CD-ROM VERSUS WEB-ACCESS TO EXTERNAL DATABASES: EXPERIENCES AND INSIGHTS

Sharon Yang and Robert J. Lackie,
Rider University Libraries

Keywords: CD-ROM databases, Web-based databases, Remote access, Database access technologies

Abstract: CD-ROM databases are still popular today, but their use is steadily declining because of the availability of Web technologies. Many companies have mounted their databases on the Web. Users can access those databases through almost any browser as long as they have Internet access. Because the databases are located remotely, libraries no longer need to purchase servers and CD towers to house CD-ROMs, nor do they need the same level of expertise in-house to maintain these systems. However, there are also inherent sacrifices, including compromised access speed, Internet traffic jams, and ISP downtime. Local customization may be limited or even impossible since libraries do not have control over the remote systems. Some of the most valued databases may not be offered by vendors through the Web. Copyright and licensing may present another problem with Web-based databases. Because of the technical complications, there are still problems in administering access to Web-based databases. IP filtering and password may not be the ideal ways to control access as they equally hinder open access. The pros and cons of both Web and CD-ROM delivery mechanisms will be discussed. However, there has been progress in network mass storage technology and remote access to shared resources in the last few years. These new advances include DVD, data caching software and hardware, and server-based computing (also called terminal services, and thin-clients). These new developments may help to eliminate some of the limitations of CD-ROM technology. We hope that by discussing the pros and cons of both Web and CD-ROM delivery mechanisms along with new developments surrounding these technologies, we can assist those professionals caught in the turmoil of deciding between CD-ROM and Web access today and in the very near future.

1. INTRODUCTION

Even now, at the threshold of a new millennium, there is still no golden path to access for our patrons concerning database technology in our libraries. In fact, there is really no agreement on the best practices, or what we may achieve in this new decade, with technology. Experts within the electronic publishing industry are saying that the CD-ROM and maybe even DVD-ROM are transitional mediums, soon to be eclipsed because better ways of distributing information are 'just around the corner.' This is not new to us—we have heard this before.
Back in 1991 in her *Library Journal* article, Carol Tenopir talked about the fear that subscribers would abandon online access to external databases in favor of CD-ROM (Ref. 8). This time, the fear is of abandoning CD-ROM networks for Web-access to remote databases. In late 1996, Bessie Carrington in her *Online* article, talked about "new and better ways of retrieving information," concentrating on electronic access to databases (Ref. 2), as did Apurva Mehta in his 1996 article in *Computers in Libraries*, discussing patron demands for remote access to full-text databases (Ref. 7). These articles, and several others, discuss the introduction of a new medium, or a new level of access to an existing one, and this can be seen as a threat to what systems libraries already have in place. Interestingly, it seems that, even now, CD-ROM and online databases, are peacefully coexisting, but many wonder for how long.

Marshall Breeding in his 1999 *Computers in Libraries* article states that we are now "witnessing an unmistakable migration away from CD-ROM-based products toward Web-based information resources accessed via the Internet" (Ref. 1). Rider University Libraries is certainly no exception to the rule. We have seen almost all of our stand-alone CD-ROM workstations and CD-LAN replaced with Web-based databases, allowing network access from computer labs and offices all over campus, as well as from home. *ABI-Inform* is a good example of a recently converted database that we now receive through the VALE Consortium, with its shared servers located at Rutgers University and Seton Hall University. However, Mr. Breeding also believes that CD-ROM and DVD-ROM technologies will still be around for awhile. Still, many of us are caught in the turmoil of deciding what to do concerning access right now, so where do we go from here? Let's first look at the pros and cons of both the CD-ROM and Web-based delivery mechanisms.

2. WEB-BASED DELIVERY VERSUS CD-ROM

2.1 Advantages of Web-based Access

Many database providers have mounted their databases, containing citation, abstract, or full-text information, on the Web. Because of this, users can now access many databases through almost any Web browser as long as they have Internet access. The main advantage to online access is the instant retrievability of large quantities of data, with some databases updated daily, hourly, or even more frequently. Most CD-ROM databases, on the other hand, are usually only updated monthly or quarterly, which makes this less appealing to users.

