Newspapers On Disc: A Survey And Critique Of Metropolitan Daily Newspapers on CD-ROM

by Mark Stover

Print newspapers have played a major part in the information dissemination process almost since their inception 500 years ago. Metropolitan daily newspapers provide current information on a wide variety of topics and can be purchased at a price that almost anyone can afford. National and international news, local events, regional coverage, sports (at all levels), humor, food, business, travel, obituaries, opinion and the arts all receive varying degrees of attention in a newspaper. In short, newspapers provide their readers with a vast array of information resources, the breadth of which cannot be matched by any other medium.

The latest trend in CD-ROM publishing is full-text access and newspapers have begun to show up in this format. This article will survey the various newspapers now available on CD-ROM and will discuss the positive and negative implications of this new phenomenon.

Electronic Newspapers

Online versions of newspapers first became available in the mid-1970s. Today, there are well over 100 electronic newspapers that can be searched full-text. Most newspaper libraries have an electronic version of their parent newspaper available in-house, but more and more metropolitan newspapers are becoming accessible online through companies like Dialog, Nexis and Vu-Text. The latest trend in research libraries is locally mounted commercial databases (such as the Chronicle of Higher Education project at the University of Southern California) and this may soon extend to mounting a machine-readable version of the local metropolitan daily on the campus networks of the larger research universities.

Online newspapers are a great boon to librarians and other researchers. They provide immediate access to the full texts of newspapers, they are generally well-indexed and they are current (usually these databases are updated on a daily basis). Two important drawbacks to these resources are their high costs and the necessity of using intermediaries to perform searches. A word should be said here about indexing and abstracting services for newspapers. These tools, available in print, microfilm, CD-ROM and online formats, are often efficient and cost-effective ways of searching for information found in newspapers. However, abstracts and indexes are not sufficient if instant access to the full text of the newspaper is required.

The year 1991 saw the sudden appearance of at least four different companies publishing full-text newspapers on CD-ROM. Although the trend is clearly toward major metropolitan daily newspapers, CD-ROM versions of national newspapers like the Wall Street Journal and the Christian Science Monitor have also begun to proliferate. Smaller, regional newspapers will probably also be affected, as evidenced by the appearance of a CD-ROM edition of the Northern Echo, a regional morning newspaper in the United Kingdom. If this trend continues, CD-ROM will soon become an established medium for newspaper publishing.

Database Evaluations

Dialog’s Los Angeles Times

Dialog Information Services offers the full text of the Los Angeles Times on CD-ROM, as well as other newspapers including the San Jose Mercury News and the Boston Globe. The installation of the retrieval software (DIALOG onDisc Manager) was quite simple and the documentation for the software was complete and accurate. Since the newspapers available from Dialog on CD-ROM are also available online, users with modems and Dialog accounts are given the option of dial-access to these databases via DIALOGLINK Communications Manager software.

DIALOG onDisc allows users to access the Los Angeles Times through an easy-to-use menu system as well as through the traditional DIALOG command line that most online searches are familiar with. The menu-driven option and the command-driven option are actually interactive with each other, so that the user can utilize some of the menu options while working within the command environment.

The Los Angeles Times onDisc indexes several different fields within each record, including subject headings, publication date, edition, section heading, headline, byline, dateline, lead paragraph, length of record, part number, page number, caption, article type and graphic type. All information from the print version is replicated in the CD-ROM version except classified ads and other advertisements.

Because of the magnitude of information, only six months of text is included on each disc (as compared to one year of text per disc on all of the other products reviewed here).

The search engine utilized for the Los Angeles Times onDisc is very effective because it can be used by everyone from the neophyte searcher to the expert searcher. The menu option has a rather short learning curve, while the command option allows for powerful and sophisticated searches (including Boolean, nested, proximity and truncated searches).

There is no tutorial available, but the help screens are somewhat useful. Response time is very fast and...
downloading of articles is available.
The Los Angeles Times onDisc from Dialog gives users a popular and critically acclaimed newspaper from the second largest city in North America, accessed by powerful software produced by one of the industry leaders in electronic databases.
Anyone interested in news from California, Pacific Rim economics and practically anything else should consider subscribing to this product.

UMI's The Washington Post

UMI offers The Washington Post onDisc, as well as other CD-ROM newspapers like the Chicago Tribune, the San Francisco Chronicle and the Atlanta Constitution/Journal. The installation process was fairly easy and the documentation was complete.

