

DIANNE ROTHENBERG

MIMA SPENCER

Co-Directors

of ERIC Digests Online

Associate Directors

ERIC Clearinghouse on Elementary and Early

Childhood Education

Urbana, Illinois

Full-Text Databases: Implications for Libraries

The dramatic increase in the amount and kinds of information available online in full-text databases has implications for all users of online information systems. The increase in the number of full-text databases is related to recent technological advances in condensed data storage, a sharp rise in the number of people accessing online services with microcomputers and modems, and a long-standing user demand for online document delivery.

Defining the Full-Text Database

In comparison to bibliographic databases, full-text databases are still in a formative stage. Design, content, format, and definitions are still evolving, leading to some confusion among users. Current definitions of full-text databases tend to emphasize different kinds of database characteristics. As early as 1984, for example, Carol Tenopir described the availability online of the complete texts of journals, books, legal documents, newspapers, and other materials. She noted that when these texts are made available on free-text systems (where every word of the text is searchable), they are referred to as full-text databases (Tenopir, 1984, p. 216).

DIALOG's *Database Catalog: 1987* describes full-text databases as composed of "complete text records" that "contain the entire or full narrative text of a journal article or source publication," and then goes on to classify directories and textual-numeric databases—which frequently contain minimal amounts of narrative text—as full-text databases (DIALOG Information Services, 1987, p. 58).

The 1987 *BRS Full Text Syllabus* defines a full-text database as a "file that contains the entire contents of journal articles, encyclopedias, textbooks, or other print publications. While bibliographic databases provide structured citations, full-text databases often provide many paragraphs of unformatted text" (BRS Information Technologies, 1987, p. 1). Since few, if any, full-text databases truly contain the "entire contents" of a print publication, this definition is misleading.

Clearly, new or more comprehensive definitions of full-text databases will continue to emerge as additional types of publications become available in full-text databases.

Full-Text Databases Available Online

Most widely available full-text databases contain the content of publications available as printed products with some notable exceptions. Database producers generally claim that more frequent updating of the online versions of their products creates added value for the online user of their products. A selective look at some of the kinds of full-text databases reveals a number of differences from the print products from which they are derived, and many of these differences have important implications for searchers of full-text files.

Newspapers

VU/TEXT, with approximately thirty-two regional newspapers, and DataTimes, with forty-five papers expected online and searchable by the end of 1987, are currently the most prominent vendors of full-text newspaper databases. The full-texts of approximately 150 newspapers, regional and national, are currently available online and used primarily by the business community. Most newspaper database producers are careful to describe their databases as containing the "editorial" content of a newspaper. As full-text databases cannot cost-effectively store and transmit pictures and graphics, the advertisements, graphs, and pictorial information are omitted. Syndicated columns, letters to the editor, and news of primarily local interest may or may not be included depending on the newspaper and the vendor. For example, while some vendors carry the Capitol Edition of *The Washington Post*, others carry the "electronic edition" which contains primarily national news. It is essential that searchers understand what the vendor has included online. The words *electronic edition* should be a clue that there are significant omissions from the printed newspaper edition.

Encyclopedias

According to Grolier, the full-text "electronic edition" of its *Academic*

American Encyclopedia provides added value for the online searcher—quarterly updates of the online edition compared to yearly updates of the print version. Again, no graphs, pictures, or captions are included. One problem with these omissions is that it may be difficult to understand an entry if the information normally contained in the graphics has not been repeated in the text. The ACADEMIC AMERICAN ENCYCLOPEDIA DATABASE is currently carried by at least thirteen vendors, with home and library users ranging from elementary school students to adults.

