

# Full-Text & Bibliographic Databases

By Carol Tenopir

A NEW ONLINE service called *The Reference Service* was announced by Mead Data Central, Inc. in February 1983. Unlike Mead's two popular full-text services (*LEXIS* and *NEXIS*), *REFSRV* provides access to bibliographic and handbook/directory databases.

Before they added this new service, MDC's first well-known venture into bibliographic databases was the New York Times *InfoBank*, which was added to *NEXIS* in 1984. It was a surprising and controversial addition at the time because MDC spokespeople often emphasized the superiority of full-text databases and MDC seemed dedicated exclusively to full-text. At first some librarians complained that MDC salespeople seemed to know little about searching *InfoBank*. Some felt that the library market was treated as an unwanted stepchild.

Expansion in 1985 into a service that emphasizes bibliographic databases is perhaps even more surprising. *The Reference Service (REFSRV)* marks MDC's official recognition of the utility of bibliographic databases in addition to full-text and provides access to more research-type information via MDC's services. *REFSRV* puts MDC into more direct competition with online services such as *DIALOG* and *BRS*.

*REFSRV* is a separate grouping of databases, but subscribers to *LEXIS* or *NEXIS* get access to it. When a searcher chooses *REFSRV* he also specifies which of the databases he wishes to search. Optionally, most bibliographic databases can be searched together as a group or most handbook/directory databases (called "structured files") can be searched as a group.

## Databases on *REFSRV*

As of the end of 1985, *REFSRV* will have databases from a variety of sources, notably all ten Information Access Company (IAC) databases and many

government databases. (Many of the IAC databases had been exclusive to *DIALOG* before 1985.) *REFSRV* databases will include: *Magazine Index*, *National Newspaper Index*, *Trade and Industry Index*, *Legal Resource Index*, *Newsearch*, *Management Contents*, *the Computer Database*, *Industry Data Sources*, *ABI/INFORM*, *BIOSIS Previews*, *BioBusiness*, *Hazardline*, *MSDS*, *Kaleidoscope: Current World Data*, *Aerospace Database*, *NTIS*, *Federal Research in Progress*, and the U.S. Department of Energy database.

Other *REFSRV* databases include *Billcast* (a database predicting the passage of public bills introduced in the current session of the U.S. House and Senate) and the *InfoBank Abstracts* database (*INABS*) and *InfoBank Advertising and Marketing Intelligence* database (*INAMI*). The two *InfoBank* databases are also available on *NEXIS*, along with the full text of the *New York Times*.

*REFSRV* is intended to be primarily a bibliographic and handbook service for research and reference use. Two full-text databases from IAC (*Magazine ASAP* and *Trade and Industry ASAP*) will become part of *NEXIS* rather than *REFSRV*. MDC spokesperson Sharon Peake explains that some articles in the *ASAP* databases are already available in full text on *NEXIS* under separate arrangements with the publishers. *Magazine ASAP* includes more popular magazines than *NEXIS*, however, so it expands MDC's full-text coverage rather than duplicating it. To further reduce duplication, *PR Newswire* stories will not be in MDC's version of the *ASAP* files because *PR Newswire* is already on *NEXIS*.

## *REFSRV* software features

The same software used to access *LEXIS* and *NEXIS* is used for *REFSRV*. This provides simple versions for truncation, Boolean logic, and word proximity (e.g., Word A within ten words of Word B). Controlled vocabulary searches can be done in those *REFSRV* databases that have a controlled vocabulary and all databases

can be searched free text. Because *LEXIS* and *NEXIS* were designed for full-text searching by end-user searchers, the software has some unique features that are not available on most of the systems designed initially for bibliographic databases.

Search words are highlighted in context. In addition, a searcher can display just that portion of a text, citation, or abstract that contains search words. The system automatically searches for both singular and plural versions of words, for words that may have British or American spellings, and for spelling variations in Chinese Romanization schemes. If the acronym for a government agency is input (e.g., OSHA) the system will automatically search for the spelled-out agency name as well.

Not all *NEXIS* search features are available on *REFSRV*. As of last May publications in *NEXIS* may be searched together in preset format categories, subject categories, or user-specified combinations. For example, searchers can elect to search only magazines, newspapers, newsletters, or wire services; search only those publications placed in the subject groups of business, finance, government, trade and technology, or news; or establish a profile of selected publication titles. Such flexibility with database groupings is not currently available on *REFSRV*.

Some of MDC's unique software features can pose problems for bibliographic searching. The automatic plural/singular feature, for example, can cause false drops. MDC's Boolean logic feature has some limitations that may bother experienced bibliographic database searchers. If more than one term is entered in a search statement, only one set is created for the entire statement. This makes modification more difficult. Other systems such as *BRS* have this limitation, but it is especially disturbing because MDC charges for each modification. Search strategy development is hampered because search words cannot be viewed online prior to searching (as with the *DIALOG* Expand command or *BRS* Root command).



Carol Tenopir is Assistant Professor at the Graduate School of Library Studies, University of Hawaii, Manoa



## Reasons for REFSRV

Why has MDC started a bibliographic service after emphasizing the superiority of full text for so long? They realize that it is not possible (or even desirable) to convert all literature into electronic form. Bibliographic databases give a "breadth of coverage" not available on NEXIS. MDC says that "abstracts provide at least a clue" to more information than will ever be available on NEXIS. Users "can go to the library or another document delivery service" for the full texts.