UMI's retrieval software, ProQuest, is well-known by users of Dissertation Abstracts onDisc or any of the other UMI CD-ROM databases. ProQuest is a menu-based program, although Boolean and proximity operators must be keyed in separately. The software is fairly easy to use, although its more sophisticated techniques are more difficult to learn. ProQuest allows truncation, but it also provides automatic truncation for some keywords (e.g., searching for "highway" will also find "highways," and vice versa).

There are excellent, context-sensitive help screens, as well as a useful tutorial. My only complaint with ProQuest is that its opening batch file sequence takes much too long (about 30 seconds).

The Washington Post onDisc is indexed by a number of different access points, including text, headline, number, byline, source, date, section, abstract, subject, name, company, type, length and illustration type. It covers all articles from this newspaper except wire service stories and stock pages.

Searches can be qualified by field name and search sets can be reused (e.g., "search #1 AND search #2"). Relevant records can be marked and later printed out or saved to disc. Response time is usually adequate, but not as fast as the other systems reviewed here. Printing and downloading of records is easily accomplished.

UMI's Washington Post onDisc combines one of the nation's most respected daily newspapers with a proven search-and-retrieval system. It is an excellent choice especially for those who are interested in political and legislative news.

Chadwyck-Healey's The Times and The Sunday Times

London's premier daily newspaper, the Times and its sister paper, the Sunday Times, are now available through Chadwyck-Healey (North American distributor) on CD-ROM. They have also announced a CD-ROM version of another UK newspaper, the Guardian, to be released in the near future. Installation of this system was extremely simple, since the software runs off the CD-ROM disc and is never copied onto the computer's hard disk. The user guide is short (24 pages) but complete.

The retrieval software, developed by the Open University, is completely menu-driven. It is easy to use yet fairly powerful, with truncation, Boolean and qualification features. The help screens are useful, but there is no tutorial. Response time is relatively good and downloading of articles is possible.

All information originally appearing in the Times and the Sunday Times is reproduced here, including domestic news, foreign news, business, sports, letters to the editor, obituaries, law reports and feature articles. Only information restricted for legal or copyright reasons is omitted. Indexed fields include text, date, section, headline and byline.

NEWSPAPERS ON CD-ROM

<table>
<thead>
<tr>
<th>PRODUCT:</th>
<th>Price:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialog's The Los Angeles Times</td>
<td>$2250 for one time cost of current year plus one year backfile; $5985 for current year plus four years backfile; $1000 to renew each year</td>
</tr>
<tr>
<td>Updates: Quarterly</td>
<td></td>
</tr>
<tr>
<td>Producer: Dialog Information Services 3460 Hillview Avenue Palo Alto, CA 94304 800/334-2564 415/858-3785</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PRODUCT:</th>
<th>Price:</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMI's The Washington Post</td>
<td>$750 per year for subscribers who also receive The Washington Post on microfilm from UMI; $2350 per year for non-subscribers to microfilm; $950 per year for 1985-1989 backfile discs</td>
</tr>
<tr>
<td>Updates: Monthly</td>
<td></td>
</tr>
<tr>
<td>Producer: UMI 300 North Zeeb Road Ann Arbor, MI 48106 313/761-4700 800/521-0600</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PRODUCT:</th>
<th>Price:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chadwyck-Healey's The Times and The Sunday Times</td>
<td>$1775 per year</td>
</tr>
<tr>
<td>Updates: Quarterly</td>
<td></td>
</tr>
<tr>
<td>Producer: Chadwyck-Healey 1101 King Street Alexandria, VA 22314 703/683-4690 800/752-0515</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PRODUCT:</th>
<th>Price:</th>
</tr>
</thead>
<tbody>
<tr>
<td>NewsBank's The Boston Globe</td>
<td>Between $1395 per year and $1750 per year, depending on type of library and frequency of updates; all backfile discs are $1095 per year</td>
</tr>
<tr>
<td>Updates: Monthly</td>
<td></td>
</tr>
<tr>
<td>Producer: NewsBank 58 Pine Street New Canaan, CT 06840 800/762-8182 802/675-2397</td>
<td></td>
</tr>
</tbody>
</table>
The Times and the Sunday Times on CD-ROM is a fascinating product for those of us on the other side of the Atlantic. It will be especially useful for those interested in news from the United Kingdom and economic developments in Europe.

