



COMMUNICATIONS
FORUM

ELECTRONIC PUBLISHING: THE CD-ROM OPTION

Linda Moulton, Comstow Information Services
Julie Schwerin, InfoTech
David Roux, Datext

March 20, 1986

Genga Arulampalam, Rapporteur

Massachusetts Institute of Technology
Communications Forum

ELECTRONIC PUBLISHING: THE CD-ROM OPTION

Linda Moulton, Comstow Information Services
Julie Schwerin, InfoTech
David Roux, Datext

March 20, 1986

Genga Arulampalam, Rapporteur

Electronic Publishing: The CD-ROM Option

Linda Moulton - Comstow Information Services

Moulton began by defining CD-ROMs (Compact Disc Read Only Memory) as laser produced hardware devices capable of storing data at much greater densities than on conventional computer disks. However, unlike conventional computer disks, CD-ROMs do not allow information to be written to the media.

She then described the relationship between CD-ROM and corporate information management. Corporations are tremendous producers and users of information. Information in the corporation is produced in the form of correspondence, memoranda, formal reports, laboratory note books, patents, etc. Corporations use this internal information, as well as many external sources of information, to support corporate analysis and decision-making. Moulton stressed the importance of information retrieval system design and management in giving corporate information seekers fast and accurate access to key information.

Moulton said that information seekers within a corporation include managers, researchers, technical specialists, data processors and information processors. They may be seeking information for themselves or for someone else. She distinguished between data processors and information processors, the former being all those people in EDP and MIS functions who deal with the physical aspects of storing, manipulating (mathematically, statistically or sorting), retrieving and reporting data. Traditionally, according to Moulton, they have not been concerned with information concepts nor conceptually related data. Rather, these areas are the traditional purview of information processors: librarians, information specialists, literature indexers, abstractors and searchers.

Moulton stated that search intermediaries are needed because all the required information cannot be obtained from "the expert" (i.e. the original source of the information). Instead, information seekers must rely on the "next best" solution, a well tested methodology for systematically uncovering all the information related to a single problem.

In a corporation most information is stored in either a printed, digitized, or analog format. However, Moulton noted that some information is in peoples' heads and "a good librarian knows whose heads have non-recorded data and makes the referral when necessary and if possible". (She stated that her comments would be restricted to the stored or recorded full text of useful information.) Information, she said, is only useable if you can find it when you ask for it. In this context she described the types of information that might be needed and the types of mechanisms that have been created to give access to that information.

Information can be compartmentalized in many ways: text vs. pictorial, analog vs. digital, full text vs. citation, paperbound vs. microformed. She stated that these are not really true analogies or even meaningful relationships anymore because we now have analog pictorial information, and digitized and analog stored together on the same device. She emphasized that the information one needs may be available in a single form or packaged in many forms. Therefore, to be a good user of information you have to know not only where to find it but how it is packaged, organized and how to access it. Key to efficient information is knowing the information source's indexing mechanism. If the user does not understand the indexing scheme, he or she must use a retrieval expert who does.

Moulton said that the information seekers can use an information system to identify appropriate resources and avoid a serial or sequential search of every volume in the library. She asserted that the heart of any retrieval system is its index. Indexes are the tool of information managers and because they are so varied in type, quality and scope, often require a specialist to use them. She then linked the following resource formats:

- Printed
- Microcards
- Microfilm
- Microfiche
- Ultra-fiche
- Hard Disc: Data Files and Text (the only read/write devices currently in active use)
- Laser or Optical Media

In each of the above cases access to information requires some type of index if it is to be timely and efficient. Also there are some forms that lend themselves better to one type of index than another. The following is a list of indexes for printed formats.

- Book
- Indexing and Abstracting Sources
- Card Catalogs
- Edge-notched cards
- Rolladexes
- Directories
- File Folders
- KWIC and KWOC (used in on-line data bases)

Moulton then commented on some of the advantages and drawbacks of each type of index.

She stated that the advantage of CD-ROM includes its ability to store large amounts of information, to distribute that information, to manipulate sites and to give the user control over information access and retrieval on a single device. New CD-ROM applications are appearing at the rate of 3-4 per week. It is now used to store encyclopedias, newspapers, directories, patent gazettes and company reports. The CD-ROM format improves on compact storage, preserves quality, increases transportability and cheapens replication. However, some

material is unlikely to be susceptible to CD-ROM publication - for example, short single works that are more easily produced in paper or fiche, trace journals that must be read very soon after publication (and that readers may wish to carry with them), company current data, text books (which need to be read cover-to-cover) and other material that requires precise indexing. Moulton noted that a good search utilizes multiple indexes, information sources, storage media and storage formats. CD-ROMs, with its ability to store a wide range of information from diverse sources, raises the need for a common indexing scheme and language. Until expert systems can replace human experts in searching, a single medium like CD-ROM will remain only one of many types of formats for information.

