TERRY NOREAULT

Database Department Manager
OCLC, Inc.
Dublin, Ohio

Optical Publishing:
Effects on Reference Services

Recent technological developments are having a major effect on the delivery of reference services. The most recent of these, optical storage technology (specifically CD-ROM), has been touted as the most important development in publishing since the printing press. While this appears to represent a level of hype that is most probably not deserved, it now appears clear that reference services are beginning to be changed in significant ways. This technology does not represent a qualitatively different service, but it will have a very dramatic effect on the economics of delivering certain types of services.

Before concentrating too heavily on optical storage, attention should be paid to an older technology which is a vital part of this revolution—i.e., the microcomputer. The advent of low cost personal computing in the last decade is an important part of the equation in allowing these new services. Presently the focus is on the newest technology, CD-ROM, but the computer is the most important element of the equation. The computer is useful in the delivery of information with the CD-ROM, but the converse is not true.

One of the immediate effects of this new publishing medium is to accelerate the trend toward end user searching. The personal workstation, without its need for telecommunication costs and with the low production cost for CD-ROM discs, is allowing libraries to offer access to electronic reference sources in much the same way that access is provided to print reference material. The term *electronic reference sources* is being used here in place of online services since CD-ROM is certainly not online.

Optical Technologies

While optical technology is having dramatic effects, it is not a pana-

cea. This discussion will explore this new technology with attention to the strengths and weaknesses of the different optical technologies. There are four basic optical technologies: CD-ROM, WORM, erasable, and video. The first three are digital technologies and the fourth is an analog technology developed for use with television.

The video disc technology has been adapted for use with digital computers, but it has few advantages compared to the digitally based technologies. It is the oldest of the four and achieved a great deal of early interest as a publishing medium but has been mostly discarded in favor of the CD-ROM.

One of the very important aspects of the CD-ROM technology is how quickly the CD audio has been adopted by the consumer market. The audio version of the CD has been the most successful consumer product ever. Its adoption has been faster than either the telephone or television. The success of the CD in the consumer market is important for two reasons—economics of scale, and availability of the hardware and mastering services. Because the consumer market is buying millions of the drives, and much of the manufacturing process is common to both, CD-ROM drive purchasers are gaining the advantage of economies of scale which would not be present in a device useful only in the computer market. In addition, the mastering facilities for CD-ROM discs cost in the tens of millions of dollars to construct, and if they had to recover the cost from data applications alone, the cost of producing the discs would be much higher. The acceptance of the CD by the consumer market also increases the life span of the technology by the shear number of people who own CD audio drives. This is a huge market which will not be easily walked away from.

The WORM (Write Once Read Many) drive allows data to be written to the drive but not erased, hence the "write once" in the name. The WORM drive appears to be most useful for archival purposes, a replacement for magnetic tape.

The erasable drive will allow data to be erased after it has been used and therefore makes the disc reusable. The first beta test versions of this technology are just coming to market.

CD-ROM

One of the prime advantages of CD-ROM is the de facto standards that exist for the hardware, software, and file structures. The Sony/Philips standard defines the characteristics for both the reader and the physical properties of the disc. There are three Sony/Philips compact disc standards, referred to as the Red, Yellow, and Green Books. The Red Book describes the CD audio system and specifies media size, layout, and mechanical characteristics as well as the organization of information on

the disc, the encoding scheme, play back signal characteristics, and so forth. The Yellow Book governs the use of the CD medium as a CD-ROM, specifies sector layout and an additional layer of ECC. The Green Book describes the use of the CD medium for CD-Interactive (CD-I) and describes the functions required on a CD-I player. None of these standards is publicly available; they are available only to licensees of the standards (essentially drive and media manufacturers), but the contents are widely known. Some of this material may be made public in order to facilitate file format standards. National standards need to be able to reference the Sony/Philips standards but cannot unless they are publicly available. This means that a disc manufactured by any mastering facility will work on any CD-ROM reader. The CD-ROM technology appeared in reference products only a short time after its introduction. In fact, the first commercially available CD-ROM product was a reference product, the *Academic American Encyclopedia* from Grolier. It was well received in the marketplace, and now there are many products available.

