Communications

Providing Access to Foreign Language Electronic Resources

Eileen Llona, Emalee Craft, Keiko Yokota-Carter, and David Pham

It may be commonly thought that CD-ROMs are a dead technology, but for many information resources produced outside of the United States, CD-ROMS may offer the only method of access to bibliographic and full-text data in non-English languages. For this discussion, CD-ROM technology refers to programs that provide bibliographic or full-text access on removable media, usually using proprietary search and interface software. Realistically, this can include DVD technology as well as older programs from floppy disks. Most products, however, presently come in CD-ROM format.

In the United States, more and more electronic resources are being made available through Web interfaces and subscriptions; by contrast, electronic resources in other countries are very often limited to CD-ROM form.[1] Electronic resources include both bibliographic databases as well as full-text digitized resources. They often include proprietary software for searching and displaying the data. Foreign-language products are written in the language of the source country, most often using localized code pages rather than Unicode. The value of these resources is often high for researchers, as they augment or replace traditional print resources. Examples include digitized newspapers from Japan, census statistics from China, digitized versions of Islamic scriptures, and election data from Russia. In the case of Japanese electronic resources, although some online full-text databases are available, CD-ROM is still the major format. Archives of newspapers, journals, and reports have been published in CD-ROMs, replacing hard copy reprints of these materials. In 2002, two major newspaper companies in Japan started publishing their archival newspaper in digital format. The whole set of one newspaper, *Yomiuri Shinbun,* consists of 126 CD-ROM disks.

For researchers in the United States, these resources are considered quite important in humanities and social-sciences research, and often represent the only feasible means for performing research on these cultures. Since primary materials are usually held by a limited number of libraries in the United States, traveling to the subject countries may be the only other alternative for researchers to do their work. Demand may be relatively low, but many resources are selective and offer valuable information or representation of cultural materials from other world regions. Since the frequency of use of these foreign-language materials outside the subject country is quite low, the cost per user is much higher than the cost of a more widely used resource such as the *New York Times* database. Not many research institutions can afford the luxury of these materials. However, the importance of holding the material is also highly valued when the alternative is to send researchers to travel to the source countries to use this material.

When obtaining resources from other countries, collection development librarians in the United States sometimes have a choice that includes print and electronic versions of a resource. While some foreign vendors publish databases on the Internet, the price for a subscription is often beyond what academic libraries can afford. Also, resources are often in the form of a one-time product representing an entire resource, bundled with retrieval and interface software (such as the *Yomiuri Shinbun* digitized newspapers). Since it is the content that is most valuable, format becomes the second choice. CD-ROM technology is still heavily used for delivery of specialized software and products.[2] The choice, therefore, is often between a print resource (such as an index or full-text resource) and its digital equivalent, if available. Since a digital form often includes superior searching methods and allows resources to be shared more readily, this format is often chosen. Cost and licensing issues must be considered, as either or both of these elements can sway a decision to stick with print, or not obtain the resource at all. If an electronic version is obtained, it may be difficult to provide easy access to the user. Programs are usually written using proprietary software and require specialized language settings. A program must be installed to view the data, so the library must make a commitment to installing and maintaining these programs and workstations. Realistically, while some of these resources do provide valuable information that is unavailable in other formats, access to that information remains a problem.

Licensing before Technology: Negotiating with Foreign Publishers

Another issue that plays a large role in obtaining electronic resources is licensing restrictions. In the case of Japanese vendors, the Japanese government has placed regulations on

Eileen Llona (ellona@u.washington.edu) is International Studies Computer Services Librarian, Digital Initiatives; **Emalee Craft** (Emalee.Craft@asu.edu) is Information Technology Coordinator, Information Technology Services; **Keiko Yokota-Carter** (kyokotac@u.washington.edu) is Japanese Studies Librarian, East Asia Library; and **David Pham** (dpham@u.washington.edu) is Senior Computer Specialist, Information Technology Services at University of Washington Libraries.

licensing agreements to protect consumers from damaged products and unethical conduct of companies. Therefore, more Japanese companies have set up legal departments and increased their sensitivity regarding the sale of their products. Licensing issues attached to electronic resources are sometimes impediments to the purchase of or subscription to Japanese (or other language) electronic resources. Universities may manage budgets to acquire materials by applying for grants, asking for donations, or even dividing payments over a period of years. However, negotiation of license issues may not be solved on an individual basis, because it deals with legal issues involving other countries.

