

CD-ROM IN THE AGE OF INTERNET

by

Bob Conrad and Dennis Depp

Information Management Services Division
Oak Ridge National Laboratory*

Abstract

Networks are hot and CD-ROM is also hot, but how do they mix? CD-ROM is a relatively inexpensive medium for storing and delivering information, and increasingly, users are connected to networks. But the technologies have developed separately, and there are obstacles to their integration. Drawing on their experience networking CD-ROMs at Oak Ridge National Laboratory, the authors discuss CD-ROM's strengths and weaknesses as a technology for delivering information to the desktop. CD-ROM networking solutions are LAN-based, not "open systems." Despite this limitation, due to the large number of information resources available on CD-ROM and the relative ease of installing and maintaining databases on CD-ROM, CD-ROMs remain an essential piece of the electronic information puzzle.

Introduction

The Information Services Division at Oak Ridge National Laboratory (ORNL) initiated a project to provide access to information on networked CD-ROMs in early 1992, largely in response to the Engineering Division's interest in accessing engineering standards electronically across the network. Traditionally engineering standards were available only on microfilm at site engineering libraries. In early 1993 18 CD-ROMs on 2 CD-ROM servers became available across the network at all three Martin Marietta Energy Systems sites in Oak Ridge. In addition to engineering information, the information products selected for the CD-ROM network included bibliographic indexes/abstracts, phone directories, and fulltext journal articles.

*Managed by Martin Marietta Energy Systems, Inc., for the U. S. Department of Energy under contract No. DE-AC05-84OR21400.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

The major factors considered in selecting the initial information resources for the network were (1) the user demand; (2) the usage history of CD-ROMs then available at the Energy Systems libraries; (3) the potential for cost savings by eliminating duplicate subscriptions of CD-ROMs held by more than one library; (4) the availability of the information resource on CD-ROM; and (5) the network license policies of the vendors. The vendors' policies were a more significant factor than one might expect. Two H.W. Wilson CD-ROMs, to which the libraries subscribed, were networked because Wilson has the enlightened policy of not charging extra to network their products. A heavily used CD-ROM, *Books In Print*, was not networked because Bowker, the publisher, included the number of buildings from which the CD-ROM would be accessed as a factor in determining the network license fee.

The original collection development decisions for the CD-ROM network proved to be relatively good, but the electronic information landscape has changed dramatically in the last two to three years. The rapid spread of networks, the shift from centralized to desktop computing, the dramatic proliferation of Gopher and World Wide Web servers, the decreasing cost of disk storage space, and the exploding number of CD-ROM titles are just some of the current realities. There is also a growing expectation by end-users that they should have direct access to information resources, preferably on their desktops. All of this, of course, has both provided new opportunities and complicated the choices for information providers.

Electronic access to the *Federal Register* is a good example of the changing information environment. Until it became available on CD-ROM in October 1990, electronic access to the *Federal Register* required a subscription to an online service such as Dialog or Legi-Slate. Now the full *Federal Register* and subject collections of federal regulations are available from several CD-ROM vendors and from more than one vendor over the Internet. The collection development issue is changing from being a question of availability to a decision about the best product and the preferred method for providing electronic access.

With the changes in the information industry and with a better understanding of the strengths and weaknesses of CD-ROM technology, we are finding that we are asking new questions. As relative novices to networking, and totally inexperienced with CD-ROM networking, our initial focus was naturally on making the technology work. Although several factors (listed above) were considered in deciding which CD-ROMs to network, the primary collection development question was, "Which CD-ROMs should we network?" We have found the following questions to be more useful when deciding how to deliver electronic information:

- What information needs to be accessed electronically?
- Is this information available electronically?
- What is the most appropriate technology or method for delivering the information?

In other words, to select the appropriate means for delivering information, information providers need to understand the users' needs, be aware of the available information resources, and be knowledgeable of the strengths and weaknesses of different

information technologies. Drawing on our experiences as information providers and with CD-ROM technology, this paper will attempt to identify some situations in which CD-ROM currently may be the appropriate technology for delivering information electronically.

CD-ROM Networking Technology

Before considering CD-ROM's place in the total information picture, let's briefly review the state of CD-ROM networking technology.

