

Digital Libraries and Electronic Publishing  
series editor: William Y. Arms

Digital Libraries, William Y. Arms

*From Gutenberg to the Global Information Infrastructure: Access to  
Information in the Networked World*, Christine L. Borgman

---

**From Gutenberg to the Global Information  
Infrastructure**  
Access to Information in the Networked  
World

Christine L. Borgman

The MIT Press  
Cambridge, Massachusetts  
London, England  
PROPERTY OF RUTGERS UNIV.  
PAUL ROBERTSON LIBRARY

The fourth trend identified is the transition from a focus on the individual user of digital libraries to the group processes involved in information-related behavior. Research must account for the social context of digital library use, must consider ways in which people use computers as a medium for collaboration, and must consider ways in which computers can assist the collaboration process. To the extent possible, digital libraries should be tailored to communities of users while also being capable of spanning the boundaries between communities. The goal is to make better use of resources and to improve knowledge discovery. Focusing on groups involves spanning the boundaries between applications; thus, it presents new challenges in drawing the boundaries between individual digital libraries, a global digital library, and a global information infrastructure.

Several lessons for the future of digital library design have been offered. One, reiterated throughout the chapter, is the fundamental challenge of balancing the need for tailoring digital libraries to a community of users with the need to construct a vast, interoperable, global digital library. Another is the reflexive relationship between human behavior and technology. It is difficult, and probably impossible, to separate the influence of one on the other. People use the technologies that are available and adapt them to suit their purposes. Subsequent iterations of technology reflect those adaptations in combination with new capabilities made possible by technological advances. In sum, technology and behavior evolve together.

The final conclusion is that usability is a moving target in the design of digital libraries. As the technology improves sufficiently to overcome earlier problems, it presents new opportunities that bring new design challenges. People have higher expectations with each technical advance and with each improvement in knowledge organization, in user interface design, or in system capabilities. As the research community addresses the challenges posed here, new challenges and new research questions will arise. Thus, access to information is an evolving concept.

## 7

## Whither, or Wither, Libraries?

When libraries first began automating their internal processes, in the 1960s and the 1970s, librarians often were asked "Why does a library need a computer?" Most people were quickly persuaded of the need when given an explanation of libraries' activities in managing information. By now, the majority of libraries in developed countries provide information resources in electronic as well as physical forms and rely on computing technology and networks for their administrative operations.

Now the reverse question is being asked: "We have the Internet, the World Wide Web, and digital libraries, so why do we still need libraries?" Sometimes the question is explicit, as when organizations plan new campuses, buildings, or agencies, or when international development agencies allocate funds for telecommunications, computing, and library services. At other times it is implicit, as when government officials, parents, or volunteers invest in computers and network access for schools that may have no libraries (Borgman 1997c).

The question "Why do we need libraries?" suggests a dichotomous choice between libraries and computer networks. The challenge for the information age is not choosing between libraries and computer networks, however. Rather, the challenge is determining how best to provide access to information, and how best to support the marketplace for ideas. Libraries in democratic societies are part of social strategy "to promote the Progress of Sciences and useful Arts," to quote the US Constitution (article I, section 8). Public access to knowledge is necessary for an informed electorate, and to promote learning and invention (Buckland 1988; Lyman 1996). These cultural values are more permanent than the media of communication, whether print or electronic (Foskett 1984). Yet the distinctive

features of the American public library system in particular are under threat from the technology and policy transitions currently underway. James Billington, the present Librarian of Congress, identifies four features currently under threat (Billington 1996, p. 36–37):

First, for democracy to be sustainable on a continental scale in a multicultural society, it must be based on the dynamic use of knowledge. . . . Second, this knowledge must be *openly accessible* to all people. . . . Third, public libraries expressed the *growing pluralism* of American society. . . . Fourth, public libraries were, nevertheless, a *unifying force* in the communities where they existed.

