JANE L. CALDWELL
Coordinator of Professional Activities
Bibliographic Retrieval Services

Management Information from Bibliographic Databases

This year's clinic dealing with the use of information for library management is very exciting. Perhaps the most exciting element is the fact that we are not talking in futuristic terms. The advent of the personal computer means that almost any library can afford an in-house database as well as access to outside databases. Sophisticated software, such as BRS/SEARCH, will soon be available for use on one's local mini- or microcomputer. This promises local access to the online catalog, an automated circulation control system, serials check-in, and the ability to store and manipulate library statistics and other types of management information. Libraries will have much better control of their own statistics and have surer methods to determine collection usage, user satisfaction, employee performance, etc.

Today, I am going to take a somewhat different, I hope, approach to the subject of management information than has been taken thus far, and describe information that is already available, easily manipulated and frequently overlooked. This information is available to all of those whose libraries are now accessing online databases through one or more of the vendors of online databases. Some of those vendors include BRS, Dialog, NLM, and SDC. There is a wealth of information available to the user of online databases that they may have overlooked.

First of all, what is a database? The simple definition is that it is a collection of data in machine-readable form. This data could be available in a batch mode or an online mode. I will limit my discussion today to online databases that are available on a dial-up basis from one of the search services I have mentioned. When I say that a database is available online I mean that it is possible to query the database, get results immediately and

interact with that information all from a local terminal. An online data-base from one of the search services could have a hard-copy counterpart such as the ERIC database that corresponds to the print indexes of *Resources in Education* and the *Current Index to Journals in Education*. Sometimes there is no exact print counterpart as in the case of the Dow Jones database; sometimes, there is no print counterpart at all, as in the case of the Pre-Med databases. A database may be bibliographic, meaning that it includes citations, often with abstracts or summaries, to printed journals, monographs, technical reports, patents, etc. A database may be nonbibliographic, meaning that it contains data, usually numeric or directory-type information. A database may be full-text, meaning that the full journal article or newspaper story is available online. Examples would be the American Chemical Society's project to put their sixteen chemical journals online. Another would be the full text of the *Harvard Business Review* that will be available online from BRS. In libraries today, all types of databases are being searched to produce comprehensive bibliographies or to answer specific reference questions.

Online database searching began, at least in an experimental mode, in the early 1960s. In the late sixties some government-produced files, notably that of NASA, were the first to be made available in an "operational" mode. In the early seventies, commercial search services offered access through telecommunications systems to, at first, only a few databases. The number of databases has rapidly expanded and from one's terminal one now has access to hundreds of different databases. When online databases first became commercially available, the softwɑre was not nearly as flexible as it is now, response time as short as it is now, and the costs for an online search as low as they are today. When online searching first became popular, databases were used primarily to produce comprehensive bibliographies on specific topics. Later, as online search analysts became more proficient and searching became an accepted research tool, online databases were regularly accessed to answer the "quick" or "ready" reference questions.

Why use an online database to answer a reference question that perhaps could be answered from a directory or encyclopedia? The answer lies in the fact that with an online database, all the information from the record—some information that may never go into the printed version—is directly *searchable*. (There may be variations between the vendors of online services—it depends upon the way that the database has been structured by the vendor and what information has gone into the dictionary or inverted file, or basic index.) Information is available in an online database that is not available in a printed index (e.g., titles, journal citation, publisher's name, price, author's affiliation, funding agencies, etc.). Often this information could be accessible through a printed source, but it would be time

prohibitive to find this information. Besides the obvious advantage of Boolean operators—*OR, AND, NOT*—or pull together synonyms or combine and remove unwanted concepts—positional or proximity operators allow the searcher to be more specific or to search for concepts that are too new to have standardized terminology.

So far, we are talking about what I consider to be the traditional approach to the use of online databases. Databases are primarily used to produce comprehensive bibliographies with the costs often passed back to the user. Databases are frequently used for reference purposes, often with the charges absorbed by the reference department. Usually money is allocated solely for this purpose and it is up to the reference department to determine if accessing an online database is the appropriate way to answer a specific question.

