

- Greer, Arlene, Lee Weston, and Mary Alm. 1991. "Assessment of Learning Outcomes: A Measure of Progress in Library Literacy." *College & Research Libraries* 52: 549-557.
- Grotzinger, Laurel. 1981. "Methodology of Library Science Inquiry." In *Busha A Library Science Research Reader and Bibliographic Guide*, 38-50. Littleton, CO: Libraries Unlimited.
- Gustafson, Kent L., and Jane Bandy Smith. 1994. *Research for School Library Media Specialists*. Norwood, NJ: Ablex.
- Katzer, Jeffrey. 1981. "Understanding the Research Process: An Analysis of Error." In *Busha A Library Science Research Reader and Bibliographic Guide*, 51-71. Littleton, CO: Libraries Unlimited.
- Katzer, Jeffrey, Kenneth H. Cook, and Wayne W. Crouch. 1997. *Evaluating Information: A Guide for Users of Social Science Research*. Reading, MA: Addison-Wesley.
- Lewis-Beck, Michael, Alan Bryman, and Tim Futung Liao. 2004. *The Sage Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage.
- Manfreda, Katja Lozar, Zenel Batagelj, and Vasja Vehovar. 2002. "Design of Web Survey Questionnaires: Three Basic Experiments." *Journal of Computer-Mediated Communication* 7(3). Available: <http://ascusc.org/jcm/vol7/issue3/vehovar.html> (accessed November 15, 2004).
- Nichols, James, Barbara Shaffer, and Karen Shockey. 2003. "Changing the Face of Instruction: Is Online or In-class More Effective?" *College & Research Libraries* 64: 378-388.
- Onwuegbuzie, Anthony J., and Qun G. Jiao. 2000. "I'll Go to the Library Later: The Relationship between Academic Procrastination and Library Literacy." *College & Research Libraries* 61: 45-54.
- Paulos, John Allen. 1988. *Innumeracy: Mathematical Illiteracy and Its Consequences*. New York: Hill and Wang.
- Peng, Chao-Ying Joanne. 2003. "Experiment." In *The Sage Encyclopedia of Social Science Research Methods*, 349-354, edited by Michael Lewis-Beck, Alan Bryman, and Tim Futung Liao. Thousand Oaks, CA: Sage.
- Powell, Ronald R. 1999. *Basic Research Methods for Librarians*. 3rd ed. Greenwich, CT: Ablex.
- Random House Webster's College Dictionary*. 1991. McGraw-Hill edition. New York: Random House.
- Rice, James. 1981. *Teaching Library Use: A Guide for Library Instruction*. Westport, CT: Greenwood Press.
- Ware, Susan A., and Deena J. Morganti. 1986. "A Competency-Based Approach to Assessing Workbook Effectiveness." *Research Strategies* 4: 4-10.

Chapter 7

Bibliometrics

"Bibliometrics" comes from the Greek words for book (*biblos*) and measure (*metron*), suggesting its focus as a research methodology that "measures books." Bibliometrics, however, is not just any approach to quantification of publications. As Paisley (1989, 707) emphasizes, bibliometrics focuses on

extrinsic facts about publications, broadcasts, and other forms of communication. For example, an article is coded by when and where it was published, by the identity and affiliation of the author, by the other articles it cites, and so on. *Intrinsic* content of the article is thought to be the domain of content analysis because of the need to develop coding categories based on a theory of the relationship of the text to intentions, effects, and the symbolic environment.

In other words, while other research methodologies are concerned with the content and interpretation of recorded information sources, bibliometrics concentrates upon those aspects of sources that do not require engagement with or interpretation of the sources' content—aspects such as

- author's affiliations, educational credentials, or geographic location
- number of circulations or re-shelvings of printed materials
- usage statistics on electronic resources
- which sources cite and are cited by other sources; or the number and types of linkages between Web pages

Essentially, if you can count something without too much debate over what it is—and it has something to do with any type of recorded information source—it could become the subject of a bibliometric study. Some people also use the terms "Webometrics," "cybermetrics," "scientometrics," or "infometrics" to refer to bibliometric studies. These different terms largely connote differences in the types of recorded information sources to which a common methodology is applied; Webometrics thus looks at Web pages, while scientometrics tends to focus on recorded information sources for scientific disciplines.