The social goals for libraries vary among democratic societies and between democratic and other social systems. They also vary among types of libraries, such as public, academic, special, and school, although libraries of all types cooperate with each other in the context of broader social goals. The means for implementing these goals were developed for a print culture, which has been the dominant form of communication during the period of modern democratic political systems. The present challenge is to support the social goals of democratic societies in a new technological environment for communication. Digital libraries have tremendous potential to enhance access to knowledge, to support learning and education, to promote progress in the “sciences and useful arts,” and to inform the citizenry. At the same time, many scholars (e.g., Billington 1996; Dervin 1994; Lievrouw 1994a,b; Lyman 1996) are concerned whether the goals of free public access to information can be sustained while encouraging a robust market for information in economies where knowledge is a form of capital. The marketplace for ideas in a democratic society is characterized by a balance between the rights of the citizenry for the broadest possible access to information and the rights of creators of information to be compensated for their work.

#### Organization and Access in a Networked World

Organization of content is a defining characteristic of libraries, digital or otherwise. Information must be described, represented, and organized in ways that allow people to discover its existence, locate it, and obtain it in a useful and usable form. Documents or document-like objects should be described uniquely, so that their identity and authenticity can be verified. The component of network architecture that supports organization and

access is metadata. Metadata serve a variety of functions in networked environments, including description, administration, intellectual access, intellectual organization, technical specifications, and preservation management.

Libraries have refined the theories, principles, and practices for creating metadata and for accomplishing these functions throughout the twentieth century. American libraries began sharing metadata for cataloging records around 1900. These practices spread around the world, and by the late 1960s international standards for sharing metadata in computer-readable form had been established and implemented. By now, libraries have more experience with creating, maintaining, and distributing metadata than any of the other links in the chain of producing and delivering information. They also have addressed a wide array of portability, interoperability, and scaling problems that others are just now encountering.

#### Library Principles and Practices

Among the core responsibilities of libraries are to select, preserve, and provide access to published materials on behalf of their user communities. This is usually a shared responsibility, as few individual libraries can acquire all materials of potential interest to their users. At a national level, one organization usually is responsible for preserving the published record of the country. Legislation supports the preservation of cultural heritage, typically requiring each publisher to provide a specified number of copies of every publication to a designated organization, which is usually the national library. In the United States, legal deposit laws ensure that the Library of Congress receives copies of US publications. In the United Kingdom, comparable laws require that British publications are deposited in the British Library. French publications go to the Bibliothèque Nationale de France, German publications go to the Deutsche Bibliothek, Hungarian publications to the National Szechenyi Library, and so on. Deposit libraries normally are responsible for cataloging these materials and creating the authoritative metadata for each document received. Collectively, the metadata they produce become the national bibliography of the country. Those metadata, in turn, become source material for other libraries that acquire the materials.

Throughout the world, legal deposit laws have been fairly effective in ensuring that printed books from major publishers are preserved in national

collections. Acquiring books from new publishers, small publishers, or self-publishers who are not aware of their legal responsibilities always has been more problematic. Acquiring serial publications, pamphlets, and ephemeral materials that may be important parts of the national cultural heritage is yet more difficult. Electronic publications raise a host of new problems, not the least of which is determining what is a publication. Identifying what exists, who is responsible for it, and whether and how it should be acquired and preserved are among the other challenges to maintaining national collections and national resources of metadata.

Within each country, academic libraries and public and other government-supported libraries typically function as a collective entity to provide access to the materials held within their countries. In principle, the residents of a country should be able to use most materials held by any participating library in their country. Special libraries within private organizations also rely on these resources and may contribute as well. National cooperation, in turn, relies on international cooperation for sharing resources.

#### *Managing Metadata*

Metadata are essential to managing library collections and operations. Libraries use many types of metadata in their operations and services. Of these, cataloging data are the primary form of metadata used to manage local collections and to discover, locate, and obtain materials held elsewhere. Principles and practices for creating and managing cataloging metadata have evolved throughout the history of print publications and continue to evolve to support new media.

Cataloging is one of the most expensive operations in libraries: each item must be described physically, so that it can be identified uniquely, and intellectually, establishing responsibility for the work. Subject terms or classification are assigned so that works can be identified by topic. Early in the twentieth century, the Library of Congress (LC) began distributing cataloging records as a means to subsidize the cost of cataloging for the nation's libraries (Billington 1996). In the United States, the LC record is considered the most authoritative metadata for a work. Libraries can use the LC record, or a record from another authoritative source, rather than creating a new record for its own copy of that work. Further, if all libraries use the same catalog record, then the work will be described consistently

across libraries, making it easier for readers to identify and locate. Substantial efficiencies and control of internal costs have been gained by sharing cataloging data (Cummings 1986; Becker, Dodson, and Yoakam 1984; Maciuszko 1984).