Journal Articles

Many of the most familiar full-text databases contain the narrative text of journal articles such as the HARVARD BUSINESS REVIEW database on DIALOG. Recently DIALOG listed nearly 400 journals in its “Full Text on DIALOG” list and indicated that this is only a partial listing. The growth in usage of readily available journals online indicates the willingness of the business community to pay for the convenience of using full-text online. It is likely that full-text databases will become substitutes for on-site collections of at least some journals in corporate settings. But business users are not the only ones interested in the convenience of online. Recent conversations with BRS Customer Service representatives indicate that in some instances members of the medical community also use full-text databases in lieu of medical journals available in the same building.

Directories

Directory databases are generally classified as full-text databases by BRS and DIALOG on the grounds that the user is given the complete text of each entry rather than a surrogate bibliographic record. Many directory databases are updated more frequently than the print products can be and represent true added value for the user.

“Conglomerate” Databases

For lack of a better term, this label may be attached to a small but growing number of databases that contain information from several different kinds of print sources, reflecting the trend toward diversification of database content. HAZARDLINE on BRS, for example, contains the full text of selected book chapters, journal articles, and court decisions concerned with toxic chemicals. CCML, the Comprehensive Core Medical Library on BRS, contains selected full-text information from textbooks and journal articles for physicians.

Databases with No Widely Available Print Counterpart

In the long term, the full-text databases that provide information that is difficult or impossible to find in print may prove to be more useful than

other kinds of full-text databases. For example, a new file named DIOGENES, soon to be mounted on DIALOG, will contain news stories and unpublished documents relating to U.S. regulations of pharmaceuticals and medical devices. MAID (Marketing Analysis and Information Database), available on Pergamon's INFOLINE, contains the full text of over 30,000 market research and company reports, digests of advertising expenditures, and information on company brand names. MAID's news reports contain stories from foreign trade publications, often in translation, that are difficult to find in U.S. libraries.

Databases carried by end-user oriented services such as CompuServe and The Source, which view themselves as providers of full-text information, must also be considered in this context. When the authors asked how many full-text databases were mounted on CompuServe, the immediate response was that it would be easier to determine how many were *not* full-text databases. Most librarians need to know more about vendors and databases not traditionally used by the library community. For example, CompuServe and The Source provide financial, personal, entertainment, travel, and shopping information, and both vendors also make available hard to find information on computers through bulletin boards of special interest groups that often contain databases. SPECIALNET and ED-LINE are examples of sources of information on education which have pioneered electronic communications for educators and made full-text information available online to their user groups.

Implications for Libraries

The option to retrieve the full text of a document (rather than a bibliography) as a search result will inevitably lead to increased searching of full-text databases by librarians and patrons in the next few years. In an era when library budgets are reduced, and library journals are heavily used, and too often vandalized or missing, many libraries already recognize the online "copy" of a journal or journal article as the ultimate archive for the patron who must have an item immediately.

Different allocations of funds will be possible—and necessary—as full-text database usage grows in libraries. Instead of four subscriptions to a popular journal, for example, an academic library may be able to maintain just one or two and use the online full-text database as an always available reserve copy from which articles can be downloaded and printed out.

Ready reference or online searching initiated by a librarian to answer a reference question will increase as the efficiency of retrieving from full-text databases is measured against staff time required to search indexes and sources manually (Diodato, 1986). In fact, libraries will be able to provide

better service as full-text databases containing information not available elsewhere become available to answer questions.

Increased searching of full-text databases is also likely to reduce the number of interlibrary loan requests for items which are available full text online. For many libraries, the cost of fulfilling interlibrary loan obligations should be reduced significantly; some libraries already will not provide interlibrary loan of items that can be retrieved online by the requesting library.

Another advantage of full-text databases is that their use provides access to the full text of documents for patrons in locations which do not have extensive on-site collections. School libraries, for example, will be able to have access to the latest editions of encyclopedias, scientific reference works, and directories of all kinds.

Funding for use of full text online continues to be a major problem. As Diodato (1987) points out, there may be a great deal of merit in thinking of the book budget as part of a general information services budget that includes funds for online access. Online funds are used to purchase information directly requested by patrons, while book funds often buy information based on an expectation that patrons may request it (pp. 49-50).