The redirector approach is the most common method for networking CD-ROMs. It requires (1) a local area network (LAN), (2) a dedicated CD-ROM server, (3) CD-ROM drives (which are preferably attached to the CD-ROM server with a SCSI—small computer systems interface), (4) CD-ROM networking software, and (5) CD-ROM subscriptions. CD-ROM networking solutions for peer-to-peer LANs and NetWare Loadable Module solutions for networking CD-ROMs do not require a dedicated server.

The following sources are good starting points for information on the "how to" of CD-ROM networking:

James Karney, "Sharing a CD-ROM," *PC Magazine*, 13(4), February 22, 1994, p. 171.

Frequently Asked Questions List for CDROMLAN@IDBSU.IDBSU.EDU
The FAQ is available via anonymous ftp at ftp.cdrom.com in /pub/cdrom as cdromlan.faq

Marshall Breeding, "Competitive Outlook: Shared CD-ROM for the Workgroup," *Datapro on CD-ROM: Computer Systems Analyst*, May 1994.

In his article, James Karney cautions,

It would seem to be the perfect match: networks and CD-ROMs. Networks allow people to share resources throughout an organization. CD-ROM technology uses an inexpensive medium that can store and retrieve vast amounts of data. As enticing as it is, though, bringing the two together is no small matter. You can get there from here, but the road to sharing a CD-ROM on a LAN—at least right now—is marked by hardware-connectivity and drive-mapping pitfalls.¹

Part of the challenge in networking CD-ROMs is that the technology was not developed for networks but to run on individual computers. The major operational problem we encountered in this area is "RAM cram." CD-ROM applications generally require a large amount of conventional memory to run. On the network, by the time all the drivers, network programs, and Microsoft Extensions are loaded, there frequently is not enough RAM available on the client's workstation to run the CD-ROM applications. Users wanting access to the CD-ROMs may not have the requisite DOS skills to do the necessary memory management on their computers.

More significant in determining CD-ROM's place in the total information picture is the fact that CD-ROM networking solutions are LAN-based, not "open systems." This creates two significant, interrelated problems. First, to access the networked CD-ROMs, the individual user must be on the same LAN or wide area network (WAN) as the CD-ROM server, because the CD-ROM networking software and the retrieval software for the CD-ROM are loaded on the network drive. This is not an obstacle when everyone in the organization who wants to access the CD-ROMs is on the same LAN or WAN. However, in most institutions there are many different types of computers and networks. Most networking solutions use the LAN's protocol or Netbios for communicating, not TCP/IP, the Internet protocol.

Even if users could telnet to a DOS-based LAN from their Macintoshes, they would encounter the second problem. Although the data format on virtually all CD-ROMs complies with ISO 9660, the international standard for CD-ROM file structuring, the retrieval software on the CD-ROM does not comply with ANSI/NISO Z39.50, the Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. The retrieval software is proprietary and designed to be run on a single type of computer, (i.e., IBM compatible, Macintosh or UNIX). As of 1992, 73% of CD-ROM titles were available only for IBM compatible computers, 14% only for Macintosh, and 13% were available in either flavor, IBM compatible or Macintosh.² At first it would seem that if vendors put both Macintosh and DOS-based retrieval software on the CD-ROMs (and some do), this would solve the problem. But in fact, most CD-ROMs require the networking software to be loaded on the LAN server so the compatibility problems are not resolved.

CD-ROM networking technology clearly has some limitations as a means for delivering information to the desktop. Information providers should consider alternatives such as Internet and in-house databases as alternative solutions. There is no single solution for electronic information. Having determined that we would like to provide electronic access to a particular information resource, three areas to consider in selecting the medium are availability, access, and administration. The focus in this paper will be on the strengths and weaknesses of CD-ROM in the total information matrix.

Availability

A major strength of CD-ROMs as a medium for delivering information electronically is the large and growing number of CD-ROM titles. In 1993 there were 4000 CD-ROM titles;² by the end of 1994 the number of information-related commercial titles is expected to reach 9000.³ The September 19, 1994, issue of *Fortune* magazine explodes the myth that most CD-ROM sales come from consumer multimedia disk titles. It reports that information-related commercial CD-ROM titles produce over 80% of total CD-ROM revenues, according to InfoTech, a market research company in Woodstock, Vermont.³

The following three publications are helpful in identifying information resources available on CD-ROM:

Paul T. Nicholls, *CD-ROM Buyer's Guide & Handbook: The Definitive Reference for CD-ROM Buyers*, 3rd ed. Wilton, Conn., Eight Bit Books, 1993. (Especially useful

for collection development are Nicholls' evaluative reviews of 200 core reference CD-ROM titles.)