With the advent of online systems, it became possible to exchange records directly and immediately among many libraries, rather than the one-way flow from a central distribution point to individual libraries. Metadata from national bibliographies and other authoritative sources form the core of online shared cataloging systems. Participating libraries search the database to determine if a catalog record exists for a new document they have acquired. If so, the record is captured for local use; if not, the library catalogs the document and contributes the record to the database, where it is available for other participating libraries. Records in the shared database are extensible. Libraries can add data for their local catalog about their copy (or copies), such as the physical location of the item or extra access points (names, subjects, etc.) considered important for local uses.

The first major online shared cataloging system was OCLC (then Ohio College Library Center, now the OCLC Online Computer Library Center), established in 1967 (Becker, Dodson, and Yoakam 1984; Maciuszko 1984). Others quickly followed, such as RLIN (Research Libraries Information Network) and WLN (Western Library Network) in the United States, BLCMP (Birmingham Libraries Cooperative Mechanisation Project) and CURL (Consortium of University Research Libraries) in the United Kingdom, and PICA (Project on Integrated Catalogue Automation) in the Netherlands (Tedd 1993). In Europe, at least 30 cooperative networks existed in 14 countries by the early 1990s, offering a variety of services including online shared cataloging (Dempsey 1990; Hein 1991; Holley 1993). The European Union promotes multinational library cooperation, including the development of shared databases of major European collections (Day, Heery, and Powell 1999; Dempsey 1990; Rau 1990).

Shared cataloging systems, also known as bibliographic utilities, have become very large operations. OCLC alone serves more than 27,000 libraries in 64 countries, and its database contains about 40 million bibliographic records (OCLC 1998a) in more than 400 languages (OCLC 1998b). Analyses of library cataloging practices indicate that the vast majority of catalog records in the United States and the United Kingdom are

obtained from these online shared cataloging systems. Records for materials in English are still more widely available than those for other languages, but the breadth of these databases is expanding rapidly as more countries contribute records (Chapman 1999; Leazer and Rohdy 1994). Thus, metadata exist in standardized, computer-readable form for a growing proportion of extant published materials.

#### **Rules and Standards**

Rules and standards are essential for managing metadata in networked environments. The establishment of standards for exchanging bibliographic data in computer-readable form in the 1960s led directly to the development of the shared cataloging utilities (Avram 1972, 1975, 1976; Buchinski et al. 1978; Library of Congress Network Advisory Group 1977a, 1977b; RECON Working Task Force, 1969, 1973). American and British libraries now have nearly 40 years of experience exchanging metadata in computer-readable form, based on international standards and rules. The Library of Congress began work toward the MARC (Machine-Readable Cataloging) format in 1961, conducted a pilot project from 1966 to 1968, and began the MARC II distribution service in 1968. The British National Bibliography and the Library of Congress began work toward a UKMARC pilot project during the LC MARC pilot study, leading to the international exchange of records between the United States and the United Kingdom beginning in 1969 (Avram 1975). Also in 1969, the International Federation of Library Associations and Institutions (IFLA) convened an international meeting of cataloging experts that led to the establishment of the International Standard Bibliographic Description (ISBD) (Avram 1975; Gorman 1969). As other countries began to establish their own MARC formats, IFLA developed UNIMARC (UNiversal MARC) in the late 1970s to assist in international cooperation and data exchange between MARC formats (McCallum 1989).

Underlying these metadata standards are rules for how to describe and represent published materials. The Anglo-American Cataloging Rules (AACR), USMARC, and UKMARC were developed in concert in the late 1960s and are revised continually to accommodate changes in technology and in practices. AACR now supports the description and representation of various media, from videotapes to CD-ROMs, that did not exist in the

1960s. Despite common rules, national practices do vary, which is why AACR is mapped to multiple MARC formats (e.g., USMARC, UKMARC, CANMARC). AACR has been translated into multiple languages, with some adaptation to local practices. UNIMARC is used widely in Europe and elsewhere, especially with cataloging rules other than AACR. Various international efforts are underway to achieve convergence, or at least bridging mechanisms, among the multiple MARC formats. IFLA and the European Union are leading many of these efforts, in coordination with national libraries. The existence of multiple rules and formats is an example of the tradeoffs between tailoring to local practices and the interoperability of systems and services.