The relationship between bibliometrics and content analysis is complex. Some have equated bibliometrics and content analysis (Miller and Stebenne 1988), while others have made bibliometrics a type of content analysis (Bibliometrics 2005), and yet others have described the two as separate methodologies (Paisley 1989). This book obviously takes the latter view, largely because content analysis requires reading and often interpretation of sources' content, while bibliometrics does not.

Bibliometrics is one of the oldest research methodologies in library and information science. Histories of bibliometrics often trace its earliest use to 1913, when Felix Auerbach first articulated Zipf's law while formulating a relationship between the rank and size of German cities (Rousseau accessed 2007). It seems probable, however, that at least some studies prior to 1913 used counting of extrinsic facts about publications to test their hypotheses or support their arguments. Despite its history, bibliometrics is not without its critics. Their objections tend to center upon what Borgman (1990, 11), herself a frequent user of bibliometrics, describes as the methodology's "application of mathematics and statistical methods to books and other media of communication." Some critics question the degree to which citations and other aspects of recorded communications actually and accurately reflect the intentions and behaviors of their human authors. Showing that a group of co-authors cited sources X, Y, and Z, for example, does not prove that they thought X, Y, and Z were the best sources on that topic. These co-authors might have preferred A and B as their sources, but they could not find their personal copy of A while they were writing, and they waited until the last minute to do their work—just to find out that B could only be gotten by interlibrary loan.

Other critics claim that bibliometrics obscures individual experiences and differences within numerical averages (Edge 1977, 15–16). While one author might have cited source X in the belief that it was the "best" source on a topic, another author might have cited it simply because a coworker wrote it, and yet another author might have cited it because a peer reviewer noted its absence from his bibliography. Emphasizing that all three authors used the same source hides their very different reasons for doing so. Still other critics object to the typical focuses of bibliometric studies on formal scholarly publications, as well as the tendency of bibliometrics to view scholarly communication as predictably patterned. Edge (1979, 113) claims that "in emphasizing formal communication through the published literature, quantitative methods [such as bibliometrics] perpetuate a 'rationalized' view of the nature of [scholarship]."

Most bibliometricians do not fundamentally dispute these claims. Rather, they reiterate one of bibliometrics' key assumptions: "A scientific paper, article, or book is a rich resource of data on . . . communication patterns and cognitive processes" (Parker and Paisley 1966, 1067). Note that Parker and Paisley claim only that bibliometrics is "a rich source of data," not that it should be the only

source of data. Indeed, it is not uncommon for researchers to use bibliometrics in conjunction with other methodologies better suited to gauging individual perspectives and differences.

Provided that beginning researchers do not assume they have learned everything there is to know about a topic from a single bibliometric study (for reasons that will become clearer later), bibliometrics can be a particularly good methodology for a first research project. Several characteristics of bibliometrics make it easier for beginning researchers to get started with bibliometrics than with some other methodologies. First, bibliometrics is unobtrusive; its focus is upon the products of human activity (books, articles, Web pages, and so on), not upon humans themselves. This means that there is no need to control for experimenter, interactional investigator, or other similar effects arising from the influences of researchers and human subjects upon each other. Institutional review board approval is also typically not required for bibliometric projects, removing another hurdle to getting started. Second, bibliometricians' data sources preexist the study, and they are usually readily accessible. Third, sampling is generally only of the simple random or systematic sort, not stratified. Fourth, bibliometricians work from a number of shared operational definitions, research designs, and measurement instruments, freeing novice researchers from the complexities of constructing these from scratch. Fifth, data tends to be numerical, meaning that it is less enmeshed in interpretive ambiguities. However, although the data is numerical, knowledge of inferential statistical techniques is not required for interpreting or presenting it. All of these factors make bibliometrics one of the more straightforward ways to get started in research.