Not long after the introduction of online database searching, many librarians, and researchers in other fields as well, discovered that there were ways that online databases, and the statistics they generated, could be approached in less traditional ways to provide all types of information. Donald Hawkins wrote a paper in 1977, published in *JASIS*, describing some of these unconventional uses of databases. The title of his paper was "Unconventional Uses of On-line Information Retrieval Systems: On-line Bibliometric Studies."[1] In it, he described ways to use databases to quantitatively analyze the bibliometric features of the literature. By using online databases we can explore, at no great expense in terms of time or money, aspects of the literature in a specific discipline. Some of the characteristics of the literature that we can explore that will have great impact on one's life as a library decision-maker include information of an evaluative nature— particularly with applications for collection development. Other information is available such as information useful for planning future use of the library collection, information to plan and budget for use of library services, and information to use in predicting trends in book and journal publishing. Information from online databases can also be used to evaluate personnel—both library personnel and staff outside the department.

The obvious use of databases for library management is using databases with the subject emphasis on management, such as *Management Contents* or *ABI/INFORM*.

In my talk today, I want to take a very practical approach and describe ways that one can use online databases to provide information that has great implications for library management and for management of one's institution. I am not going to talk about the obvious—using databases with the subject emphasis on management, such as *ABI/INFORM* or *Management Contents*, that provide references to printed literature on management topics. Rather, I am going to discuss the use of databases to quantitatively explore aspects of the literature that have implications for

collection development, for determining how clientele use the services, and how their use of the literature will impact one's library. I am also going to describe how the statistics generated by online database searching can be effectively used for planning, budgeting and evaluating one's library services. The information I am going to discuss today is available from other databases and other search services, but I will be speaking primarily of access to BRS databases. I am, of course, more familiar with BRS databases and slightly prejudiced toward BRS as an innovative search service. However, I do believe that BRS has been the most responsive to the needs of managers in developing extra services such as our collection devlopment service, faculty bibliography service, electronic newsletter service (BRS/Alert), and our online accounts file.

Getting back to the primary subject, I would like to show how online databases can be used for different evaluative or bibliometric studies. Online databases can be used for citation analysis which could have valuable applications in collection development.

Bibliometrics has existed as a field of study since 1917. Bibliometrics became more sophisticated when Bradford, in the 1930s, established a basic law of bibliometrics—Bradford's Law of Scattering. Bradford discovered that, by arranging journals in descending order of productivity of articles on some subject, it is possible to identify a "core" of journals that publish most of the articles on a particular subject. Trueswell took this one step further and suggested an 80/20 rule—20 percent of the journals in a given field account for 80 percent of the relevant literature. The manual work that can go into identifying some core list is beyond the resources of most library managers. Online systems can provide a method to simplify the identification of this core. Here is a method developed in part by Donald Hawkins, Jane Caldwell and Celia Ellingson:

1. Create a set of all journal articles on a topic. This would be a broad search to account for variations in terminology. Use paragraph qualification to make sure the articles themselves will deal with the subject matter rather than having the terms appear in the journal titles.
2. Sort by journal titles (if this feature is available on your search system—it is on BRS) and print journal citations.
3. Run the same search on all relevant databases.
4. Rank order journals by number of articles on topic.
5. Identify a core journal list/secondary journal list.
6. Compare lists with your own holdings to determine how well the core journals and secondary journals are represented.[2]

This approach will identify the core journals in a specific subject discipline. This is an abstract list in that it will not necessarily correspond to users' needs. BRS has taken this approach one step further and provides a

special service to assist managers in collecting relevant data for quantitative collection analysis. BRS will save all citations that are requested from offline print requests for one calendar year. These offline prints are the ones that have been generated in response to users' queries and should represent the needs of actual users. Because users often use online services when looking for information on new or "hot" topics, the collection development service will represent current demands on the literature. BRS will save the journal citations and generate an annual or semiannual report, depending upon the service option that was chosen. The report, broken down by database, will consist of two sections: (1) an alphabetical listing by journal title; and (2) a rank order by frequency of citation.

In my former life (before joining BRS) I was at the University of Minnesota's Education/Psychology/Library Science (EPLS) library. One of our continuing projects was the evaluation of our collection and we used the BRS collection development service as one of the tools to judge the strength of our journal collection and to identify specific titles that should be considered for inclusion in the collection. This study was reported by Celia Ellingson and Lori Hedstrom. Their methodology:

1. Select databases that are primarily related to the collection and are used enough to warrant further analysis. This limits the focus to ERIC, Exceptional Child Educational Resources, National Clearinghouse for Mental Health, and Psychological Abstracts.
2. Eliminate from study journal titles that were cited fewer than six times. This threshold was perhaps arbitrary, but it relates to the fact that interlibrary loan can be used for the first five requests.
3. Check each title with six or more citations for holdings information and location within the University of Minnesota libraries. Journals that were not located within the EPLS library, but had multiple campus locations, were noted for their location most physically available to to EPLS.
4. Count the total number of titles to determine the percentage held within the EPLS library, held in other campus locations, and not held. Example: Online serials list, like the California Union List of Periodicals (CULP), could provide an even simpler alternative. Example:

ERIC

Total	544		
Held at EPLS	364	66.9%	
Held elsewhere	93	17.0%	} 83.9%
Not held	87	16.1%	

NIMH

Total	66		
Held at EPLS	34	51.5%	
Held elsewhere	25	37.8%	} 89.3%
Not held	7	10.7%	

5. Compile a merged list of all journals held outside the EPLS library, but within the University of Minnesota (twin cities campus) libraries. Example:

Architecture Library
 Planning for Higher Education (21)
 Progressive Architecture (14)
 Environment and Behavior (10)
Biomedical Library
 Journal of Autism & Childhood Schizophrenia (359)
 American Journal of Psychiatry (145)
 Journal of Medical Education (102)
 British Journal of Psychiatry (90)
 Behavior Therapy (80)
 American Academy of Child Psychiatry Journal (71)
 Journal of Nervous & Mental Disease (71)
 Hospital & Community Psychiatry (67)
 G/C/T (91)
 Bureau Memorandum (56)
 Pointer (56)
 AAESPH Review (50)
 Australian Journal of Mental Retardation (42)
 Psychologie a Patopsychologie Dietata (33)
 Child: Care, Health & Development (32)
 Teaching at a Distance (30)
 Special Education in Canada (29)
 United States Air Force RHL Technical Report (28)

6. Compile a merged list of all journals that were not held by any University of Minnesota library (checking holdings was a relatively simple process because of the existence of the *Minnesota Union List of Serials*. A student assistant was able to do the checking during his/her spare time).[3]

  The University of Minnesota is one of the many libraries across the country that has taken advantage of the BRS collection development service. To date, most users have come from academic libraries, although there would be many advantages for corporate libraries.

Before I go on, I would like to review the assumption on which the collection development service is based and that is that citations retrieved in literature searches represent the actual demands of our clients. We have to assume that our users take the journal literature and the literature search seriously enough to want to read the articles that are retrieved. We have to assume that the users of the literature search service are representative of all our users. We also have to assume that the searches were relevant to the subject and that the literature represented was actually about the desired subject.

Surveys like the one from the University of Minnesota point out several uses for the BRS collection development service. It can be used to:

1. Judge the relative strengths of holdings within a specific library or group of libraries.
2. Identify core journal titles that should be acquired by a specific library.
3. Identify low-demand titles that could be cancelled if necessary.
4. Compare the requests of our users with other types of surveys that identify core literature in a specific field.[4]

The collection development service provides a quantitative measure of potential use of a library collection. It is not meant to be used without the professional judgment of the librarian or subject bibliographer. There are, of course, other factors to consider—i.e., the service measures only the requests from users of a literature search service. In libraries that charge for access to the service, these statistics would not represent needs of all users. In an academic library, undergraduate students would not be as well represented as graduate students and faculty. Rich departments that could afford more literature searches might be overrepresented. And, most importantly, the service is valid only in disciplines where online databases adequately cover the published journal literature and where journal publication is the accepted means of communication between scholars. The collection development service provides a quantitative measure of potential user demand and, when used in conjunction with qualitative measures, provides a rational method of collection evaluation.

Another online tool that can be used for a bibliometric study is the overlap study. Several years ago, I was involved in a study the goal of which was to compare overlapping retrieval between the ERIC and *Psychological Abstracts* databases. Before our study, we knew that there should be a high percentage of overlap between the two files in subject areas—such as educational psychology—which were equally represented in both databases. Our findings, published in the journal *Database*, June 1979, were surprising.[5] The percentage of citations that were duplicated...was less than might be expected based upon the percentage of overlap in journal coverage between the two databases...." After comparing the journal titles

produced from the retrieved citations with the list of journals the database producers claimed to index, we discovered that "more than twice the number of citations actually duplicated could have appeared as duplicates." Our findings in the case of ERIC and *Psychological Abstracts* indicated that for these two databases, a search in a cross-disciplinary topic warranted a multiple file search.

At the time of our study, it was necessary to do a time-consuming comparison between the results of the several online searches. BRS now offers a unique feature—MERGE—that allows the searcher to merge together searches from multiple files. Duplicate citations can be easily revealed, and in addition to the obvious convenience offered to the consumer of the search, the MERGE feature offers valuable information on overlap between databases.