#### **Catalogs and Content**

A by-product of the sharing of cataloging records is the creation of union catalogs that reflect the collective holdings of participating libraries. OCLC alone has records on the existence and location of more than 660 million copies of 40 million distinct works that have been cataloged through the system (OCLC 1998a). Albeit the largest, OCLC is only one of many shared cataloging systems. Thus, these systems represent a significant step toward creating a global digital library of metadata indicating what information resources exist and where they are located.

Digital libraries of metadata facilitate discovery and location processes. They also can assist people in gaining access to the full content. Toward this end, shared cataloging systems provide services that execute resource-sharing agreements among participating libraries. Public and academic libraries in most countries lend materials to other libraries, and in return, they borrow materials from other participating libraries on behalf of their user communities. Interlibrary lending agreements often are supported by laws and by supplemental funds to redress imbalances between lending and borrowing libraries.<sup>1</sup> In this way, all the participating libraries in a country may function as a "virtual library collection" available to its citizenry.

The virtual library collection is accomplished by a variety of means. In some cases, library staff search databases such as OCLC or the British

1. In the context of librarianship, the words "loan" and "lend" are often used interchangeably. American libraries typically use the term "interlibrary loan"; British libraries and IFLA use "interlibrary lending."

National Lending Library at Boston Spa to determine the existence and availability of materials. These systems process the requests and submit them to the lending libraries, which deliver the materials to the requesting library. Materials such as books are usually sent by post, whereas journal articles may be sent by fax. The requesting library, in turn, lends the material to the requesting user. Materials delivered by fax or email can be provided directly to the requesting user. As more components of this process are automated, the number of steps decreases, the amount of handling increases, costs drop, and delivery time shortens.

Commercial document supply services, for example, employ automation effectively to build upon existing resource sharing practices. They broker agreements with libraries to supply copies of articles from journals to which those libraries subscribe. The document suppliers' database of available journal articles is drawn from the online catalogs of participating libraries. Users request documents from the service, provide a fax number, email address, or other delivery address, and pay a fee. The service forwards the order to an owning library, which supplies the material directly to the requester. Fees paid are divided between the library for its subscription and staff costs, the publisher for copyright permission, and the document supply service for brokering the transaction. Libraries increasing rely on these sources as a form of outsourcing of interlibrary lending requests.

At present, electronic exchanges of materials are feasible only for short documents such as journal articles that are easily digitized, and only in parts of the world where telecommunications and computer networks are sufficiently reliable and available. Physical materials such as books and videos continue to be exchanged via postal mail owing to the expense of digitizing, limitations of bandwidth, and copyright policies.

### Generalized Principles and Practices

The library principles and practices outlined above are predicated on the above-mentioned social goals for access to information in democratic societies. National bibliographies, shared metadata, resource sharing agreements, and copyright laws, taken as a whole, are the underpinnings of a "global virtual library collection." The form in which virtual collections are realized varies considerably within and between countries; this discussion has focused on common features. Many of these principles and prac-

tices also can be found in nations with other forms of government. These practices are by now well tested, are supported by available technology such as automated library systems and computer networks, and by resources such as shared cataloging systems that are accessible without political prejudice. "Automated library systems" is shorthand for enterprise management systems that support and integrate library operations such as acquisition of materials, serials management, cataloging, online catalogs, circulation of materials, and associated accounting processes (Barry, Bilal, and Penniman 1998; Borgman 1997a; Tedd 1993).

These practices were initially designed for print publications. Many of the assumptions on which they were predicated, such as one work existing in multiple copies, or the ability to specify precisely whether a document is subject to deposit laws, do not apply to digital documents that exist in networked environments. The world of electronic publishing and distributed access to information is far less ordered than is the world of print publication. Yet virtual collections are needed now more than ever. The necessity for mechanisms to identify, locate, and obtain information resources is increasing with the proliferation of digital content in many forms, in many languages, and in many places. Ways are needed to balance the rights of readers with the rights of authors, regardless of the format of information resources or the means by which they are delivered.