One of the common legal issues for libraries is license agreements between a university and foreign vendors. Problems may arise over where jurisdiction lies and who is responsible for legal fees in case of legal disputes. Each provider and subscriber specifies its own jurisdiction for legal settlement. For example, the University of Washington (UW) specifies *King County, Washington State;* Japanese companies specify *Tokyo, Japan.* In addition, Japanese companies require all legal fees associated with disputes to be paid by subscribers. Usually, the institutional policies of American universities do not allow representatives to sign agreements with a foreign jurisdiction or that oblige the university to pay such legal fees, making it difficult to obtain subscriptions.

In *Yomiuri's* case, the vendor refused to negotiate on treatment of a legal clause at the initial stage of negotiation. When communications through international telephone calls, faxes, and e-mails failed, the Japanese-collection librarian, a native speaker of Japanese, flew to Japan on a special mission on behalf of the director of the UW libraries. The librarian met the representatives of *Yomiuri* newspaper and negotiated with them personally, and was able

to persuade them to delete the clause with a promise of maintaining a cordial, professional relationship with their business partners.

This face-to-face meeting was a breakthrough. The Japanese company, of course, conducts its business based upon business calculation, but before real negotiation could be established, it was essential that a high-ranking representative of the university express—in their native Japanese—the respect and faith necessary to build a trustworthy business relationship with them.

In this sense, working with foreign electronic resources not only involves technology issues, but also legal and cultural issues. Both information-technology staff and collection-development librarians offer diverse knowledge and experience that contribute in different ways towards providing access to these resources. Knowledge of the social-cultural situation of a provider's country, as well as technology issues, is necessary in bringing these resources to American researchers. Such team building between systems librarians and subject librarians is crucial to improve access to foreign electronic resources.

Common Practices Used for Providing Access to Foreign Language CD-ROM Resources in Libraries

Once the decision has been made to obtain a CD-ROM resource, libraries must provide access. Often this involves nothing more than making the disk(s) available for circulation. Often, however, these resources are quite expensive, and librarians are reluctant to circulate the materials. In such cases, libraries have to set up equipment for users to work with these resources. In the case of foreign-language media, this often involves

language settings, and sometimes language-specific versions of operating systems (the Japanese version of Windows 98, for example). Individual workstations may be set up with the required language settings, often with separate workstations for each language. Obviously, this becomes unmanageable if every language is to be supported.

Foreign-Language CD-ROM Resource Access at the UW Libraries

The use of foreign-language CD-ROM resources is mainly at the East Asia Library at UW. There are about 200 CD-ROM resources in three languages (Japanese, Korean, and Chinese). In addition, Chinese resources may be in either traditional Chinese script, or simplified Chinese script. These require different language settings. All of the resources will run on some variant of the Microsoft Windows operating system.

At the UW–Seattle campus, library public PCs are managed with a one-size-fits-all approach across the seventeen Seattle-campus libraries. Part of the public PC management is a tight security model at the local desktop. While the strong security provides a stable platform for public use, as well as access control to meet licensing requirements, these security settings create configuration problems for programs requiring customized settings (such as language settings and proprietary software). In order to provide access to East Asian-language CD-ROM resources, workstations must be configured for language-specific code pages, keyboards, and font settings. In addition, much of the proprietary software conflicts with local security-policy settings. Thus, setup of local workstations and programs often requires tedious troubleshooting and configuration.