James H. Shelton, ed., *CD-ROM Finder: The World of CD-ROM Products for Information Seekers*, 5th ed. Medford, N.J., Learned Information, Inc., 1993.

Special Interest Group on CD-ROM Applications and Technology, *SIGCAT CD-ROM Compendium*. Washington, D.C., U.S. Government Printing Office, Office of Electronic Information Dissemination Services, April 1994. (The compendium tracks and highlights CD-ROM products that contain federal government information.)

Just because the library has a CD-ROM network and a particular product is available on CD-ROM does not make CD-ROM the platform of choice. Some products, such as *Current Contents* from ISI, are available electronically on tape, providing an alternative to CD-ROM. Another option is to contract with vendors to provide information resources across the Internet. However, many commercial and government information resources are only available electronically on CD-ROM or from an online vendor.

Despite all the interest in Gophers and World Wide Web servers, Internet is not currently "the" information solution simply because most of the best information resources are not available via the Internet. One of the major weaknesses of Internet is the current lack of comprehensive and commercial databases. It appears information vendors were caught off guard by the rapid expansion of networks and Internet. Many are still focusing their resources on developing CD-ROM products, ignoring the inherent problems with networking CD-ROMs. Many vendors appear oblivious to the potential for developing Internet services, mistakenly thinking that the value of their product is the CD-ROM interface they developed rather than the information they provide. Because of open architecture and common user interface, Internet technologies such as Web servers and WAIS (Wide Area Information Service) search engines may also be attractive options for delivering commercial information products on in-house servers. Hopefully, in time more vendors will offer alternatives to CD-ROM.

Access

When there are alternatives for providing electronic access to a particular resource, several access issues should be considered in determining if CD-ROM is the best choice: What types of computers and networks will be accessing the CD-ROM network, how often will they access it, and how large are the databases they need to access?

Because CD-ROM networking solutions are not "open" but tied to LANs, they are best suited for a homogeneous network in which all users access the same LAN from the same type of computer. Even in a such a uniform environment, there are problems with access. With each CD-ROM, the vendor provides data, a search engine, and a user interface all packaged together. Though the vendors' search engines are often robust and the proprietary interfaces generally easy to use, the user is confronted with multiple interfaces if the network has CD-ROMs from multiple vendors. These multiple interfaces are naturally confusing for the users, and the quality of interfaces and search engines varies from vendor to vendor. Because of the way these CD-ROMs are currently packaged, one must use the

vendor-provided interface. To reduce the number of interfaces which the user must confront, the best the network administrator can do is to attempt to limit the number of vendors from which CD-ROMs are purchased, since some products are available from multiple vendors.

Two major CD-ROM vendors appear to be addressing the proprietary interface problem. SilverPlatter has developed a client/server product called the Electronic Reference Library (ERL) 1000, which will allow clients on different platforms to access CD-ROMs from UNIX-based servers. Currently only a DOS client exists, but there are plans to release Windows, Macintosh, and UNIX clients. ERL currently uses a proprietary protocol, Data Exchange Protocol, but SilverPlatter is planning a Z39.50 version.⁴ Also, in a recent meeting at ORNL, Dialog's vice president and general manager of the CD-ROM division, Barry Richman, stated that Dialog is committed to Z39.50 compliance for the Dialog Ondisc products. These are promising developments. Nonetheless, the representatives of the CD-Recordable industry, which supplies the software for creating CD-ROMs, generally appear unaware of networking issues.⁵

The limitations of CD-ROM technology in the nonhomogeneous network environment have already been described. When possible, institutions such as ORNL, which have many different networks and types of computers, need to consider alternative means of providing electronic access to information. For this very reason—the desire to provide the resource to a broader number of users—ORNL is currently implementing *Current Contents* on an in-house server.