Deposit laws ensure that documents are acquired when they are available so that they can be preserved for future generations. In short, these laws ensure the persistence of content (Day, Heery, and Powell 1999). The lack of persistence is among the greatest challenges of managing digital documents; they may exist only as long as someone maintains access to them, and even if so, they often lack a persistent identifier or location.

### Managing Metadata

Electronic documents rarely follow the "one work, many copies" rule. A work may exist in only one form and in only one place. The same work may exist in multiple versions, whether in one or many locations. Documents may exist in duplicate copies in multiple databases, such as digital libraries that are mirrored on multiple continents. If documents are unique, then only one copy of a metadata record is required, in which case shared cataloging is not an effective model. However, other services of

shared cataloging systems continue to be necessary, such as registering what exists, where, in what form, and the means to obtain access. The question then arises of the appropriate methods to accomplish these functions, whether through centralized systems and services or through searching mechanisms that create dynamic virtual collections on demand.

One option is to shift responsibility for creating metadata to different stages in the publication process. Responsibility for cataloging monographs has been moving "upstream" in the process for some years already. "Cataloging in publication" (CIP) programs, for example, enable publishers to submit prepublication copies to national centers to generate partial cataloging records that can be included in the print publication. New programs in "Electronic CIP" (ECIP) enable publishers to submit prepublication copies electronically and have the records returned electronically, which shortens the turnaround time (Davis-Brown and Williamson 1996; Day, Heery, and Powell 1999).

Authors, creators, publishers, and other producers of electronic content can generate their own metadata, using standards such as the Dublin Core. Various international projects are underway to experiment with new ways of assigning metadata so that information resources can be discovered, located, and obtained in networked environments (Day, Heery, and Powell 1999). Other approaches involve the search engines that sweep the World Wide Web. These engines catalog, to varying degrees, what exists online at any point in time (Kelly 1999).

Though the principles and practices associated with legal deposit and shared cataloging are far from perfect in capturing, maintaining, and preserving a nation's cultural heritage, they go a long way toward providing a virtual collection for its citizenry. Most of the approaches proposed or tested to date for organizing electronic materials on computer networks are piecemeal. They also are more effective at providing access to what exists today than to creating a permanent and comprehensive set of resources.

### **Rules and Standards**

Rules, standards, and practices continue to evolve with changes in technology, economics, and policy. The online digital world presents significant new challenges. One of these is the transition from cataloging physical

objects to cataloging digital objects. The rules were intended for "book in hand" cataloging: the assumption that the cataloger can handle the item and inspect it for certain attributes. The Anglo-American Cataloging Rules, for example, originally presumed that the cataloger could determine the physical dimensions of the object and the number of pages, and could identify author, title, publisher, place, and other data on the title page and verso. In a digital world, catalogers are creating metadata for objects they cannot handle, and are reliant on other means to identify attributes of those objects. Creating metadata for electronic documents is easier and faster than for print documents only if the documents are structured in a way that data can be extracted automatically. Often, creating metadata for digital documents is more labor intensive than for printed documents. Rather than spending a few moments handling the object to identify metadata elements, the cataloger may need to visit a web site. Or, software may need to be installed and run to determine descriptive elements such as author, title, publisher and date, much less additional elements such as technical specifications.

The forms and genres of electronic publication are not stable, and may never be. This is the blessing and the curse of digital documents. Rules, standards, and practices for managing metadata are becoming more generic to accommodate the variance. Digital documents are often unique, and no equivalent yet exists for unique and persistent identifiers such as the International Standard Book Number (ISBN) or International Standard Serial Number (ISSN). Uniform Resource Locators (URLs), the current addressing mechanism, are notoriously unstable. Other proposals such as Persistent URLs (PURLs), Uniform Resource Names (URNs), and Digital Object Identifiers (DOIs) have yet to be adopted widely (Kelly 1999).