Often, it is a management decision to set policy on whether multiple files are routinely accessed to answer specific queries. The merge feature can assist managers, after studying results of various searches, to determine whether multi-file searches are necessary. In most library settings, it is often necessary to elect the most cost-effective single database search. Overlap studies and use of the BRS merge feature can be very useful to justify these decisions.

I have discussed uses of online databases to evaluate journal collections and the content of various indexes. There are several other nontraditional uses of online databases that can provide insights into the research habits of your clientele and thus better determine priorities of library service.

It is possible to use online databases to discover the research interests and citation habits of your clientele. Author searching can give you the subject specialities of your clientele, but by examining their citation habits through the use of a citation index like *Science Citation Index* or *Social Sciences Citation Index* you can unearth their ongoing research interests, the journals that they regularly scan (and probably subscribe to), and other authors that they favor. It is also possible to see if a researcher cites the same papers over and over again, or if he is adventurous and strays from his standard sources. It is then possible to speculate as to whether your library is really used!

Trends in publication can be determined by use of databases like *Books in Print*, Brodart's *Booksinfo* file and *Ulrichs*. It is possible to pull up all books recently published by a particular publisher to scan quickly their current output, or to take the opposite approach and do a subject search and determine which publishers favor particular subjects. Price information is searchable and within the capability of BRS files so it is also possible to use online databases to determine the average price of a book in a particular subject area, to determine what percentage of books cost more

than a specific price, etc. It is possible to use online databases to determine how much literature exists in a topic before adding a new program or course in an academic institution. Online searching can provide the decision-maker with a starting point for justification of a book budget.

If the researcher cites papers from journals held by your library and not readily available elsewhere, or cites journals that you know he has requested or that have appeared in a computerized literature search, you can be relatively sure that the researcher has been using your library services.

To identify the research interests and publication output of an organization, search under the institution name to determine its publication output, see in what format it usually publishes, its subject area strengths, and related interests. Searching online is often the only way that institution names are accessible and it provides a quick and easy way to measure an organization's research activity. Do not forget to use files that cover patent information—often patents are the best measure of ongoing research in technical fields.

You can determine the "invisible college," or group of scholars who communicte directly with each other, by using cocitation analysis. Cocitation analysis is a sophisticated technique. For the purpose of discussion, we will simplify it to an examination of which authors cite each other. Because these cocitations are often based upon prepublication correspondence between experts, it is possible to speculate upon what core of researchers form an "invisible college." Computerized searching can make it possible to crack this core of expertise. Databases like *Science Citation Index* and *Social Sciences Citation Index* can be used to determine cocitation. Databases like *American Men and Women of Science* can be searched to discover experts in a specific field.

A different, and sometimes controversial, use of online databases is to use them to gather supporting evidence for faculty promotion and tenure decisions. Online databases provide a cost-effective method to quickly pull together complete bibliographies of all members of a department, whether in a university or other research organization. Individual productivity can be measured. The strengths of an entire department or organization can be evaluated and compared with similar departments or organizations. Databases that index material other than periodical literature, such as ERIC, NTIS, DOED, or SSIE can show evidence of the ability to successfully compete for research grants.

Citation indexes provide a way to determine the relative value placed upon the research produced by one's own faculty or researchers. Relative rankings or accreditation surveys of departments or of graduates of departments can be compiled by examining how frequently the department faculty or graduates are cited. Although it is possible to produce similar

evaluation studies manually, online searching makes this type of survey economically feasible.

I have discussed how various types of "bibliometric" studies can be used to provide management information in libraries. Another aspect of online database searching is that it provides a sometimes overwhelming amount of information to library managers in the form of statistics generated by the online search service. The statistics, a byproduct of the service, provide information on who uses one's services and why they use them. The search statistics can provide cost figures—how much an average search costs in connect time, print charges, and personnel time—that can help the manager justify a budget. The annual increase in search volume can help one argue for an increase in staff—or at least justify the present staff. Sometimes it is necessary to use search statistics to determine which search analysts do not make cost-effective use of online time—perhaps it is necessary to invest in additional training.

I would like to review some of the information that should go into a search log, and ways one can use that information.