The other important standard is the one originally developed by the High Sierra Group and adopted by ECMA (European Computer Manufacturers Association) and NISO (National Information Standards Organization). (The High Sierra Group was formed by the major CD-ROM suppliers to produce a draft file format standard. The Fall 1985 meeting produced a draft upon which ISO, ANSI, and ECMA standards are based.) This standard specifies how the files are to be structured on the disc. Software which is developed to this standard will work with any of the CD-ROM hardware on any computer which provides a driver designed to the standard. This is important to the library community in that it increases the likelihood that one hardware configuration will be able to run software packages from different vendors.

There are two primary advantages of the CD-ROM technology—i.e., low cost and high storage capacity. At the present time, the price of a CD-ROM reader is about $700. It is likely that the price will drop in the near future, but it still will remain several hundred dollars higher than the CD audio drives. The higher cost is due to the cost of the interface card and special requirements of data applications. When looking at possible technologies which could replace CD-ROM, it is important to remember that one of the important reasons for the low cost of CD-ROM is the huge success of CD audio. This means that a replacement technology would find it difficult to compete on a cost basis without also finding a consumer niche.

The storage capacity of the CD-ROM (600 million bytes) opens up the publishing of large databases at low prices. At present most of the mastering facilities are only capable of placing 550 million bytes on the disc, but this is still a large database. This means that with complete indexing it is

possible to place 500,000 to 700,000 MARC records on a disc. Of course this depends on the size of the records, the overhead of the particular indexing package used, and whether data compression is used.

While the capacity is large, it is of fixed size. This places limits on the number of records which can be placed on the disc and forces the system designer to make packaging decisions when the database size exceeds the capacity of one disc. It is possible to build multi CD-ROM reader applications, but at some point the size of the database becomes too large to be manipulated by a microcomputer if Boolean queries are performed. It is unlikely, with today's technology, that a CD-ROM-based microcomputer system will be used to search the entire MEDLINE database since it would occupy approximately fourteen discs.

Another important advantage of CD-ROM is that the variable production costs of the discs are very low—about $5 each. It is likely that this cost will still drop a little. This low unit cost is important in something which is basically a publishing medium with a large market. While the variable costs are low, the fixed costs of production are a slight barrier. These are large databases and the cost to prepare the original data, put it in machine-readable form, and index it, are quite substantial. In addition, there is a cost to prepare the master. In the past two years the mastering charge has dropped from $15,000 to under $3,000. This trend should continue and even lower prices can be expected.

What these advantages mean is that the CD-ROM is best for fairly large databases with many potential subscribers. The databases need to be large enough to require this technology, but not so large that they exceed its practical limits. If the application will only have a few users, then each user must be willing to pay a large share of the fixed production cost. An application with many users can be offered at very attractive prices.

The CD-ROM readers are, by computer standards, quite slow. Drives are now offered with average access times from 250 milliseconds to 1,000 milliseconds. This may seem fast, but when compared to the Winchester disc technology used on today's microcomputers it is almost 20 to 100 times slower. The slowness of the drive is noticeable in information retrieval applications. This slowness places more demands on the software developer and is a very real limit on the performance of such systems. The CD-ROM drives will not, due to technological limits, match the performance of magnetic drives.

Write Once Read Many

Another optical technology which shows promise for some applications in the reference area is WORM. As the name says, each area on the disc can only be written to once. It can be read repeatedly, but once data are

written to a part of the disc, it can never be rewritten. The most obvious use of this medium is for archival storage—i.e., as a replacement of magnetic tapes. Compared to tape, it is more durable, easy to store, and it can provide quicker access to archival data when placed in a jukebox.

One of the major advantages of the WORM disc is that there is a low fixed production cost—no mastering is required. The WORM will be cheaper for applications where only a few copies are to be distributed. Given the cost reductions in CD-ROM mastering, this gap is becoming very small. The problem with WORM is that the variable costs are much higher than CD-ROM. The disc itself costs $100 or more and each copy of the disc must be written on the computer and not pressed as in the case of CD-ROM. This means that more labor is involved in the production of the disc.