Some software and hardware products provide partial solutions to the problem of multiple platforms. For example, Logicaft's LANCD networking software allows multiple protocols to access the CD-ROM server; however, this does not solve the problem of CD-ROM software not being platform independent. Both Logicaft's 486Ware and Virtual Microsystems' CDaccess allow VAX/VMS users to run DOS-based CD-ROM products. For a nonhomogeneous environment, Logicaft's Omniware server allows Macintosh and UNIX users to telnet to a UNIX server and then access the DOS-based CD-ROM network. The Omniware server eliminates emulation problems and comes in single user or four-simultaneous-user versions. ORNL has purchased one of Logicaft's Omniware servers to allow UNIX and Macintosh users access to the CD-ROM network, but it is a partial solution, at best.

Another access-related issue to consider in selecting a medium for electronic information is how often users will access the information. Cost and performance are both involved in this decision. CD-ROM network licenses can be purchased based on the number of simultaneous users. The greater the number of simultaneous users, the higher the cost of the subscription. A subscription for loading information on an in-house server is usually an unlimited use site license. Cost comparisons must be made for each product. In our experience, a site license for information on tape is not necessarily more expensive than a limited simultaneous use license on CD-ROM; however, the administration costs (discussed in the next section) must be included in determining the total cost of delivering the information. Because of the slow speed of CD-ROM drives, performance will degrade more rapidly as the number of simultaneous users increases for a CD-ROM than for information on a hard drive. Heavily accessed CD-ROMs might be copied to a hard drive (which can still be accessed across the CD-ROM network), or the information moved to an in-house server. CD-ROM is likely to be the better solution for less frequently accessed information.

CD-ROMs can store a large amount of data very cheaply. Each CD-ROM holds approximately 650 MB. Although this is a considerable amount of data, is this enough? Because Microsoft Extensions requires a drive letter to be assigned to each CD-ROM being searched, in the network environment there is a limit to the number of disks which can easily be loaded. On products which span multiple disks, vendors take three basic approaches. The first is to have each disk as a stand alone product. Each disk is selected, loaded and searched as a separate database. The second approach is to initially load multiple disks when the database is selected, but then each disk is searched separately. With this approach, there is usually a feature which permits a search strategy to be easily reexecuted on the other disks. The third method is to both load and search all the disks simultaneously with a single search. This treats multiple disks as a single database and is probably the least utilized approach. Because each disk loaded must be assigned a drive letter with the second and third approaches, this complicates the management of extremely large databases on CD-ROM.

When considering the different PC configurations and taking into account assigned network drives, we consider five to six disks to be the most that can practically be loaded simultaneously, without having to remap network drives. This means that approximately 3 GB of data can be accessed on CD-ROM conveniently during any given search. Some large databases, exceeding 2-3 GB, may be better suited for an in-house database because of their large size and access pattern. An example of this is the DOE Energy Science & Technology database which Dialog sells on 12 CD-ROMs. It would not be practical to load this many disks simultaneously; yet, splitting the database may be inconvenient for users.

To summarize the access issues, networked CD-ROMs work best in a homogeneous environment. There are products available to assist in providing access to a CD-ROM network in a nonhomogeneous environment, but they are temporary solutions. Licensing agreements and performance limitations make networked CD-ROMs best for low and moderate use environments. High use and extremely large databases may be more easily managed and more conveniently searched when loaded on an in-house server.

Administration

A major strength of CD-ROM networks is their ease of administration. Here we will consider only those tasks relating to the administration of the CD-ROM network, not the LAN administration.

Installation of a CD-ROM network requires a moderate level of technical expertise. The CD-ROM network administrator must have a basic knowledge of the LAN software, a basic knowledge of DOS, and the ability to install the CD-ROM server and client software. No programming knowledge above basic DOS .bat files and no knowledge of database design are necessary for implementation. At ORNL the CD-ROM network was installed by a reference librarian who had a good understanding of DOS. Communication between the network administrator and the CD-ROM administrator is essential to a smooth implementation and operation.

Installation of new products is relatively simple. When installing CD-ROMs, the retrieval software is normally installed to a network hard drive. Some programs offer the

option of placing all, some, or none of the retrieval software on the hard disk, in which case placing the software on the hard drive usually provides the best performance. Once the software is installed on the network, the product needs to be installed in the menu software. This involves writing basic DOS commands. Finally, the software must be tested to ensure users can access it.