**NewsBank’s The Boston Globe**

NewsBank has announced a new product called CD News, which places the full text of major city newspapers, including the *San Francisco Chronicle* and the *Sacramento Bee*, on CD-ROM. For this review, I examined the *Boston Globe*.

Although I experienced some problems with the installation process, it was not terribly difficult or cumbersome. Presently, only a preliminary manual is available, but NewsBank promises that in the near future more complete documentation will be sent to subscribers.

The retrieval software provides for Boolean searching, truncation and one proximity operator (users can search for two terms within the same sentence). The software is menu-driven, although the Boolean and proximity operators can be keyed in directly. The system is fairly easy to learn and use, although the help screens could be improved. The response time is good, but unfortunately downloading is not available.

Fields that are indexed in the Boston Globe include text, headline, lead paragraph, keywords, date, section and author/source.

I encountered several obstacles in searching NewsBank’s *Boston Globe*. One problem was that search terms were not always highlighted in the text, especially after Boolean searches. A second puzzling feature of this database was the fact that a few records only included a headline and the text of the article, omitting the author, the date and the section of the paper.

A final problem that I experienced was a quirk in searching the index. Pressing the F5 key pulls up a window on the right side of the screen with a list of every word indexed in the database, allowing the user to browse through the index and choose selected terms. However, before one can type in a word, one must first backspace one character, which is somewhat distracting. NewsBank has assured me that all of these problems are bugs in the software and will be fixed in the next version released. Despite the existence of these problems, the *Boston Globe* on CD-ROM from NewsBank is a viable option for those considering CD-ROM newspapers. Potential subscribers in the New England area and fans of the Boston Celtics will be especially interested in this database.

**SUMMARY**

While no one product stands head and shoulders above the rest in this review, my personal favorite is Dialog’s Los Angeles Times. Its

---

**Newspaper CD-ROMs Tell The Full Story**

by Mick O’Leary

Newspapers are the latest publication type to appear on CD-ROM. Papers that are currently or prospectively available in full text include leading national titles like the *New York Times*, *Wall Street Journal*, *Christian Science Monitor* and *USA Today* as well as major urban dailies like the *Washington Post*, *Los Angeles Times*, *Chicago Tribune*, *New York Newsday*, *San Francisco Chronicle*, *Philadelphia Inquirer* and the *London Times*. CD-ROM is the fourth medium for these newspapers, which are also available in print, microform and online.

Having these important publications on CD-ROM is welcome, but does this new distribution medium mean yet another set of complications about database content? Will there be another set of editorial policies about what goes onto the CD-ROM and what does not? Happily, the answer is generally “No.” CD-ROM newspaper users will find that their learning curve is mercifully short. Editorial policies from one publisher to another are surprisingly consistent. They are also very inclusive, with virtually all of the text from the paper appearing on the CD-ROM.

NOT A FACSIMILE

The key point in comparing newspapers on CD-ROM with the actual print paper is that “full-text” does not equal “facsimile.” “Full-text” in fact means “full ASCII text,” or the text of the newspaper’s articles. This excludes two important parts of the original printed paper: images and certain kinds of non-article text.

Photographs and other graphics, such as line drawings, illustrations, maps and tables, which are not convertible to ASCII are, of course, omitted. Textual material that is both ephemeral and not related to a story, including advertising and tabular material like stock quotes, sports box scores and calendars, is also excluded.

These guidelines are exactly those which apply the online versions of these papers. The CD-ROMs are produced from the online newspaper database with no changes. Online newspaper searches will be happy to know that they are dealing with exactly the same content when they move to the CD-ROM version.

**VIRTUAL FULL-TEXT**

In general, a newspaper on CD-ROM includes every news and feature story from every section of the paper. The few variations from this principle concern only the smallest and most insignificant stories. “Every” story includes not only so-called “hard” news from the front section, but also everything from sports, entertainment, lifestyle and other sections. In the case of the *New York Times*, for example, this includes the four regional sections for New Jersey, Long Island, Westchester and Connecticut, according to Jim Patterson, Director of Index Operations. The *Wall Street Journal* CD-ROM, explains Pat Rodewald, Associate Director of Business Development, includes all unique stories from the four regional editions of the *Journal*.

When there are several similar versions of the same story,
flexibility, ease of use and power make it appropriate for experienced searcher and inexperienced end-user alike.

**Benefits Of Newspapers On CD-ROM**

One of the most important benefits of newspapers on CD-ROM is access to the complete text of the primary sources. This feature sets it apart from most bibliographic databases which generally require a "two-step lookup" process to reach the indexed document. It can also relieve the frustration at not having immediate access to the original source.