A bonus associated with having the databases accessed remotely by our patrons is that our library no longer needs to purchase additional servers and CD towers to house CD-ROMs, nor do we need to allocate a large percentage of time and expertise to maintain these systems. Most CD-ROM databases are proprietary. Each CD-ROM database has different system requirements and needs special attention in its installation. Those individual CD-ROM programs, in combination, take up a lot of system resources on a computer. Sometimes, there is the issue of compatibility among those CD-ROM databases running side by side on the same machine. It takes a lot of expertise to maintain CD-ROM databases and their associated programs locally. It is definitely easier for our systems librarian and UNIX administrator here at Rider University to support the use of Web-based databases, instead of the existing CD-ROM LAN, especially since almost all of our public workstations now provide a Netscape or Internet Explorer Web browser. Many of our databases are accessible on our two campuses from our Library Home Page through IP address recognition. As much as it may hurt to say this, our students and faculty do not have to physically come into our library to obtain a fairly large percentage of the information they need, as long as the Web servers/ISP are all up and running, that is.
2.2 Advantages of CD-ROM Access

Yes, as good as Web-based access may sound, accessing information databases through the Web also entails inherent sacrifices, including compromised access speed, Internet traffic jams, and ISP downtime. When many of our most frequently used databases are very slow to respond or are not available at all, reference service and library instructions can grind to a halt, and the satisfaction of those users attempting to access our databases remotely from off-campus plummets. That is why we believe we should not 'put all of our eggs in one basket.' For instance, several of our more heavily used databases, such as ERIC, Social Science Index, General Science Index, MLA, and Humanities Index, are housed in the Library on our ERL server. These are in-house, Web-based databases which we receive on CD-ROM from SilverPlatter. When there is a problem with the ISP, access to the whole Internet is down, and we find ourselves totally isolated from the outside world. All of the remote databases housed with the vendors will not be available to us. However, since the five ERL databases are housed within the Library, their access will not be affected by the ISP. We can still access them and several other CD-ROM databases, such as Simons Research and Market Study - Choices II, OECD Health Data 99, and National Trade Data Bank, on our CD-LAN. This is important because when the Internet is very slow or the connection fails, we still are able to successfully continue serving our faculty and student information needs in many areas—as you would probably agree, demand for our services does not drop just because Web-access in down.

Another advantage of CD-ROMs is that some of the most valued databases may not be offered by vendors through the Web, at least not now or in the foreseeable future, and even if they are, the CD-ROM versions may be much cheaper to buy than their online counterparts (we’ll cover more on cost later). Additionally, the user-friendliness of the interface might be a deciding issue. We ran into this initially with the conversion of our CD-ROM version of ABI-Inform (Business Periodicals On Disc) to an online version offered through VALE. Many of our librarians and teaching faculty were not happy with the VALE interface—fortunately, we were able to make changes and are continuing to do so, but that was because we had a say in changing and upgrading the consortia’s common interface. Most local customization is not possible when libraries do not have control over the remote systems, such as with most Web-accessible databases.

2.3 Other Factors Associated with Both Web and CD-ROM Access

Speed of access presents other problems; for instance, Peter Jacso states in his 1999 Computers in Libraries article that one solution many libraries are using to give greater access (especially when the library is closed) to CD-ROM resources is to add these servers on WANs and make them “accessible through Internet connections, but in practice [he] saw many of them working so slowly and crashing so often under the load that patrons just gave up on using them” (Ref. 6). We must concur with Mr. Jacso, because although we, too, have made use of WANs to house some of our databases, providing the greater access demanded, we have seen how slow and cumbersome this method can be at times for students and faculty—we continually hear this complaint at the Reference Desk.
With the advent of Web-based databases, copyright and licensing has posed a challenge. It has become very easy to download an electronic full-text article and harder for vendors to maintain tight control. Many vendors of Web databases lack the means to administer the licenses. The use of passwords is not a convenient way to maintain license agreements, especially in an academic environment where a huge number of users are involved. IP filtering is more successful, but it prohibits those faculty and students living off the campus from accessing resources, in most cases. So far, the more successful method is a combination of IP filtering and password use. But not every vendor can provide it due to technical limitations. A better way to control copyright and licensing has to be developed by vendors for Web-based databases.

Copyright still presents a problem for us, even when the database vendor providing the full-text articles has a contract with the publishers of the periodicals. Yes, as Mr. Mehta says, our library "patrons can print and download full-text articles without being overly concerned about breaking copyright laws..." (Ref. 7) because the responsibility mainly lies with the vendor; however, database vendors can easily lose the rights to the full-text, leaving us in the lurch. We, then, have to rely on other methods to retrieve the desired information, such as with interlibrary loan, again, putting our library or the requesting library's staff back into the copyright spotlight.