FINDING A TOPIC

As one of the oldest and most commonly used methodologies in library and information science research, bibliometrics displays a wide range of topical applications. Broadly speaking, bibliometric studies can be categorized into the following four groups:

- *Studies that seek to learn about information sources*, such as the contents and functionality of different databases or the time frames within which scientific research results in different types of publications (conference papers, preprints, journal articles, review articles, and so on). When focused upon the value of individual journals or other publications to specific disciplines and fields, such studies are often used in institutional decision making about what to acquire, keep, discontinue, or weed. The study by Fosmire and Yu (2000) listed in Figure 7-1 is an example of a bibliometric research project that seeks to learn more about information sources *per se*, while the study by Smith (1981) describes

bibliometric studies of information sources in support of local decision making.

- *Studies that seek to learn about institutional trends*, such as the impact of decreased library spending for print monographs upon patrons' use of library resources over time. These studies often factor in institutional decision making, particularly collection development, and in outcomes assessment projects. Smith's study (listed in Figure 7-1) of whether the usefulness of the University of Georgia Library's collection has changed over the past ten years because of the introduction of electronic resources and increases in the periodical budget is an example of such a study, as are the studies by Edwards (1999) and Walcott (1994).
- *Studies that seek to learn about people's behavior*, such as what sources undergraduate students, laboratory scientists, or other researchers use in their work, or to what degree researchers from different fields, institutions, countries, professional ranks, or sexes co-author publications. Especially when focused on the products of undergraduate researchers, some such studies are conducted for assessment purposes (e.g., seeing how the sources cited relate to changes in assignments, instructional methods, or materials). Bahr and Zemon's (2000) study listed in Figure 7-1 examines the extent of co-authorship within academic librarianship, especially the number, gender, and institutional settings of co-authors.
- *Studies that seek to learn about socio-intellectual phenomena*, such as the formation of disciplines or of interdisciplinary research fronts, the spread of ideas between disciplines or geographic regions, or the impact of print, electronic-by-subscription, and open access publishing models on scholarly communication. Youngen's study (1998), listed in Figure 7-1, for example, tracks the increasing acceptance of electronic preprints in the physical sciences.

Fosmire and Yu's article, "Free Scholarly Electronic Journals: How Good Are They?" published in *Issues in Science and Technology Librarianship* in 2000, illustrates many typical aspects of bibliometric research. Fosmire and Yu conducted their research as a follow-up to a study by Harter five years earlier on the impact of open access journals on various scholarly fields. Like Fosmire and Yu's study, most bibliometric studies draw heavily upon prior studies in finding their research topics (and in interpreting their data, as will be seen below). The reason for this is simple: a single, self-contained bibliometric study is no more widely interesting or generally informative than a photograph of particular people at a particular time in a particular place.

Suppose you found a photograph from the 1960s of your parents' college roommates at Ohio University. You would probably be somewhat interested in the photograph because it showed people connected to your parents; others who graduated from the same institution or lived through the 1960s might also have

some interest. Most people, though, would have no interest in the photograph based simply on its content. Its meaning would interest more people, but it is hard to supply this meaning without knowledge of the time, place, and people in question. Had people always dressed like that? Do we dress differently now? Did students in the U.S. do their hair in the same way as students in France or China? Was roommate Bob's hairstyle a relic of the 1950s and completely atypical for 1960s hairstyles? Knowing more about the context of your photograph helps make it more interesting and understandable to others. Similarly, situating bibliometric research projects within the context of prior studies helps give meaning to the data—and to see what is researchable. Bibliometric studies tend to be particularly concerned with addressing whether what was true then holds true now, as well as whether this category of authors resembles that category of authors in the materials they cite and produce. Fosmire and Yu ultimately found that now, unlike five years ago, there "are several free scholarly journals that have a significant impact on their respective fields."