The WORM technology has faster access speeds—more like magnetic discs—than CD-ROM. It was designed with data applications—not audio—in mind. The readers are also more expensive—from $1,500 to over $10,000. The capacity of the drives is much larger; the range is from 150 million to over 3.5 billion bytes per disc.

One of the most serious drawbacks of this technology is the lack of standards. There are no standards for the format of the disc, the way data are encoded on the disc, or even on the size of the disc. This means that it is very likely that a reader purchased for one application cannot be used for other applications. This is not a great problem for archival purposes, but for reference uses it is a definite problem.

It does not appear that WORM technology will have a major impact on reference services. It will be used primarily in those cases where only a few copies of the database are needed such as a database of local material. But if the mastering charges for CD-ROM keep falling, even these applications may not be justifiable when the cost of the reader is included.

Applications for CD-ROM

It is important to remember that CD-ROM is primarily a publishing medium. The important attributes in evaluating a potential CD-ROM application are: (1) size of the database; (2) number of potential users; and (3) types of access required for the information. Each of these factors will be discussed in turn.

The database must be large enough to require a CD-ROM but not so large as to make its use impractical. A database of a few hundred thousand bytes can be distributed on floppy discs and would not require the users to purchase additional hardware for their microcomputers. In contrast, a database of 5 billion bytes, which cannot be logically segmented, would be

difficult to use because of the number of CD discs involved in a complete search of the database—between eight and sixteen depending on the level of indexing.

The number of potential users has a great influence on the per unit cost of the product. The high fixed cost and low variable cost associated with the production of CD-ROMs mean that applications which have only a few users will have a high cost for each user. Successful applications will have a large user population—the larger the better.

Information which does not benefit from the more flexible searching available on the CD might be better suited to print. The primary advantage of print media is that no hardware investment is required. One disadvantage of CD-ROM is that it requires the user, or the user's library, to purchase a workstation. The cost of this purchase will certainly discourage some potential users from purchasing CD-ROM products.

Given these characteristics, it seems that there are a number of areas in which the technology can be successfully applied—e.g., as a supplement to online searching, for reference material, and for full text.

The CD-ROM is initially being viewed as an alternative to online searching but will eventually lead to an increase in online searching activity. The advantage of CD-ROM over online searching is the lack of telecommunication and connect charges. This technology allows access to the database to be sold on a fixed cost basis enabling end users to conduct their own searches. Increased use of information systems by end users will lead to increased demands by these users to access online systems since online has advantages over CD-ROM for some applications. The online systems will be the access method of choice when: (1) the databases are large; (2) currency of the information is important; or (3) the database is not highly used. The last of these points is important to remember; an organization or individual is not going to purchase a database which is only used a few times a year. This technology will create users who are more informed about advantages of electronic sources and thereby increase the demand for electronic information services.

Another important application of CD-ROM technology, in addition to bibliographic databases, is that of reference material. As noted earlier, the first CD-ROM product was a reference work, Grolier's *Academic American Encyclopedia.* Microsoft has recently announced a product which is a writer's workbench of reference material: a thesaurus, dictionary, and quotations. This is a natural progression for libraries, which are providing access to bibliographic databases, to purchase electronic versions of reference material.

The last type of information to become CD-ROM products will be full text—both serials and monographs. While publication of these materials on CD-ROM is not ideal for reading the material, it does open up the

possibility of full-text searching. It now becomes possible to treat books or journals as reference material. With full-text searching available, it is possible to look in a journal to find the answer to a specific question.

The distribution of full-text material, and some reference material, presents special problems. The main problems are graphics and an acceptable level of readability of hardcopy output. OCLC has been working for a number of years on these problems and its Graph-Text project has developed solutions to both of these problems. It is possible to scan the graphics from journals and to use computer typesetting languages—such as TeX— to allow the workstation to produce a copy with print quality almost identical to the original.