The use of passwords is not an ideal way to control access to online databases. Without a single login solution, users often must remember multiple passwords, and that is really asking for trouble. When librarians are forced to recall several different passwords, and we sometimes forget them even when we use them fairly often, it is not hard to see how many of our patrons can forget, too. Students living on campus at Rider are continually asking for help with resetting their original e-mail passwords because they forgot their new ones. Unfortunately, those who "remember" theirs often do so by writing them down and leaving these notes in plain sight, and that is also true for passwords that gain access to remote databases.

However, to gain access to the databases without password verification, in general, libraries must access the databases from static IP addresses that are registered directly with the vendor. We then purchase a site license, allowing us to offer Web-based access to a particular database for everyone on campus. This sounds great, except when the vendor inadvertently drops an IP range, for instance, which just happened recently to one of our databases. Then, not just one or a few users are inconvenienced, but all of our users, including the librarians! Usually this is a fairly quick fix, but our control is shaky here. Another problem associated with the IP filtering technique is that faculty and student living off the campus have little access to those Web-based databases, if at all. Still, to many, it is worth any problems that come along with it, although site licenses can be very expensive.

Cost is definitely an ongoing battle. As we stated earlier, the main advantage for Web-based services is the instant access to huge amounts of data that can be updated very frequently. However, the high cost associated with frequent updating and continuous access has resulted in a relatively high price for the data. The basic costs incurred in delivering and accessing information via the Web are represented by telecommunications charges, which cover data transfer over local lines and networks, and by connect-time charges, which cover the costs of networked servers, maintaining the databases, and royalty payments to full-text providers. In order to serve many users accessing a database simultaneously, these vendors require large,
costly computer systems that need to be operated 24 hours a day, 365 days a year, with enough capacity to comfortably handle service peaks. The cost of operating and maintaining these data centers is passed on to our libraries.

In comparison, the cost of delivering and accessing information by CD-ROM is different. This involves a variety of development, hardware, and manufacturing costs, such as the cost of data preparation and mail delivery costs. There is also a hardware component (the CD-ROM drive) cost; however, this may only add about $50 to the cost of a PC. Therefore, this type of access can be much cheaper than Web-based access, as long as you have a CD-ROM network in place and do not need to buy additional servers or towers to offer or maintain access service.

3. OTHER OPTIONS OR NEW DEVELOPMENTS IN ACCESS TECHNOLOGIES

As you can see, the choice between CD-ROM and Web-based access to external databases does not have to be mutually exclusive. Each has its advantages and disadvantages, and as you read articles about these technologies, you will see that each is constantly evolving, as is the technologies surrounding them. One fairly new development in access technology that can help us is housing our CDs on a local server with all the advantages of remote web-based databases.

3.1 Web-based Databases Housed Locally

It is possible for libraries to have both the advantages of Web-based databases and in-house CD-ROMs if they have the expertise and financial resources. A few database vendors, such as Ovid and SilverPlatter, offer this option to libraries. With this option, CD-ROM serves as a medium to carry data between vendors and libraries. Libraries can cache the data from CD-ROMs to hard drives and speed up access time. With the search software provided by vendors, libraries can set up their own Web servers and provide access to their in-house, Web-based databases. This option also provides libraries with flexibility for customization and control. When the ISP has a problem and the Internet is down, those in-house, Web-based databases are still accessible. The maintenance cost is not high in comparison with that of CD-ROM databases. In this case, libraries only maintain one server (or servers), and there is no maintenance cost for clients. And an important note: systems administrators don’t need to worry about proprietary software installation and compatibility issues using this method.

3.2 DVD

One of the latest developments in network mass storage technology is DVD-ROM. When it first came out, it stood for Digital Versatile Disc Read Only Memory. The words have changed, but the DVD acronym has remained, although it has dropped the "ROM" extension. DVD has a much bigger capacity than a regular CD-ROM. While a CD-ROM can hold approximately 650 MB of data, a single-sided, single-layer DVD can hold up to 4.7 GB of data, a double-sided, single-layer DVD can hold 8.8 GB of data, and a double sided, double layer DVD can hold 17 GB of data. In addition to increased mass storage, DVD drives are backward compatible and can accommodate CD-ROMs. Even though DVD drives cost more than CD-ROM drives, it pays to buy a DVD drive so that one can use both formats. We think everyone will agree that the invention of DVD is certainly a big step forward in network mass storage technology.
Librarians are aware of the capacity of new DVDs and are taking DVD technology into consideration when they make new purchases. Unfortunately most library database providers are slow in adopting this new technology. Most of the databases are available on CD-ROMs, but not on DVDs. Unless database vendors promote the use of this new technology, it will not be used widely in libraries. We believe, though, that DVD technology will eventually become popular in libraries. It may not replace CD-ROMs totally, but undoubtedly this new technology has a lot of advantages over CD-ROMs.