Julie Schwerin - InfoTech

Schwerin estimated that 31 firms are producing or about to produce 58 products on either CD-ROM or digital video discs. Of those 58 products only 7 are on digital video discs. Many industry observers have projected that CD-ROMs, the information delivery version of the compact audio disc, will be as outstanding in the publishing field as the compact disc has been in the consumer entertainment field. Many users see CD-ROM as a convenient (i.e. compact), personal and cost effective medium for the distribution of digital information. Although CD-ROMs may benefit from the development expenditures, hardware production infra structure, and economies of scale of the compact audio disc, one cannot assume that the optical publishing infrastructure will match that developed for CD commercialization.

Schwerin mentioned that many of her remarks would be based on primary and secondary data collected over the last six months for a market research study conducted by InfoTech. The study concentrated on the library market, because the researchers felt that librarians were innovative enough to "grasp" a new technology and represented a well-defined market.

Schwerin stated that the computer, entertainment and information industries are converging into a difficult-to-differentiate industry. The optical medium is at the center of this new industry. CD-ROM, a digitally encoded read only optical storage media, has two main advantages over other publishing media - price leverage and standardization. The entertainment industry subsidised CD-ROM technology by developing laser video discs and compact audio disc technology, and by building a manufacturing infra structure for discs and drives which CD-ROMs can use. The computer industry has provided the fully developed hardware/software system which CD-ROM requires to selectively process and capture information. Phillips and Sony have successfully developed a world CD standard which will ensure compatibility between disc drives and various common computers. The end result, according to Schwerin, is very favorable

economics for the CD-ROM.

While CD-ROM is viewed by many in the information and computer industries as a means of breathing new life into existing products, CD-ROM ventures cannot be undertaken without significant commitment and confidence. Therefore she said, CD-ROM must be evaluated within three planes. (1) Other options for optical publishing. (2) Current means of electronic and print publishing. (3) Business and market objectives and strengths of the organizations seeking to exploit it. There are many roles being played by the suppliers of CD-ROM products and systems, and the buyers are various institutional and end user markets. The library market has been the earliest to develop as a cohesive market both because of their receptivity to new technology and because of the attractiveness of fixed price local access to machine readable databases.

Of the three industries mentioned earlier she claimed that the information industry has the greatest opportunity to capitalize on optical publishing because information customers are less concerned about the underlying hardware and software than in the actual usefulness provided by the system. In this context, Schwerin stated that CD-ROM offers a level of information storage, access and retrieval that is compelling to end-users. (Schwerin defined the information industry as one which captures, selects, packages, and delivers information products.) Schwerin claimed that of the hundreds of information based organizations which have seriously examined optical publishing, most are focussing on the bottom line. While they are doing pilot testing, in-house and some outside demonstrations, few are jumping in without a deliberate business analysis. She felt that CD-ROM will become an integral part of these organizations' businesses only if it enables them to increase market share or to break into a new market. From an operations standpoint, CD-ROM is evaluated for its potential to lower production and distribution costs, thereby improving marketing effectiveness and increasing product control. Because the cost of transferring existing machine readable databases to CD-ROM is much less than transferring to hard copies, products currently on paper or microforms are among the strongest candidates for CD-ROM. Schwerin emphasized that CD-ROM could be an important vehicle for the expansion of the information industry. However, because its impact may take a few years to develop, she recommended that venders of hardware, software and services invest more effort in encouraging this emerging industry lest momentum and business opportunity be diminished.

In terms of the computer hardware industry, CD-ROM can be viewed as the logical extension of the computer product line. However, Schwerin stated that few hardware firms have the vision and commitment necessary to really capitalize on the CD-ROM potential. In contrast the computer software industry is much more entrepreneurial and therefore has the potential to apply their expertise to create innovative CD-ROM information products. Because of low entry costs, Schwerin predicted that many new software

firms would spring up around CD-ROM applications.

In terms of the entertainment industry Schwerin predicted that when the growth rate from converting audio tapes slows, the entertainment industry, producers and distributors will begin looking around and making related programming. She said that it is difficult to imagine CD-ROM existing were it not for the technical and market success of the CD audio. The entertainment industry sees information products as new programming for its established customer base and will seek to exploit those products that are oriented to the consumer mass market. This would probably happen with the third generation CDI (Compact Disc Interactive).