One of the primary target audiences of REFSRV is a segment of the current NEXIS customers rather than new users: the corporate information specialist/librarian who has been using NEXIS for full-text searches and DIALOG, BRS, or Orbit for bibliographic searches. MDC sees REFSRV as "an alternative so the information center doesn't have to leave our service" for the more extensive literature searches that are possible with the bibliographic databases and fact finding possible with handbooks.

According to MDC, with both NEXIS and REFSRV they can be a "one-stop shop" for online searching in the corporate environment. They say that end users now using LEXIS or NEXIS will probably also be attracted to REFSRV, such as lawyers, occupational health professionals, and journalists. Academic librarians are a secondary audience. Very few public libraries subscribe to NEXIS and MDC's addition of REFSRV is not expected to change this.

The cost of MDC's services is one reason why public or academic libraries have not been major customers. There is a \$50-per-month fee to access NEXIS and REFSRV. In addition, REFSRV costs \$20-per-hour connect time, \$8-per-hour telecommunication fee via Telenet or MDCnet, \$9 per search on a single database, and \$3 for each search modification. (Search costs are cut in half between 7:30 P.M. and 7:30 A.M. local time.) There is no charge for online prints. NEXIS has similar pricing: single, \$7; group, \$15; all, \$21. A dedicated terminal leased from MDC is no longer required. LEXIS, NEXIS, and REFSRV can now be accessed via microcomputers or the MDC terminal.

## Full text on BRS and DIALOG

The Reference Service represents a new direction for MDC Data Central. It puts them into more direct competition with online services such as DIALOG and BRS, although as yet REFSRV does not have nearly the number of bibliographic databases as the others. It assumes that the software features designed for easy full-text searching will also make bibliographic searching easier.

As MDC is adding bibliographic databases to their once exclusively full-text service, DIALOG and BRS are continuing to add full text to their once primarily bibliographic services. Both DIALOG and BRS have the full texts of many directories and reference works such as *Marquis Who's Who*, *Peterson's Guide to Colleges and Universities*, and *The Academic American Encyclopedia*. Both increasingly provide access to the full texts of journals.

BRS began adding full-text journals three years ago with the *Harvard Business Review*, followed by the American Chemical Society journals and the biomedical reports of the International Research Communications System. Medical journals and textbooks from many publishers are in the BRS Comprehensive Core Medical Library, including the *New England Journal of Medicine*, *Lancet*, the *British Medical Journal*, *Gray's Anatomy*, *Triage Manual Bluebook Series*, and *Principles and Practices of Emergency Medicine*. DIALOG added *Harvard Business Review Online* in 1983 and *Magazine ASAP* and *Trade and Industry ASAP* soon after. Now BRS and DIALOG regularly add full-text databases.

## BRS and DIALOG software

Unlike MDC, DIALOG and BRS were designed for bibliographic searching where citations are short and controlled vocabulary is often used. Some software enhancements were made on these systems to facilitate full-text searching.

BRS especially made many enhancements when they began adding full-text databases. Searchers can now specify that words appear in the same textual paragraph in addition to the other proximity operators that allow specification of word adjacency and words in the same sentence. BRS recommends that one of these proximity operators always be used instead of the Boolean operator AND in full-text searching to avoid needless false drops.

When a search is complete on a BRS full-text database, the searcher can view an "Occurrence Table" that presents the paragraph and word position of search words in each document. It is assumed that the documents containing the most occurrences of the search words will be the most relevant. Alternatively, without using the Occurrence Table a searcher can choose to view just those paragraphs in which the search terms occur by using the PRINT HITS command. PRINT HITS is similar to the MDC feature that displays search terms in context. Both the occurrence table and print hits make full-text searching more satisfactory. To help users formulate full-text search strategy, BRS now conducts a half-day full-text seminar.

DIALOG also allows words to be searched in the same text paragraph using the (S) proximity operator. Their new NEAR operator (N) allows bidirectional adjacency searching in full-text and bibliographic databases. As of this writing no special full-text display features have yet been added to DIALOG, but they are promised.

## Bibliographic and/or full-text?

Full-text databases have opened up new possibilities in document delivery and online searching. They allow obscure facts in a document to be retrieved and provide ready access to selected publications. They do not appear to be replacing bibliographic databases, as was once feared. Even the full-text pioneer MDC now recognizes the unique role of bibliographic databases.

Studies conducted by BRS and the American Chemical Society show that full-text and bibliographic databases complement each other and are often used for different reasons or to retrieve different types of material. End users are more likely to use full-text databases for fact retrieval or current awareness, often turning to the information specialist for the more extensive literature searches possible in bibliographic sources.

Effective searching of full-text databases seems to require different search strategies than bibliographic searching. The Boolean AND may result in many false drops, so more precise delimiters such as words in the same paragraph, same sentence, or within a certain number of words are essential. Display features must also be changed to accommodate full-text searching. The BRS Occurrence Table and features that allow search words to be viewed with their surrounding text save a user from having to read an entire retrieved document.

Software should accommodate the best-search strategies for bibliographic database searching as well. Studies have shown repeatedly that a combination of controlled vocabulary and free-text searching offers the best search results in most bibliographic databases. Fields must be identifiable and separately searchable. Modifications should be easy to make to best use the interactive power of online searching.

There is no longer a complete separation between full-text systems and bibliographic systems. Addition of bibliographic databases by MDC, full-text databases by DIALOG and BRS, and handbooks or directories by all three have made these systems less distinct. Searchers will need to pay more attention to appropriate search strategy development and to special search features as they access both full-text and bibliographic databases on the same system. The systems will need to continue to provide software enhancements that allow searchers the most flexibility.