Unfortunately, libraries will find that the costs for online training for staff, hardware and software, and increased online connect time will more than offset any anticipated savings. Improved service is therefore a more realistic goal than reduced costs.

Another cost-related issue is the provision of access to online information for those who do not have ready access to computers or sufficient funds to use online resources—particularly full-text databases—which tend to be more expensive. The growing segments of the population whose information needs are underserved (or unserved) must be considered in the context of equity issues in income and education and the empowerment of traditionally underprivileged segments of society. Libraries have a special responsibility to make information available for everyone and to work toward making online access possible for all library users. Increased budgetary support and reallocation of the library budget, subsidized services, use of new storage technologies like CD-ROM to reduce costs, and other creative approaches are needed.

End Users as Searchers

In designing and offering databases to their clientele, producers and vendors in the past have taken for granted a certain level of user education and training, including a disposition to read manuals and database documentation, and hands-on experience. This kind of user knowledge and experience can no longer be taken for granted, and the design of the

vendor's user interface therefore becomes a challenge of greater proportions. The existence of a large group of end users searching for information from their homes or offices is bringing about changes in the user interface.

The number of end users is growing dramatically. Fully 80 percent of the new subscribers to DIALOG in 1985 were individuals rather than organizations (DIALOG Information Services, 1986, p. 1). It is likely that BRS statistics for this same period equal or surpass DIALOG's figures since BRS markets heavily to end users. Despite the training sessions, end user oriented materials, and technical support provided by vendors, many experts see a greatly expanded role for librarians in advising inexperienced online searchers.

Vendors view end users as a new and practically unlimited market for online services, while the library market is finite and becoming saturated. Most experts forecast that librarians will play a major role in advising end users on such issues as selecting vendors and databases, developing search strategies for full-text and bibliographic databases, and downloading. Ongoing training and updating of librarians' skills will clearly be necessary to satisfy end user needs.

Many factors have contributed to the dramatic increase in the number of end users searching databases. Consumer-oriented services such as CompuServe, The Source, and topically-focused vendors like DELPHI (political science) and ED-LINE (education) are demystifying online searching for thousands of end users who will consequently be more interested in using other online information services.

The growing number of students learning to use online services will also enlarge the population of end users. DIALOG's fast growing Classmate Instruction Program is helping to create the next generation of end users. Hundreds of high school and even some junior high and elementary school teachers and librarians are using the database Classmate's specially targeted training materials and lower-priced online connect time charges to teach information retrieval skills to students and faculty (A. Caputo, personal communication, March 1987). Dow Jones NEWS/RETRIEVAL also offers special rates for schools but requires that a surcharge be added to its hourly rate for full-text database usage. The concepts involved in online searching are also being taught in a number of other ways. Students are using computer programs that include searchable databases of facts in history, science, and other subjects and are learning to create their own databases.

Vendors are aware that more full-text databases—especially on previously neglected subject areas—are necessary to encourage growth of the end user group. The increase in the number of "niche databases" which concentrate on a fairly narrow topical area provide evidence that as the user base diversifies and grows, so will database coverage. The KING JAMES

BIBLE full-text database on DIALOG, for example, is one of several recently added databases of little apparent commercial value but great research value. SPORT and CINEMA are other examples. These two databases, which index primarily the scholarly literature on these popular topics, are probably of most interest to those who have an academic or career-related interest in these subjects.

The development of user-friendly software such as Sci-Mate and Pro-Search that simplifies logging on and the mechanics of searching are also important. Already in place are subsystems, often menu-driven, designed for generalist end users, such as BRS/After Dark and DIALOG's Knowledge Index.

Among those subsystems structured for users in specific professions is BRS/COLLEAGUE, intended for use by physicians and researchers. COLLEAGUE currently represents approximately one-fourth of BRS's business. Major features of BRS/COLLEAGUE are full-text databases of the medical journal literature and the full-text AIDS KNOWLEDGE BASE. Other full-text databases in a variety of subject areas are planned. BRS profiles its typical end users as people familiar with computers who have immediate information needs and who work in environments where other people are using computers.