3.3 Server-based Computing

Another development that may help to overcome some of the limitations of CD-ROM applications is the revived interest and growing popularity in terminal services and thin-client technology, also referred to as server-based computing. As we mentioned in previous sections, one of the major setbacks of CD-ROM technology is its limitation in open access. While many CD-ROMs are still DOS based and can only run on standalone computers, most of CD-ROMs can be networked on a LAN. A few leading CD-storage vendors began to provide solutions to provide databases over the Intranet such as CD Anywhere 100 Server from Cutting Edge Bridgette, Inc. (Ref. 5) and DiscZerver from Microtest, Inc. So far it is still a technical challenge to provide CD-ROMs over the Internet without local installation and maintenance. As all the CD-ROM databases require individualized access and proprietary searchware, it is almost impossible to provide a uniform user interface, such as a Web browser, for remote access. A technology that may help to bring CD-ROM databases to the Web is server-based computing.

With server-based computing, applications are executed and managed completely on a server. A client can be a dumb terminal or a PC with full processing capabilities. One example of server-based computing is Citrix's Metaframe 1.8 for Windows NT 2000 with terminal services (Ref. 3 & 4). Windows NT 2000 with terminal services allows multiple users to logon and run applications in separate, protected sessions on a single server. Metaframe is installed on Windows NT 2000 to provide resources in a transparent, bandwidth-independent way. The Citrix ICA client allows users to connect to Metaframe servers. The Citrix Windows Web Client enables Netscape or Internet Explorer to access CD-ROMs and other applications over the Internet. Even though the client is easy to install as a plug-in or help program, users still have to download and install it. It is not a perfect solution to provide CD-ROM applications, yet it does provide great help in overcoming some limitations imposed by the proprietary nature of CD-ROM databases.

Server-based computing may also provide a consistent user interface and eliminate most of the maintenance on clients. Usually it takes a lot of labor and skill to maintain each CD-ROM database and make sure they can run side by side on the same server and workstations. It requires more system resources to run many CD-ROM programs on a computer than Web databases. To access Web databases, all a systems administrator has to do is to provide Web links. The only requirement for Web databases is a browser. As hardware and software are not housed locally, there is no maintenance on clients. With server-based computing, system requirements for clients are the same as those for Web-access. Instead of running many different programs for CD-access, only a Web browser is needed. Systems work is mostly done on the server, and it still has all the advantages of local customization.
3.4 Caching Software and Hardware

Other new developments in CD-ROM mass storage technology involve many types of hardware and software especially designed to increase the access to information on CD-ROMs. For instance, there are CD towers with servers that you can plug into the network and have instant access to CD-ROMs. There is no need to configure or install the server operating systems. As the cost for hard disks has dropped considerably in the last several years, it is now a common practice to cache CD-ROMs to the hard drive for increased speed in access. Technically the databases are not administered through CD-ROMs but, instead, through hard drives. There is even software that can make any workstation on a network become a CD server. This option has certainly made it easier for us to run CD-ROMs on a LAN today, compared to five years ago.

4. CONCLUSION

As computing technologies continue to rapidly develop, the line between the Web and CD-ROM database access is blurring. In some cases, CD-ROMs are just used as a medium to pass information between vendors and libraries. Sometimes, it is even hard to tell if the databases are on CD or the Web. For instance, as we mentioned earlier, SilverPlatter's ERL databases are subscriptions on CD, but libraries can cache the databases from CD-ROMs to a server's hard drive. Those databases then become Web-accessible databases through software provided by the vendor. And with terminal services and thin-client technology, other CD-ROM databases can become Web-based databases when made available over the Internet. Our options will continue to grow with time.

CD-ROM as a storage medium will probably not die out totally, but we believe that its use in accessing external databases will continually be reduced as Web-access continues to increase in popularity. Unless CD-ROM technology can catch up with and integrate with Web technology, CD-ROM as a database carrier will continue to face serious challenges. Right now, though, we recommend that libraries, if possible, split up all the resources between in-house and remote locations. Then, if the ISP has problems and the Internet is down, libraries should still have enough resources to successfully serve their users.

5. NOTES

1. VALE is the Virtual Academic Library Environment, a consortium of 45+ New Jersey academic libraries, formed to help meet the demands for access to scholarly materials. VALE was funded by New Jersey Higher Education Technology Infrastructure Bond funds, matching institutional dollars, and a New Jersey State Library grant.

6. REFERENCES


