ONLINE DATABASES

BY CAROL TENOPIR

Electronic Access to Periodicals

[This column is based on a speech presented at the American Society for Information Science 1992 Annual Meeting, October 1992, Pittsburgh.—Ed.]

FULL TEXT IS NOW the most prolific type of database. Just a few years ago, full text surpassed the tried-and-true bibliographic files in terms of sheer numbers. According to the 1992 edition of *Computer-Readable Databases*, full text now make up 44 percent of textual databases and over one-third of total databases (including numeric and textual). (Out of a total 6800 databases, 30 percent are full-text; 78 percent are textual.)

Not all of these full-text databases are periodicals; there are also encyclopedias, statutes, novels, press releases, and so on. About two-thirds (68 percent) of the full texts listed in *Directory of Online Databases* are periodicals.

Perhaps more meaningfully, there are now well over 3000 periodical titles available online (according to *Fulltext Sources Online*). They may be as stand-alone titles (Consumer Reports, Commerce Business Daily, or Wall Street Journal) or, more commonly, part of a large, multimagazine database like Magazine ASAP, Chemical Journals Online, or Comprehensive Core Medical Literature.

CD-ROM started with bibliographic databases, just as online did. Today, about 37 percent of CD-ROMs are full text or include some full text—many of which are periodicals (over 330). That number is growing and changing weekly.

Electronic full text on other distribution media such as diskette or magnetic tape for local loading are much fewer in terms of numbers right now (excluding FAX), but are increasing and will play a bigger role very soon. In the 1992 *Directory of Portable Databases*, approximately 11 percent (45 of 430) of magnetic tape titles and 28 percent (154 of 545) of diskette databases are full text. Just a few of these are periodicals.

Numbers are fine and are academically interesting, but to really understand the current state of the electronic periodical database marketplace you have to look deeper. To take that deeper look there are five questions that need answering.

Why is full text increasing?

Full-text databases are not just increasing in numbers; they are the fastest-growing type of commercial database and have been for the last few years. This is happening for several reasons.

1. Improvements in scanning and OCR—Scanners have gotten less expensive and more reliable. Optical character recognition software has also improved, yielding acceptable error rates for most kinds of texts, although some proofreading/editing is still required. Many database producers have switched the bulk of their full-text conversion operations over to scanning/OCR in the last few years, and others have gotten into full text for the first time because of affordable scanners.

2. Increases in storage capacity—CD-ROM represents one kind of increase in storage capacity for micro-based databases; also, the storage capacity of fixed disk drives for online systems and locally mounted databases has steadily increased while the cost has steadily gone down. This makes large full-text databases and more full-text databases feasible. That is the hardware side, but there is also a software side to this—better compression/decompression software means much larger databases can be stored in the same amount of space.

3. Faster speeds—The increase in speed has two elements: 1) faster computer processing speeds that make search and retrieval of lengthy texts reasonable on CD-ROM and locally mounted databases, and 2) faster telecommunications speeds for remote online. The upper limit for most commercial online systems is still just 9600bps, but that is much faster than just a few years ago and is steadily going up. Access over Internet connections may be faster for some already.

4. Demand—Profits drive most of the periodical database market; even nonprofit segments are driven by priorities as set by their customers or constituents. Librarians, researchers, and users are saying loud and clear, *"We want more full texts."*

What periodicals are available?

The five categories of periodicals covered in *Fulltext Sources Online* are:

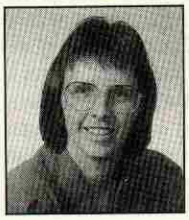
1. Scientific/technical journals
2. Popular magazines
3. Newspapers
4. Newsletters and
5. Newswires

All of these are widely available on the major online systems such as DIALOG, MEAD Data Central, BRS, STN, etc. The top three categories also are on CD-ROM and increasingly on magnetic tape. A few single titles of magazines or newsletters are available on diskette (such as Commerce Business Daily and some specialized newsletters such as Lloyd's Maritime database, which tracks ship movements on diskette or magnetic tape.)

What formats are available?

What form or formats are available for electronic periodicals? This is actually a broader question of what do the databases include and what searching powers will they support? Right now there are three main options for form and format: full texts can be either ASCII text only, scanned images only, or a combination of ASCII text and scanned images.

ASCII texts predominate in the online world and the diskette or locally mounted tape world. There also are many ASCII-only products on CD-ROM, such as EBSCO's Magazine Articles Summaries with Full-Text Elite and Ziff-Davis's ComputerSelect. The advantages of ASCII text are well known to online searchers—they take up much less storage space and transmit faster, require no special display hardware, can be searched by every word, can be downloaded (documents



Carol Tenopir is Associate Professor at the School of Library and Information Studies, University of Hawaii at Manoa, Honolulu

or portions of documents) and transferred to word processing or database programs, etc. But they are missing graphics, usually exclude charts or tables, and ASCII text just doesn't look like periodicals we are all used to. The page numbers are different, the type fonts and layouts are weird—the result is unaesthetic and incomplete.