A related challenge is to adapt techniques developed for static documents such as books to dynamic or mutable documents that are updated continuously (Bishop and Star 1996) and to documents that may include links or pointers to other documents. Cataloging rules do account for relationships between documents that may be manifested in multiple ways (e.g., variations in copies, different editions) and in multiple forms (e.g., text, video, film). Research on the relationships that exist between documents in a world of static materials, such as equivalence, derivative,



descriptive, whole-part, and shared authorship, are providing insights into the types of linking relationships needed in an online world (Leazer 1994; Smiraglia and Leazer 1999; Tillert 1991, 1992).

### *Catalogs and Content*

In practical terms, the shift from print to digital forms only eliminates the last step in the information access process: obtaining a physical copy. Metadata continue to be essential for discovering and locating information resources, regardless of form. With print materials, the last step is to remove the item from a library bookshelf or request it via interlibrary lending or a document-delivery service. With digital documents, the last step may be simply to click on the name of the item, initiating the transfer of a copy to one's local computer.

As can be seen from the above examples, however, the shift from sharing cataloging data to sharing content is more than a technical matter. In the print world, legal deposit laws, shared cataloging systems, and interlibrary lending agreements combine to provide public access to information. The primary means for maintaining a balance between the rights of the citizenry for access and the rights of creators to be compensated adequately is through copyright and other intellectual property laws. US copyright laws include the "doctrine of first sale," which enables libraries to lend the print materials they have purchased, for example. They also include tests of "fair use" that permit a small number of copies to be made for educational purposes without additional payment to the authors or publishers. Copyright laws in other countries draw the balance somewhat differently, and international agreements govern sale and use between nations.

Copyright features such as the doctrine of first sale and fair use are being challenged for digital materials. Many publishers prefer to lease rather than to sell electronic content, for example. Others are providing electronic content only under contracts in which libraries forfeit certain copyright privileges. These are among the threats to library services to which Billington and others refer (Battin and Hawkins 1998; Billington 1996; Hawkins 1998; Lyman 1996, 1998). Finding the appropriate balance between the rights of citizens and the rights of producers may be the greatest challenge we face in providing access to information in democratic societies.

### **Form and Function, Space and Place**

Libraries are more than collections and buildings. They exist in many forms, have many functions, and serve a variety of purposes, not the least of which is fulfilling social goals of their countries.

#### **An Institution or a Function?**

The simple definition of a library—an agency that selects, collects, organizes, preserves, conserves, and provides access to information on behalf of a user community—says little about how these activities are performed, the relative emphasis on each, or the relationship between them. Indeed, libraries come in so many types and sizes, encompass such a wide variety of activities, and vary so much by social context, that no single agreed-upon definition of "library" appears to exist. Libraries differ along such lines as type of institution ("academic," "school," "public," and "special" are the usual categories), and by politics, such as the contrast between public library services in democratic and totalitarian systems. Why they differ is little studied, however (Buckland 1988).

Libraries are social institutions that have evolved over a period of many centuries. They serve the information needs of their user communities, adapting collections and services as those needs change. Libraries tend not to be autonomous institutions. Rather, most are funded by governments to serve a defined community; by schools or universities to serve students, teachers, and staff; or by businesses, hospitals, museums, or other organizations to serve their employees and other constituents. Most libraries have a mission statement and collection development plan that identifies who they serve and what they collect, thereby drawing boundaries around their responsibilities.

Although libraries usually are the primary social institution that selects, collects, organizes, preserves, conserves, and provides access to information, few, if any, of these functions are unique. Archives and museums, for example, serve many of the same functions, for the same or similar user communities. The line between libraries and classrooms is blurring as courses are conducted via computer networks. Distance-independent learning environments rely on a steady flow of information resources between



libraries, teachers, and students, and between libraries, homes, workplaces, and virtual classrooms.

Thus, libraries are both institutions and functions. Much of what distinguishes libraries from other information organizations is their professional principles and practices, rather than their functions.