Menus, which most vendors believe are easier for inexperienced searchers to use than command languages, moved to prime time on the standard DIALOG system with the 1987 addition of AMERICAN BANKER NEWS, a full-text database that offers menu-driven access to its contents as well as keyword and full-text searching of individual database records. The implications for special librarians who serve these new end users of databases are numerous: an increased advisory and training role, staff access decisions, and online costs as they impact on other library services.

The Future of Full-Text Databases

Deterrents to the increased use of full-text databases include the small number of databases available and the high costs of full-text retrieval. As more databases cover subject areas previously neglected, the user base will grow. The growth of subject specific "niche" databases is a good indication that the number of users for all online services is increasing. Niche database development seems to contradict earlier speculation that full-text databases in commercially important subjects would be among the first to be developed.

Another trend is the development of databases created by organizations for the purpose of making their own print products available to online users. For example, FIND/SVP originally produced market reports

which are now available online full text. Similarly, the ERIC system has taken short articles on education called "digests," which are produced by the sixteen ERIC clearinghouses, and created a database called ERIC DIGESTS ONLINE.

While the number of full-text databases is increasing, the production of full-text databases has proved to be labor-intensive and expensive. Only a few years ago, for example, Information Access Corporation (IAC), a major full-text database producer, was still keyboarding text from published journals to produce its databases. Vendors also are affected by the growth in the number of full-text databases; they continue to invest in new software features intended to make these databases easier to search. Inevitably the increased production and marketing costs are passed on to the user. Therefore, full-text database usage costs are frequently higher than those associated with bibliographic retrieval.

Over time, database production costs are likely to decrease and the number of users to increase. If these changes result in lower prices, full-text databases can be expected to fulfill the promise they hold to improve library services. More immediately, since most information providers' profit derives from print and online versions of their products, the income lost in one sector may have to be recovered in the other, sometimes resulting in rising online costs. Libraries may find that local storage options can help to meet these rising costs. For example, full-text databases are candidates for CD-ROM or alternatives such as Grolier's Storage On-Site Program, which plans to offer access to Grolier publications and other databases for a fixed fee by storing them on campus data processing facilities and networking these databases to the library (Grolier Customer Representative, personal communication, April 1987). CD-ROM in particular has the potential to make full-text databases available to end users in libraries, such as the full text of popular journal articles.

Despite revisions already being made by vendors in search software, more improvements are needed. Database vendors and front-end software designers could make full-text database usage more attractive to users by creating user friendly interfaces specifically for full-text files. Such interfaces could deal with the special problems of full-text searching with the goal of minimizing false drops by combining a keyword-in-context approach with a full-text search of user-specified terms.

Conclusion

Full-text databases are changing the way libraries "package" and store information for their patrons. The implications for libraries range from staffing and equipment issues, to reallocating budget funds, to finding ways of providing equal access to information for all patrons. In a recent

television advertising campaign, *The Wall Street Journal* claimed that "information isn't power; knowing where to find it is." The answer to "where" is more likely to be "full text online."

REFERENCES

- BRS Information Technologies, Inc. (1987). *1987 BRS full text syllabus*. Latham, NY: BRS Information Technologies.
- DIALOG Information Services, Inc. (1987). *DIALOG database catalog: 1987*. Palo Alto, CA: DIALOG.
- DIALOG Information Services, Inc. (1986). From the director. *Chronolog*, 14(1), 1.
- Diodato, V. (1986). Eliminating fees for online search services in a university library. *Online*, 10(6), 44-50.
- Tenopir, C. (1984). Full-text databases. In *Annual Review of Information Science and Technology*, New York: Knowledge Industry Publications, Inc. pp. 215-246.