1. Search number. There are two different theories on how a search should be numbered. One theory is that a search should be a single intellectual query. It does not matter how many databases are accessed to answer the question. (The analogy would be that in traditional reference services you do not count each printed tool consulted as a separate search.) Another method is to count access to each database as a search. This is a better representation of the amount of work that is done, but it can be very misleading. Some institutions count access to different online segments of the same database as separate searches. A sensible compromise would be to keep counts of both.
2. Search requester. This could be keyed to keep track of department use and status of each individual user.
3. Purpose of search. Is this search to answer a ready reference question or to compile an extensive bibliography? Is the requester asking for information to write a short paper, to give a talk next week, or to review the literature prior to starting a dissertation?
4. Who will be charged for the search? Sometimes it is advantageous to be able to go back to your records to determine what departments have been using your services. You might need their support later! Another key point, if you are charging for your services, is to determine the level of support coming out of your users own pocketbooks.
5. Search question. A brief statement will do; it is helpful to code for broad subject categories.
6. Databases accessed.
7. Vendors used.

8. Prints—both online and offline.
9. Sign on and sign off time for each database.
10. Calendar day.
11. Elapsed time.
12. Estimated charges.
13. The name of a search analyst.

When the search comes back, one will want to add the actual charge and after the search analyst and requester have looked over the search, one may want to add evaluative comments.

Online search statistics, like statistics generated from traditional library services, provide hard data for library managers to use in evaluating their services and budgeting for the future. The difference between traditional reference services and online services lies in the fact, unfortunately sometimes, that a bill is generated each time we go online. For internal accounting procedures, we need to justify our invoice to actual services provided. This produces accurate record keeping because we are forced to log in parts of the reference process that are normally not recorded. Thus we can better judge the time it takes to handle a specific question and judge the costs of providing information.

BRS provides an online accounts file that helps searchers to tailor the search service to fit their specific needs and provides managers with the most up-to-date figures on online usage. The accounts file will display usage on specific files or access to all the BRS databases. The online usage will be broken down by current month, last month and the year up to last month. One will always have access to information that will help determine the online usage and expenses.

BRS, always trying to meet the needs of its users, is now developing an online record keeping system. It will provide a way for one to store this management information—the same information that would go into a search log. Monthly reports could be generated that will indicate how one is using online services, who is using the service and for what purpose, average time and cost per search, what databases are being searched, etc. Many libraries that have access to their own mini- or microcomputer already have the means to store and manipulate this type of information. The statistics online file will give all libraries that advantage.

I have discussed ways that online searching can provide one with a wealth of management information. Much of the information provided is *quantitative* in nature. This is not a negative feature for we need objective measures to evaluate our services. However, quantitative surveys should always be used in conjunction with concern for quality as well as quantity, and with the measured judgment of the trained librarian.

REFERENCES

1. Hawkins, Donald T. "Unconventional Uses of On-line Information Retrieval Systems: On-line Bibliometric Studies." *JASIS* 28(Jan. 1977):13.

2. Caldwell, Jane L., and Ellingson, Celia S. "Online Bibliometrics." *BRS Brief Paper Series* no. 4, June 1980, p. 1.

3. Ellingson, Celia S., and Hedstrom, Lori A. "Using Online Databases as a Tool for Collection Development." *BRS Brief Paper Series* no. 15, May 1981, p. 3.

4. Caldwell, Jane L., and Ellingson, Celia S. "A Comparison of Overlap: ERIC and *Psychological Abstracts.*" *Database* 2(June 1979):66-67.

5. Ibid., p. 66.

BIBLIOGRAPHY

Broadus, Robert N. "The Applications of Citation Analyis to Library Collection Building." In *Advances in Librarianship*, vol. 7, edited by Melvin J. Voight, and Michael H. Harris, p. 299-335. New York: Academic Press, 1977.

Caldwell, Jane L., and Ellingson, Celia S. "Online Bibliometrics." *BRS Brief Paper Series*, no. 4, June 1980.

Garfield, Eugene. "Citation Analysis as a Tool in Journal Evaluation." *Science* 178(3 Nov. 1972):471-79.

Goffman, William, and Morris, Thomas G. "Bradford's Law and Library Acquisitions." *Nature* 226(6 June 1970):922-23.

Halperin, Michael. "Preparing Online Service Reports with Packaged Programs." *Online* 5(Oct. 1981):62-68.

Line, Maurice B., and Sandison, Alexander. "Practical Interpretation of Citation and Library Use Studies." *College & Research Libraries* 36(July 1975):393-96.

Narin, Francis, and Moll, Joy K. "Bibliometrics." In *Annual Review of Information Science and Technology*, vol. 12, edited by Martha Williams, pp. 35-58. White Plains, N.Y.: Knowledge Industry Publications, 1977.