earthquake data from Encyclopedia Britannica
(j) water quality data for the Great Lakes from Pollution Probe

9. Thinking, Inquiry, Problem Solving  Use at least two sources to find information for the topics that follow. Explain what you would do to determine whether the information that you have found is reliable.

(a) the cause of the depleting cod stocks in the Grand Banks
(b) immigration patterns into Canada between 1945 and 1990
(c) the effects of famine in northern Africa since World War II
(d) teenage tobacco use in Ontario over the last 20 years
(e) gambling in Canada over the last 20 years
(f) the economic impact of war on the population of Bosnia

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

10. Knowledge and Understanding  Describe Boolean operators and what they are used for.

11. Application  For each of the following situations, suggest two sources that could be searched to provide reliable data. Check to see if your sources produce the desired data. When you find two sources, decide which source has the most reliable data.

(a) the average income of Canadian females since 1960
(b) individual player statistics for the current Montreal Canadiens
(c) launch dates of the last five space shuttle flights
(d) the best-selling domestic car in North America over the last decade

12. Thinking, Inquiry, Problem Solving  Find three examples of an article, report, or study, and decide whether each bases its conclusion on unreliable data. Find a reliable data source that could have been used in this situation.

13. Communication  When searching for data from secondary sources, the reliability of the source must be taken into consideration. List several Canadian sources that could be considered reliable. What factors should you consider when judging the reliability of a source?
Chapter Problem
Mystery Most Mathematical—Part VII

Subject: Re: Puzzle - Part 7
Date: Wednesday 06:09:04 -0700 (EDT)
From: v=d?/dt@homework.com
To: jto@coldmail.com

Normally, a scavenger hunt requires the player(s) to find as many items as possible. For this one, choose only those five items that you find most interesting.

Use sources such as school and libraries, the Internet, local archives, post-secondary institutions, and city hall.

- the number of athletes participating in the first modern Olympics in Athens in 1896
- the average daily high temperature reading for July in Sault Ste. Marie, ON
- the dimensions of the Saturn V rocket that propelled Apollo 14 to the moon
- the female and male winners at the Wimbledon Tennis Championship in 1934
- the loss of life when the airship Hindenburg exploded over New Jersey on May 6, 1937
- the number of Rembrandts held in the collection in the Louvre in Paris
- the estimated population of China in 1960
- the number of goals, assists, and penalties in minutes for Maurice “The Rocket” Richard in the 1955/1956 season
- the amount of salmon (in kilograms) caught by British Columbian fishers in 1999
- the magnitude (on the Richter scale) and location of the largest earthquake in North America in 1992
- the number of road fatalities attributed to impaired driving in Ontario for 1995