Although it might seem obvious, it should be noted that another great advantage of CD-ROM is the ability to work within the electronic environment. Free text and Boolean searching, date qualification and a host of other features are available on CD-ROM and generally make the search much faster and more productive.

Newspapers on CD-ROM provide users with relatively current information as they usually publish information as soon as it is received. Thus, CD-ROM newspapers will generally be at least one year more current than most other databases on CD-ROM.

Second, CD-ROM newspapers give users electronic access to local and statewide news, information that would almost never be found in traditional databases. This type of information is extremely valuable for students and researchers especially in the social and behavioral sciences and can also be useful in legal and technical research.

Third, newspapers on CD-ROM furnish a wider view of societal issues than is generally found in academic-oriented or "scholarly" databases. For example, the opinion section of a daily metropolitan newspaper contains "person-on-the-street" viewpoints in letters-to-the-editor and in opinion/editorial pieces. Very few other kinds of databases could provide this sort of information, which conceivably could be quite valuable to researchers.

Fourth, CD-ROM newspapers give users a unique type of information in the form of "news." News, especially written from a journalistic slant and dealt with on a daily, "as-we-see-it" perspective, can be highly useful to historians, social scientists and even a college student writing a freshman term paper. Newspaper articles are often considered primary source materials by researchers because they usually provide eyewitness accounts and direct quotations from active participants.

Reading a newspaper article, whether in print, microform, or electronic format, is usually the "next best thing to being there."

**Drawbacks Of Newspapers On CD-ROM**

A major disappointment with the products reviewed in this article is that none of them contains the full images of some papers may violate the letter of this law without harming the principle. Richard Geiger, Library Director of the San Francisco Chronicle, explains that redundant earlier versions of a story are omitted in favor of the latest one. At New York Newsday, describes Director of Libraries Mary Ann Skinner, successive versions of a story are compared line-by-line. Unique information from an earlier version is added to the final, saved version.

There are two categories of story which some papers consistently omit: Wire stories and small "fillers." Most newspapers do include all stories which they obtain from news wires like AP and UPI, though there are exceptions to this pattern. Pat Rodeawald of the Wall Street Journal says that the online and CD-ROM versions of the Journal omit stories from Reuters, but include those from all other wire services.

Fillers, or very short stories of minor interest, are often uniformly saved. The New York Times, Wall Street Journal and New York Newsday all keep even the most insignificant one or two line stories. When fillers do not make it onto the CD, it is usually no great loss. The San Francisco Chronicle's Geiger points out that only fillers of non-local interest are excluded, thereby preserving the rarer and more valuable local fillers.

**BEETTER THAN PRINT**

There are two cases where the CD-ROM newspaper will be an enhancement over its print predecessor. The Wall Street Journal CD-ROM is based on the text which appears online in Dow Jones News/Retrieval, where it is defined as "...all articles scheduled for publication in domestic editions..." News/Retrieval Associate Editor Diane Thieke explains that some of these stories may be omitted from the print paper because of space limitations. Thus, the CD-ROM and online Journal will have stories that the newspaper itself does not, though Thieke adds that these articles are of lesser importance. Publishers generally prepare enhanced corrections for their online and CD-ROM databases. When a correction is made, it is linked with the original story. Online and CD-ROM searches, therefore, will automatically retrieve the corrected version of a story, which is impossible with print and microform research.

**FACSIMILES ON THE HORIZON**

The guiding principle when moving into full-text CD-ROMs is inclusion, with the aim of making them as complete as possible, not only for researchers, but for the reporters and editors who also use them. Fastidious researchers may want to investigate the editorial policies of individual newspapers, if it is important to know specifically what exclusions or modifications are made on the CD-ROM. In time, even these small variations may disappear.

Newspaper publishers are anticipating optical storage of true facsimiles of their papers. The stories would be linked to a searchable ASCII database of the text, thereby, combining the best of both worlds. This hybrid technology is now available, but is too costly for commercial application. In the next four to five years, however, "full-text" and "facsimile" may be synonymous.