The library community is very progressive in using new technology to improve their responsiveness to patrons. The future of optical publishing in libraries is to offer more information to patrons, save space, and upgrade hard copy collections and reach more patrons through the integration of user friendly software. She said that the most desirable applications for librarians include:

- full text periodicals
- full page information including graphics
- static portions of online databases
- bibliographic databases and
- direct substitute for micro film.

David Roux - Datext

Roux began by briefly describing Datext as a two year-old start-up aimed at taking advantage of the new generation of mass storage growing up around the optical media. Datext observed three important features of the marketplace when it was first organized. First, there was a huge reservoir of relatively untapped demand for electronic business information. Second, both large and small firms were beginning to rely heavily on PCs. Third, Datext observed the continuing evolution of very cheap low cost mass storage.

Datext, he said, is a distributor of value added information. Their aim is to get the information directly to departmental end users. Roux claimed that while librarians and information specialists dispense a large amount of electronic information that comes into a corporation, the overwhelming information budget (i.e. the amount really spent on consulting studies, contract information, multi client research studies, magazines, etc.) is contained in departmental budgets - Datex's main target.

Datext produces all of its own PC software for information access, retrieval and manipulation and does all of its own database preparation. Datex's direct sales force located in Boston, New York, Chicago and Los Angeles demonstrates,

installs, and provides all the post-sale support for its products. Datex's product is a four disk industry data base aimed at corporate end users who must track competitive opportunity. For \$20,000, the customer obtains an annual subscription to Datex's data base which is published monthly via delivery of four updated disks. Included in the annual subscription is use of the CD-ROM reader, a year's license to the associated software, documentation and support. Their target market includes very large corporations, major consulting firms, big banks, and financial service firms.

Roux said that in developing and designing products, Datex counts on relatively easy acceptance to avoid spending a lot of time marketing the product and educating the users. The product has to be exactly what the user needs. The corporate end user requires a system that provides "one stop shopping", local processing and control, common indexing across databases, simple report generation, easy document formatting, simple file transfer and subscription pricing.

The Datext system gives instant access to comprehensive business information regarding companies, industries, lines of business, and corporate executives. The information covers all publicly listed companies in the US (about 10,000 public companies), 900 different lines of businesses, 50 industries, 8,000 different senior officers and executives in those companies. Each of the four CD-ROM disks hold 250,000 pages of business information. The annual subscription provides a monthly update of the information.

Roux said that every week they receive 5-10 proposals from publishing companies and database producers recommending what they believe to be an absolutely perfect product for CD-ROM applications. Datext has a very active product development program. He said that they had just announced their plan to market a database for high technology companies. He claimed that Datex will be introducing a new product about every three months though 1987.

Roux concluded by stating that there is no shortage of CD-ROM applications, but that the real trick is converting a good idea into a marketable product. Buyers are unconcerned over the technology used and care instead about the business value delivered. Pricing is key to this concept of marketing value. Publishers will succeed if they can capitalize on their knowledge of publishing, add value and market a truly new product. Innovation is the name of the game.

Speakers' Comments and Responses to Questions

A question was asked as to whether it was possible to store graphics as opposed to text or numerics on CD-ROM. Schwerin responded that there were two firms that produce

strictly graphics. First, Laser Track who put maps (like airport approach maps) onto CD-ROM to make it compact and convenient. Second, Geo Vision who produce topographical and geographical information for developers, and oil drilling companies.

A member of the audience inquired regarding the possibility of 'bootlegging' (i.e. publishing in book form). CD-ROM stored data. Schwerin stated that encryption techniques are being developed to prevent this kind of pirating.

In response to a query regarding current production costs Schwerin stated that \$4-10,000 is required to produce a master disk and about \$10 for replicating each marketable CD-ROM disk. She noted that although CD-ROM is very similar to CD audio it currently costs 5-8 times as much as CD audio. In her opinion, prices are bound to come down.

A question was asked relating to a corporation's in-house data publishing (e.g. CAD/CAM), to which Roux responded that it was not an "internal" or "external" issue, but rather that it depended on the size of the project - very much like printing.

Roux, responding to a question about the storage of classified information on CDs, stated that he did not foresee a problem because encoding could be used to protect classified information. Regarding the training period required to use the system he said that a) though Datext offers training, "no one takes it". Most users "get upto speed" using reference guides and on-line tutorials. He further added that the price of their product is determined by what the market is willing to pay. A member of the audience asked what legal risk the CD-ROM publisher bore with regard to the accuracy of the published information. Roux responded that publishers make no independent guarantees about data accuracy, and that the issue does not arise.