One of the stumbling blocks with networking CD-ROMs is that frequently they were designed as standalone products and lack adequate documentation for network installations. Occasionally, a CD-ROM installation program will want to write user files to an area of the network to which you do not want to give the users access, or it attempts to write program files to the local drive rather than the network. Though frustrating, these types of installation glitches can normally be resolved by modifying path statements in .bat files or through the use of DOS variables. Because the networking instructions of CD-ROMs are often incomplete, a phone call to the vendor's technical support is not uncommon. The software for CD-ROM vendors' products is usually the same for all their products. As a result, once you solve a problem for a particular vendor, you will not have the problem with other disks from that vendor. We have observed that the quality of vendors' installation programs has improved significantly over the last two to three years.

Once CD-ROMs are installed, updates to products are performed with ease. Installing an update can be as simple as removing the old CD-ROM from the drive and replacing it with the new, updated disk. Some products require that the installation program be rerun with each update.

A brief comparison of administration of CD-ROM databases with administration of the same information on an in-house server highlights the strength of CD-ROM in the area of administration. Assuming the computer system hardware exists, a suitable database program is required with which to manage the information. On the CD-ROM, the search software is an integral part of the product. To install the information on an in-house server, the information must be uploaded from tape into the database. This requires knowledge of database design to set up the database, specific computer knowledge to upload the tapes, and possibly some programming knowledge to format the tapes. In addition to creating and loading the data, a suitable user interface will need to be purchased or developed. None of these steps are required with the CD-ROM product, and the user interface, like the search engine, is provided.

Updating the database on the in-house server requires uploading additional tapes and, possibly, reindexing the database; these tasks are performed by the vendor for the CD-ROM product.

Installation of additional information products on the in-house server is likely to require additional development work. The advantage of the in-house server solution is that, whether the interface was purchased or developed, the interface will be consistent throughout each product installed.

From the standpoint of administration, the strength of CD-ROM for delivering information is in the minimal technical expertise required and the relative ease with which products can be installed and updated. This led to the decision at ORNL to subscribe to the DOE Energy Science & Technology database on CD-ROM, despite the access limitations and drive-mapping problems caused by the size of the database. In the long-term, access issues may dictate moving the database to an in-house server.

Conclusion

CD-ROM technology is not "the" solution to providing desktop access to information, but currently it is a significant piece of the electronic information puzzle. The large number of CD-ROM titles available and the relative ease with which CD-ROM databases can be installed and maintained are its greatest strengths. Network access issues are the primary weakness.

What might the future hold for CD-ROM technology? As Dialog and SilverPlatter begin to move to Z39.50-compliant search engines, other vendors will likely follow suit or be left behind. Vendors will realize that the value of their product lies first in the information and second in the search engine. Because users want a common interface they are familiar with, the value of proprietary interfaces will decrease. Z39.50-compliant search engines on CD-ROM will allow for seamless integration of CD-ROMs with other means of providing desktop access to information, such as the Internet or in-house servers. Where the information resides will be transparent to the user. With this transparency, CD-ROMs will provide an economical means of providing desktop access to less frequently accessed, yet valuable, information resources.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

1. James Karney, "Sharing a CD-ROM," *PC Magazine*, 13(4), February 22, 1994, p. 171-178.
2. Paul T. Nicholls, *CD-ROM Buyer's Guide & Handbook: The Definitive Reference for CD-ROM Buyers*, 3rd ed. Wilton, Conn., Eight Bit Books, 1993, p. 17-18.
3. Stephanie Losee, "Watch Out for the CD-ROM Hype," *Fortune*, 130(6), September 19, 1994, p. 127-143.
4. Marshall Breeding, "Competitive Outlook: Shared CD-ROM for the Workgroup," *Datapro on CD-ROM: Computer Systems Analyst*, May 1994.
5. Discussions between Bob Conrad and CD-Recordable vendors at SIGCAT 1994 conference in Tysons Corner, Va., April 25-28, 1994.

Bob Conrad is both a reference librarian and the administrator of the library's CD-ROM network at Oak Ridge National Laboratory. Prior to coming to ORNL six years ago, he was a law librarian at the Tennessee Valley Authority. He frequents the information superhighway and the Appalachian Trail.

Dennis Depp has been working at ORNL for the past two years and is currently the supervisor for Library Support Services and an administrator for a Banyan Vines LAN. Prior to his current position, Dennis developed databases primarily for IBM clone computers.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."