*Communications to the author may be addressed to Mick O'Leary, 10230 Harmony Road, Myersville, MD 21773: 301/293-2799.*
newspapers on disc. Full imaging of an indexed document is expensive and space-intensive and it may be several years before newspapers receive the same treatment as UM's Business Periodicals OnDisc full imaging product. Although CD-ROM metropolitan newspapers are more current than most other databases, they are usually at least one month (and sometimes up to three months) out of date. While this lack of currency will usually not be a problem for most users, it stands in stark contrast to the daily updates that most online newspapers offer. As with many other electronic information products, most users will decide to use the CD-ROM version of a newspaper as the mainstay and the online version for the most recent information.

Another drawback of CD-ROM newspapers is the relatively large up-front costs that are involved. Users of these products will need to spend money on a personal computer (IBM or compatible), a printer, a CD-ROM drive and a subscription to the newspaper (usually between $1000 and $2000 per year).

Rhetorical Questions

The realized concept of newspapers on CD-ROM makes us ask several questions vis-a-vis the whole idea of electronic serial publications. One question concerns the problem of ownership versus leasing. Are libraries and other subscribers of these products the owners or simply the temporary users of this information? Subscribers to print or microform newspapers own that information, but most companies that offer CD-ROM newspaper subscriptions require a lease agreement.

Subscribers are spending a great deal of money on information that may turn out to be ephemeral. This does not seem very fair and it probably is not prudent for libraries to spend money on valuable information which they will not be able to keep should their lease be terminated. However, many producers of CD-ROM information are rethinking their policies in this area and the situation may change in the near future.

A related question is that of the conservation and archiving of older electronic information. One of the tragedies of the print newspaper medium is its extreme fragility, but one of the beauties of microfilm is its durability and its ability to preserve information that was originally produced in a fragile medium.

Electronic newspapers on CD-ROM must deal with analogous problems. The question of how long an optical disc will last is almost a moot point since electronic information can be easily manipulated, refreshed and/or transferred to another medium. A more important question is that of ownership; it is almost irrelevant to ask about preservation of electronic information if you will not be able to keep it in the end.

Libraries that subscribe to CD-ROM newspapers may want to consider the interlibrary loan ramifications. Since more and more libraries are stressing interlibrary cooperation as a way of fighting local budget cuts, it seems reasonable to assume that even electronic information should be made available outside the institution.

How would libraries share electronic information? Would a hard copy be printed and then mailed (or faxed), or would the information remain in electronic form throughout the entire process, perhaps traveling from one library to another through an electronic network? Would copyright restrictions limit the sharing of electronic information? These questions need to be addressed if libraries are to continue to forge cooperative agreements even in an electronic age.

How do we define or categorize a CD-ROM newspaper? Do we catalog it as a regular serial (which in fact it is) or relegate it to the netherworld of "machine-readable data files"? Is it an indexing tool or a periodical, or both? These are really theoretical and philosophical questions, but ultimately they will affect the way we think and talk about these electronic tools.

Should publishers of electronic newspapers pursue advertising as a source of revenue (which would decrease subscription costs)? If advertising was pursued, which market(s) would be targeted and how would advertisements appear on the screen? Perhaps an advertisement could be attached to each record? If advertising was not chosen as a viable option, why not?

It would certainly increase revenues for the publisher and advertisers would benefit as well. Would some users be offended by the utilization of the "pristine" CD-ROM environment for less-than-pure purposes? Since the CD-ROM pricing structure would suggest that money plays a major role in this market already, why not allow advertisers to throw their hats into the ring?

Potential Users Of Newspapers On CD-ROM

Many types of libraries could benefit from a subscription to one of the available CD-ROM electronic newspapers. Newspaper libraries can use these products as an occasional alternative to online searching, although newspaper librarians often need to search for more up-to-date information than CD-ROM newspapers are able to provide. Other special libraries, such as research think-tanks and corporate libraries, could also profit from these tools.

While individuals may never be able to afford CD-ROM newspapers, libraries and other institutions should be able to support at least some of these products. Future upgrades to newspapers on CD-ROM will hopefully include better user interfaces, compacted storage of information on discs and eventually full imaging. Online newspapers will probably continue to exist side by side with their optical counterparts, but newspapers preserved on microfilm may someday become obsolete, superceded by full image CD-ROM newspapers with better indexing, faster access time and higher quality reproducing capabilities.

Communications to the author may be addressed to Mark Stover, Library Director, California Family Study Center, 5433 Laurel Canyon Blvd., N. Hollywood, CA 91607; 818/509-5959. Internet: ENQ4MES@MVS.OAC.UCLA.EDU

104 CD-ROM PROFESSIONAL NOVEMBER 1991