Fosmire and Yu (2000) are also typical of other bibliometricians in their operationalization of research terms by using established definitions for topics of interest. They take their definition of "impact factor" from the Institute for Scientific Information® (ISI), whose definition of impact factor (as the number of citations to articles published in a particular journal in a two-year period divided by the total number of articles published in that journal in the same period) is widely used by other bibliometricians. They also used standard directories of information sources to select their research population. By focusing on science, technology, and mathematics journals listed in the *Directory of Electronic Journals; Newsletters; and Academic Discussion Lists* (7th ed.), they protected themselves against skewing their study's findings through a personal and idiosyncratic population selection. Because there were only 85 such journals, they studied the entire population, not just a sample. If they had been forced to sample, they—like most bibliometricians—would likely have chosen a simple random or systematic sample. Because "impact" was defined in terms of the number of citations to a journal, Fosmire and Yu searched Web of Science® to determine the number of citations to each journal. Citations, or references by one work to other, earlier works, are a particularly common source of data in bibliometric projects. Fosmire and Yu were well acquainted with how Web of Science® works (e.g., it "searches citations based on a 20 character code"), and they described their searching or data gathering mechanisms in detail ("for example, *Emerging Infectious Diseases* was searched as 'eme* inf* dis*'). Such knowledge of data-gathering tools and detail about data-gathering procedures is essential to ensure the validity and reliability of results. If the database's workings skew data gathering, then the validity of the study is questionable; if insufficient details are provided about how the data were gathered, the study's reliability is questionable. Fosmire and Yu present their data in seven tables that

give the impact factor, immediacy index, and total number of current articles for each journal. This data underlies the study's conclusion that "overall, it appears that several high-quality, productive, free scholarly electronic journals exist currently. These journals scored very well in impact factor and immediacy index, and they have reasonable numbers of articles published."

Figure 7-1 lists some other exemplary bibliometric studies by practicing librarians and information scientists.

Figure 7-1: Studies Using Bibliometrics

Bahr, Alice Harrison, and Mickey Zemon. 2000. "Collaborative Authorship in the Journal Literature: Perspectives for Academic Librarians Who Wish to Publish." *College and Research Libraries* 61 (5): 410-419.

Researchers used articles published in College & Research Libraries and the Journal of Academic Librarianship between 1986 and 1996 to track co-authorship among academic librarians. Tracked number and percentage of co-authored articles in each journal per year, also tracked the number, gender, and institutional setting of co-authors. Included a lot of information about prior studies of co-authorship in librarianship and other fields.

Davis, Philip M. 2005. "The Ethics of Republishing: A Case Study of Emerald/NCB University Press Journals." *Library Resources and Technical Services* 49 (2): 72-88.

Researcher examined degree to which Emerald (formerly MCB University Press) engaged in republication without notification, as well as whether articles were republished in journals with the same or similar subjects. Identified a number of republished articles via keyword searches, then examined them for notices about republication and tracked the journals in which they appeared. Also included some data about the library holdings of the journals republishing articles.

Fosmire, Michael, and Song Yu. 2000. "Free Scholarly Electronic Journals: How Good Are They?" *Issues in Science and Technology Librarianship*. Available: <http://library.ucsb.edu/istl/00-summer/refereed.html> (accessed May 18, 2007).

Researchers calculated the impact factor and immediacy index of 85 science, technology, and medicine journals listed in the Directory of Electronic Journals, Newsletters, and Academic Discussion Lists (7th ed.) as a way of determining whether free scholarly electronic journals have more impact than they were found to have in a study by Stephen Harter five years earlier. Impact factor and immediacy index were based on standard calculations of these constructs.

Germain, Carol Anne. 2000. "URLs: Uniform Resource Locators or Unreliable Resource Locators?" *College & Research Libraries* 61 (4): 359-365.