Scanned images, on the other hand, adhere to the user's mental image of a magazine. They look just like printed versions (and in most cases today are just digital "photographs" of print-based titles). They are popular with many users. UMI's trendsetting products, such as Business Periodicals Ondisc/ABI-Inform, can be used in conjunction with a searchable bibliographic database. As such, the scanned periodical database is a document delivery tool. The negatives are that they require high-resolution monitors and laser printers and one loses the searchability and manipulatability of full text. (Some say they are just one step up from microfilm.)

A combination of ASCII text for full-text searching, downloading, or altering with scanned images for display is, from the user's point of view, probably the best of all possible worlds. It is also the most demanding of hardware and software. CD-ROM encyclopedias such as Grolier's and Compton's Multimedia Encyclopedia are the leaders in combination products. Look for many more, including periodicals, soon. The American Chemical Society and OCLC's CORE project are combining formats and media.

Who are the big players?

There is much disagreement in the industry over whether scanned-image or ASCII files are the best, especially in the CD-ROM arena, and proponents on both sides are claiming their solution is the only one. That brings us to the people involved—the companies that consumers (purchasers) will need to deal with to lease or buy electronic periodicals.

In the online world many of the names are familiar—they've been around in most cases for a decade or two. *Fulltext Sources Online* tracks 19 companies in the United States, Canada, and Europe, including such major players as DIALOG, Mead Data Central, STN, etc. These organizations sell access to periodicals that almost always have a print equivalent. The advantage of electronic access to titles with print equivalents is that the printed form is almost always indexed in one or more indexing/abstracting publications.

Publishers and libraries are beginning to go beyond print equivalents. A big early player in electronic-only periodicals was OCLC, which, along with the American Association for the Advancement of Science (AAAS), developed the award-winning online Current Clinical Trials. Current Clinical Trials is a refereed, text and graphics scholarly journal that is available only in electronic form.

In addition, dozens, perhaps hundreds, of electronic-only newsletters and journals are being created by researchers and librarians for distribution over Internet. Internet has grown quickly as an important conduit for electronic-only journals and not-for-profit textual databases. For the many titles that are now on Internet, the key player is the "little guy."

In the CD-ROM world, UMI is the king of scanned images, just as it is for microfilm; other producers or vendors heavily involved in ASCII, scanned, or a combination include Information Access Company (IAC) and Ziff-Davis; DIALOG; EBSCO; Elsevier Science Publishing; Elsevier & Blackwell; and many other scientific publishers with the ADONIS project.

For-profit producers such as UMI/Data Courier and IAC are a bit involved. But the most interesting developments are happening in the government and libraries sector with involvement of organizations such as NLM (Chemline), OCLC (CORE with the American Chemical Society), and the American Chemical Society.

What impacts are there on libraries?

Although widespread use of electronic periodicals is just beginning, libraries are already feeling some impact. First, libraries are paying subscription costs twice. Remember, almost all of the electronic periodicals today have printed versions. This will not always be the case, but even for those for which it remains true, it doesn't make sense to buy something twice. Most libraries have kept their print subscriptions along with electronic access. By next year, with the continuing rise in subscription costs and longer time with full-text access under their belts, more libraries will begin to switch over and cancel print. (Even if it is not *exactly* the same, by the way.) Second is the "pay per drink" model vs. the standard pay-per-subscription plus photocopier. Electronic versions, even some on CD-ROM, are often sold on the pay-per-drink model where every user who accesses an article incurs a per document or per page cost. In the past, of course, the photocopy machine got that per article use cost. Nothing further went back to the publisher except perhaps copyright clearance fees through interlibrary loan (ILL) departments.

Libraries are responding differently to the pay-per-use charge. Some pass back the direct costs to the user, some absorb all fees and take it out of the materials budget, some differentiate by medium (they absorb CD-ROM costs but pass back online). Whatever method you use, electronic journals with a pay-per-drink rate can get very expensive, especially in large libraries, because the copyright fee attached to each copy may run as high as $18 (for more on this whole issue, see Mounir Khalil's "Document Delivery: A Better Option?" *LJ*, February 1, p. 43–47).

Third is the impact of hardware requirements for full-text periodicals. I've already mentioned that image files require high-quality monitors and printers, both image and ASCII full text take lots of storage space when loading inhouse, and huge full-text files may adversely impact processing and response time. If you get into inhouse loading in a big way you will need hardware upgrades. One database producer told me he gets many inquiries about leasing full text for loading locally, but most libraries back out when they discover the hardware overhead requirements.

Lastly, libraries will be affected by the demand for more and better products. When users get used to one full-text product—be it CD-ROM or online, be it image files or ASCII—they want more. And they want the texts accessible in their offices or dorm rooms (or wherever your OPAC or bibliographic files are available). Furthermore, most people cannot or do not want to pay.

Electronic full-text periodicals are already a big deal. In the online and CD-ROM area especially they are available, accepted, and used. In the near future, the trends over the last decade will continue, and we can safely predict more—more titles available, more acceptance in libraries, and more use (also, perhaps, more money required).