To provide access to these non-networked resources, four public workstations have been modified to accommodate language-specific settings and locally installed CD-ROM programs. This has required extensive troubleshooting due to the proprietary nature of the programs conflicting with the workstation security settings. When the public workstation model is updated, locally mounted software must be reinstalled. Obviously, this model is difficult to sustain over time.

Consider, for example, the problems involved in providing access to the *Yomiuri Shinbun* set of digitized Japanese newspapers. The Japanese-studies librarian purchased this multivolume set, which comes with a proprietary search engine and represents a very important collection for Japanese-studies researchers. Manual research methods using print and microfilm of these newspapers require many hours. The electronic version provides good search capabilities and instant access to full-page images of the newspapers. Due to the size of this valuable and expensive resource, the option of local installation on a public workstation is not feasible.

The alternative to installing proprietary CD-ROM resources on local public machines is to lend the materials to users. Problems with this solution include potential loss of materials, and requirements that users have the machines and settings needed to run the software. With a set as large as the *Yomirui Shinbun* product, circulation is not possible due to item costs and security issues.

Since the current alternatives do not lead to sustainable access, other methods of providing access to these resources have been explored. Networking is the obvious choice, with Internet-enabled access being most ideal. A review of networking options is provided by Kumar;[3] newer technologies can also be utilized to provide networking of for-eign-language resources, as has been done at the East Asia Library.

Networking of Foreign-Language CD-ROM Resources at the UW Libraries

English-language CD-ROM resources have been made available over the UW libraries network for several years. CD-ROM resources at UW have evolved in several stages. In 1999, UW replaced its locally developed X-Windows-based Willow information system with its Information Gateway (the Gateway).[4] During the employment of Willow as the libraries' primary information search interface, CD-ROMs were not generally offered as networked resources. With the inception of the Gateway, however, access to electronic resources of varied types was integrated through the use of a digital-resource registry.[5] From the time of the Gateway's inception until early 2003, CD-ROM-based products, mostly English-language resources, were offered online through a hybrid implementation of Windows-based technologies. The CD-ROM images were copied in full to a Windows NT-based network. The interface for CD-ROMs was set up on client machines located in the libraries' public areas. Any proprietary software that came as part of a CD-ROM was loaded onto a central public image and then distributed via image casting to hundreds of public PCs.

Most CD-ROM resources, and all other electronic resources, were organized through the Gateway. Through a series of HTML-embedded scripts, a user connected to a CD-ROM application by launching it from the Web-based Gateway. These scripts would open onto the desktop with locally installed application software and appear to the user to be very similar to using a CD in the local drive. The scripts, along with net-work share limits, also were employed to insure that individual licensing agreements were met.

Although this system took advantage of the technology of the time and allowed centralized management of hundreds of electronic resources, it had serious drawbacks supporting CD-ROM resources. It was difficult or impossible to update any local software once the public image was prepared and deployed, and most notably for international studies resources, almost no support was possible for non-roman-character-based (NRCB) programs.

Beginning in 2002, the libraries began to investigate other technologies to support non-Internet-based electronic resources, including NRCB products. Starting with a system for primarily English-language items, the libraries set up a Windows 2000 server with Terminal Services in Application Mode and went live on public workstations in early 2003. This most recent networked setup is similar to versions used in libraries since the mid-1990s when Microsoft licensed CITRIX and developed the technology further.[6] Preparing for Terminal Services included installing Windows 2000 Advanced Server with Terminal Server set to Application Mode.

Application Mode means that server computing resources and licensing are optimized to allow multiple simultaneous client connections. Various authentication and resource-specific configurations can be set on the server side, allowing for security settings without interfering with program-specific requirements. The most important contribution of Terminal Services in providing access to foreign-language electronic resources is the ability to set all of the language configurations at the server, once. Fonts, input methods, and code pages are set once on the server, and delivered to the user via thin client architecture. Thus, the user's machine does not need to have

language-specific settings or software. Maintenance on public workstations is significantly reduced, and security is improved. Most importantly, access to these resources is vastly improved.