Researcher randomly selected 31 journal articles published in journals in various fields (library and information science, science, computer science, humanities, and social sciences) between 1995 and 1997. All citations with URLs (N=64) in these articles were checked as to the persistence of the URL cited every three months for a three-year period (1997-1999). Tracked number and percent of articles with inaccessible citations by year.

Rhodes, Jo Ann. 1997. "Sentimentality? An Exercise in Weeding in the Small College Library." *The Christian Librarian* 40 (1): 16-17 and 20.

Researcher tracked number of circulations by call number range for one year after a project to reclassify some 50,000 items left as a separate collection when the library joined OCLC in the mid-1970s. Data on the age of the items as well as their eventual circulations underlie conclusion that some of reclassified materials might better have been withdrawn.

(Cont'd.)

Figure 7-1: Studies Using Bibliometrics (Continued)

Smith, Erin T. 2003. "Assessing Collection Usefulness: An Investigation of Library Ownership of the Resources Graduate Students Use." *College & Research Libraries* 64 (5): 344-355.

Researcher examined up to 75 dissertations from 30 dissertations in four subject fields (education, social sciences, sciences, and humanities) from 1991 and from 2001 to see the types of materials cited (e.g., book, newspaper, etc.), as well as the percentage of cited materials that were locally owned. Study intended to help in evaluating the "fit" of the University of Georgia Library's collections with the needs of its patrons.

Youngeun, Gregory K. 1998. "Citation Patterns to Traditional Electronic Preprints in the Published Literature." *College & Research Libraries* 59 (5): 448-456.

Researcher used ISI's Scisearch® database to track the number of journals publishing articles with citations to preprints and e-prints, as well as the overall number of citations to preprints and e-prints over the past ten years. Trendlines from the data show increasing citations to and acceptance of e-prints.

FORMULATING QUESTIONS

In part because the total number of publications in recorded history is so vast, bibliometric research projects involve quite narrow research questions. For example, instead of studying the frequency of citations to various types of Web and print resources by all researchers, a bibliometric research project may limit its focus to the frequency of citations to various types of Web and print resources by first-year students, undergraduate history majors, or practicing chemists. As Figure 7-2 suggests, there is simply too much recorded information to look at all examples of anything unless that "anything" is something quite small. (For example, when Fosmire and Yu [2000] did their study, there were only 85 free scholarly electronic journals in the fields of science, technology, and mathematics. Eighty-five is, in itself, a researchable number, but the number would not have been so researchable had they looked at all open access journals.) Bibliometricians also use such narrow research questions because publication and citation patterns generally hold true for only limited times, places, and populations. While there is an overall phenomenon consisting of the frequency

Figure 7-2: Estimates of the Total Amount of Recorded Information

Source Type	Yearly Quantity of Information
Books	39 terabytes
Newspapers	138.4 terabytes
Mass market periodicals	52 terabytes
Journals	6 terabytes
Newsletters	0.9 terabytes

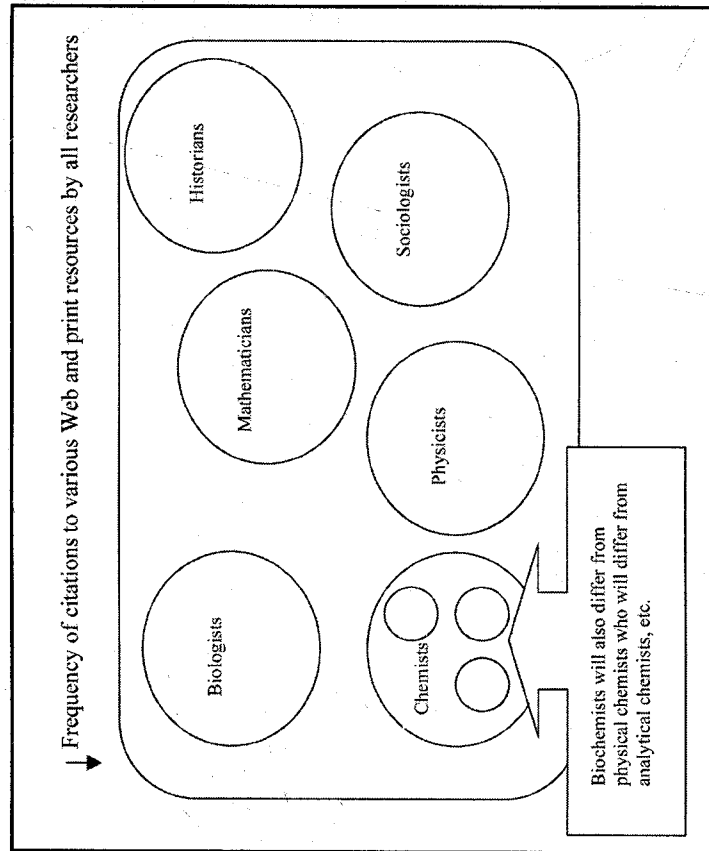
This figure is based on information from Lyman and Varian (2003).