Future Reference Services

As was stated earlier, CD-ROM is not a panacea for the delivery of reference services. While it is going to have a major impact on their availability, it will actually increase the demand for already existing electronic reference sources. In the future—and for universities the future is not very distant—there will be access to three major sources of electronic information: (1) CD-ROM, (2) local online, and (3) remote online. The distinction between the last two is who provides the machine.

The CD-ROM is well suited for medium sized databases where single user access is acceptable. If the databases are small, it is cheaper to distribute them on floppy discs unless there is a large number of small databases. Very large databases require a substantial number of CDs making it impractical to use for a complete search. The single user is really the limit of the technology. The CDs are not very fast and will not rival magnetic technology for speed. If there are multiple simultaneous users, the response time is going to be slow.

There has been a great deal of interest in providing access to CD-ROM readers on Local Area Networks (LANs). This appears attractive for three reasons: (1) providing remote access, (2) saving on hardware costs, and (3) saving on subscription costs. The first of these is a definite advantage and may allow users to access the database from a remote site. The savings on the hardware would only be relevant if the network is already in place. At present prices, the cost of the CD-ROM reader is about equal to the necessary network access hardware. The saving on subscriptions is driven by what the vendors charge for multiple copies. In addition, there are some operational costs of maintaining multiple readers. This system will also have to have either one CD per reader or a jukebox. If a jukebox is used, then some scheduling will need to be done.

If remote access by multiple users is desired, the better approach is to

use local online access. In this arrangement, a minicomputer or mainframe with magnetic storage provides access to multiple users through either a telecommunications line or LANs. The major drawback to this approach, at present, is the need to maintain the database locally. This requires a staff and a large amount of computer resources for database construction. One of the byproducts from the CD-ROM database production is a fully indexed database. Manufacturers of CD-ROM databases can sell the software and premastered databases to local sites to mount on their own machines. This would mean that bringing a new database up would be similar to installing a new software package.

The last source of electronic reference information is the remote online system—the current technology. This source is still best for very large databases, infrequently used ones, or where currency is required. If the user demands the most current information available, remote online sources will still have to be used to satisfy the demand. It is unlikely that either of the other local alternatives will be inexpensive enough to allow a user to have all the databases that will be required. To get access to these databases, remote online systems will still have a very significant cost advantage. The very large databases, as has been previously stated, will be most convenient to access remotely. Databases the size of *Chemical Abstracts* or *MEDLINE* will not make sense for local mounting for most institutions.

Conclusion

Optical storage media are going to have a much larger effect on the delivery of reference service than their direct applications might indicate. Users are going to become more aware of the benefits of electronic information sources and expect to have access to them for all their information needs. CD-ROM, while not appropriate for every information need, is going to educate end users about what is possible and thereby increase the demand for all forms of electronic information sources. CD-ROM will not cause the death of online services but will lead to increased use in the future.

The decreased cost for both storage and computing will lead not only to making searching of bibliographic sources more accessible but also to whole new areas of information searching and delivery. It is now possible, using desktop publishing systems, to do on demand printing of very high quality. This kind of system could be a very cost effective replacement for microforms. It is possible to set up such a system so that items stored in the library on CD-ROMs can be transmitted over a Local Area Network to be printed on remote systems in the user's office or department. The quality of such printing, using a $2,000 laser printer, is much closer to typeset material than to facsimile.

Full-text searching will become very popular. The first applications will be in reference material. It obviously meets the requirements of size and high use. The offerings now lack graphics and color, but there are no major technological barriers to supporting those capabilities. The next step is in offering books and journals in CD-ROM. This form of distribution has some drawbacks—e.g., you can't read it at home in an easy chair—but the full-text searching has distinct advantages. It greatly increases the usefulness of the item for reference applications. The searching allows the user to look up specific facts in the article, not simply find the article through some rather limited cumulative index.

The capabilities and cost of the technology are now at the point of making electronic searching available to the general public. The next several years are going to be a period of change. Many new types of services will become available, but more importantly, there will be a dramatic change in users' information-seeking behavior. They will be exposed to these more powerful tools and will adopt them.