The Gateway continues to be the primary interface for English-language CD-ROM resources. However, instead of launching in the local public PC desktop with locally installed software, links to CD-ROM-based resources launch a universal ActiveX Terminal Services client from within Internet Explorer, so that users with an Internet connection can access networked resources after authenticating. Because network domain permissions can be employed with the Terminal Services client, domain groups make it easier to manage resource-licensing restrictions.

While Terminal Services is a great advance in providing anytime-anywhere access to foreign-language CD-ROM resources, there are constraints. Since most non-English resources are still produced using language-specific code pages (not Unicode), switching from one language to another is not yet possible without rebooting the server. Hence, language-specific servers must be set up and maintained. At the UW libraries, one Japanese server running Terminal Services has been set up, and a simplified Chinese server currently is being configured as well. While machine costs are relatively low, ongoing maintenance and labor remains a significant consideration in deploying Terminal Services for several languages. At the present, there are no immediate solutions to this problem. While the Unicode character set is available, there are few non-English resources that utilize this character set. Since code pages are loaded at boot time in the Windows environment, switching between languages and code pages requires a reboot. It is hoped that future implementations of the Windows operating system will provide for simultaneous-code page support.

Conclusions

While many electronic resources are being delivered via Internet subscriptions, libraries are still faced with providing access to CD-ROM resources, particularly for foreign-language-based materials. Thin client technology such as Terminal Services goes part of the way to improving access to these one-off resources. However, providing access to resources in multiple languages still presents problems, and not all languages can be served equally. Libraries must make decisions, both at the collection development and the information-technology levels, on which languages to support and how to support them. As Unicode becomes more widespread, it is hoped that language-specific programs will utilize this technology, thus making language-specific requirements less cumbersome.

Migration to Internet-based resources continues in the United States, and is becoming more common in other countries. However, issues of preservation and access will remain cloudy for this format for some time to come. If subscription prices become too high, or publishers go out of business, what becomes of the resource? On the other hand, the proprietary nature of programs and data made available on removable media such as CD-ROMs presents its own set of sustainability questions. In five or ten years, will operating systems be available to read currently produced programs? What commitments do libraries need to make to keep legacy equipment and software available for such resources?

In the foreseeable future, it appears likely that libraries will have to continue to support standalone resources that represent unique data sets and cultural-heritage data.

Business models that provide instant Internet access are not feasible for many smaller companies and countries. Delivery mechanisms for providing such data are most likely to continue to be through proprietary programs and software that can be packaged and delivered on a one-time basis. In order for libraries to provide access to these often rare, unique data sets, some infrastructure should be in place for making such access sustainable.

Acknowledgments

References

1. Ellen H. Hammond, "Japanese Electronic Database Use in North America: Widespread and on the Web: Report of the Consortial Licensing Task Force, North American Coordinating Council on Japanese Library Resources (NCC)." Accessed November 13, 2003, www.library.pitt.edu/ncc/CLTFSurvey2000.pdf.

2. Marshall Breeding, "Does the Web Spell Doom for CD and DVD?" *Computers in Libraries* 19, no. 10 (1999): 70–75; Barbara Burg and Amy M. Kautzmann, "Web versus CD-ROM: Format Follows Function," *Searcher* 6, no. 1 (1998): 59–63.

3. K. Satish Kumar and P. O. Jayaraman, "Networking of CD/DVD ROMS: An Overview," *DESIDOC Bulletin of Information Technology* 22, no. 6 (2002): 3–12.

4. Debra S. Ketchell et al., "Willow: A Uniform Search," *Journal of the American Medical Informatics Association* 3, no. 1 (1996): 27–37.

5. William Jordan, "My Gateway at the University of Washington Libraries," *Information Technology and Libraries* 19, no. 4 (2000): 180–86.

6. F. Bartow Culp, "Streamlining Electronic Access with Windows Terminal Server," *Library Hi Tech* 20, no. 3 (2002): 302–7.