of citations to various types of Web and print resources by all researchers, this overall phenomenon is composed of smaller phenomena—the frequency of citations to various types of Web and print sources by specific types of researchers—that comprise but are not identical to the overall phenomenon. Figure 7-3 illustrates this in more detail.

While chemists might have something in common with physicists or biologists in terms of the frequency with which they cite various types of Web and print sources, they do not have everything in common with physicists or biologists, and they have little in common with historians or mathematicians. Even within chemistry, there may be sizeable differences between biochemists and physical chemists in the frequency with which they cite various types of Web and print sources. Failure to treat different communities as distinct in gathering and interpreting data would yield conclusions that, while purporting to include everyone, actually represent no one.

Many bibliometric studies are purely descriptive. They seek to do no more than provide numbers (such as frequency counts or percentages) that help to illuminate the topics they are discussing. All of the studies from Figure 7-1 are of this type.

Figure 7-3: Bibliometrics' Focus on Small Questions



- Bahr and Zemon (2000) provide raw numbers and percentages of co-authored articles in two journals, as well as raw numbers and percentages of co-authors by sex and institutional status.
- Davis (2005) provides frequency counts and title lists of Emerald journals that have republished articles without notification.
- Fosmire and Yu (2000) give the impact factors, immediacy index, and total number of articles published over a two-year period by open access journals in science, technology, and medicine.
- Germain (2000) tracks the total number and percentage of articles citing to inoperative URLs over a three-year period.
- Rhodes (1997) provides total numbers and percentages of reclassified volumes by age and circulation count.
- Smith (2003) gives the number and percentage of source types cited in dissertations, as well as the number and percentage of cited items locally owned.
- Youngen (1998) gives the number of citations to preprints and e-prints in journals included in ISI's (Institute of Scientific Information®) SciSearch.

These studies use their data in drawing conclusions about the topics studied, but they generally do not have formal hypotheses, nor do they subject their data to tests of statistical significance. Some bibliometric studies do have formal hypotheses (e.g., circulation percentages of books selected by faculty will be higher than those selected by librarians, or open access will lead to higher impact factors for articles). Bibliometric studies can also use tests of statistical significance, such as chi-square or Analysis of Variance tests, on their data. However, it is precisely because bibliometricians work so often without formal hypotheses or tests of statistical significance that they spend so much time replicating studies. Another researcher taking and interpreting a different "picture"—or set of data—will necessarily reach different conclusions, a fact which makes the taking of multiple "pictures" important. The more "pictures" that show the same thing, the more sure researchers can be of their interpretations.

DEFINING THE POPULATION

The immense number of publications and the differences in publications over time, space, and socio-intellectual groups also helps to explain why bibliometricians focus on specific populations. Some populations studied by bibliometric researchers in library and information science include:

- articles published in *College & Research Libraries* and the *Journal of Academic Librarianship* between 1986 and 1996 (Bahr and Zemon 2000)
- electronic-by-subscription journals published by Emerald/MCB University Press (Davis 2005)