#### Physical Place or Virtual Space?

Libraries serve as gathering points for communities, bringing together people, information resources in physical forms, access to information in electronic forms, and professionals to assist people in their information-related activities (Kent 1996; Lyman 1996; Mason 1996). In the 1990s grand new buildings or additions were constructed for public libraries in Los Angeles, San Francisco, Cleveland, and other major American cities. Britain, France, Germany, and Croatia recently have opened new showcase buildings for their national libraries. The Library of Congress has renovated its original Jefferson Building. These large investments of public monies signify a resurgence of interest in libraries as physical places (Dowlin and Shapiro 1996; Lehmann 1996). The new library buildings are designed to support the latest technologies while preserving existing materials in multiple formats. The new model for library services is not a print library or a digital library; it is a hybrid with complementary print and digital collections. The presence of complementary resources and services is having a crossover effect. People visiting a library building to borrow a book learn to use the Internet while they are there. Conversely, they come to the building to use computer networks, then browse printed materials and leave with a book (Mason 1996).

Many of the functions that libraries serve can be supported online, in a virtual space. People can search catalogs and databases from home or office, can consult with librarians or other advisors by email, and can convene meetings online. Libraries provide services to discover, locate, and obtain materials online, thus many are using libraries without even visiting their buildings.

#### Print or Electronic Resources?

People need information resources that exist in a wide variety of formats. Some wish their libraries would collect a greater proportion of print materials than they do; others would prefer a greater proportion of digital

materials. These debates often degenerate into discussions of the portion of library collections that should be in electronic form—20 percent, 50 percent, 75 percent? Approaching collection development in terms of format begs questions of content and community, however. Traditional approaches to collection development begin with questions of who we are collecting for (community) and what they need (content). The choice of format (electronic, print, or other) usually follows from answers to the questions about community and content. However, these factors interact in complex ways. The materials selected must be in formats that are usable (readable, playable, etc.) on equipment available in the library or in the community, for example.

Even if the same content exists in multiple formats, consideration should be given to the ways in which print and electronic materials are used. Buckland (1992) draws a contrast between the uses of paper documents<sup>2</sup> and electronic documents. Paper is a solo technology, typically best used by one person at a time, and it is localized: the paper and the reader must be in the same place at the same time. Electronic documents can be used by many people at the same time, and they can be located at a different place from the reader (provided that the reader has a telecommunications connection to the site).

Other characteristics of form are relevant in collection choices. Paper documents often are more effective than electronic forms for discussions by a small number of people, for example. Paper is well suited for purposes such as explaining diagrams, where people need to point and make hand motions over a document. Paper generally is more suitable for making annotations and for browsing by flipping through pages than are most forms of hypertext (Dillon and Gabbard 1998; Silberman 1998). New technologies such as “electronic white boards” and “electronic books” try to replicate these features of paper. Paper is a very flexible medium, with many features that are difficult to support in digital environments.

The economic tradeoffs between print and electronic materials are complex and evolving. One factor is the difference in business models for print and electronic publication. Print materials normally are sold as individual copies or annual subscriptions, whereas electronic materials are subject to

2. Buckland (1992) regards other hard-copy materials, such as microform and clay tablets, as inconvenient variations on paper.

a variety of elaborate pricing schemes that may include annual fees, fees per use, fees based on size of user population, and combination fees based on acquiring sets of print and electronic resources from one publisher. Other economic considerations for libraries include the continuing costs of maintaining the materials, whether stored on a bookshelf, mounted on local computers, or access provided to remote computers, and the conservation and preservation of the materials as technology evolves.

#### The Public's View of Libraries

Libraries are a near-sacred social institution, and yet many question the need for their continued existence. The availability of information technology to create content in digital form, to digitize existing printed materials, and to distribute these resources over computer networks is provoking a broad public debate over the role of libraries. The debate involves other public and private institutions as well, and it is part of a more general discussion about the marketplace of ideas in a democratic society (Lyman 1996). Mason (1996) reframes the debate from whether we *can* do without libraries to whether we *should* do without them.

Many people have strong personal feelings about libraries, whether or not they use them. Every library has an intended community of users, such as those residing within the jurisdiction of a public library, or the students, faculty, and staff of a university served by an academic library, or the employees of a business or other organization served by a special library.

#### Public Libraries

Public libraries are the most familiar and visible type of library to most people. Their collections typically include popular reading materials for children and adults, basic reference sources, classics, and resources specific to the local economy and culture. Local history and genealogy are especially popular. They often lend materials such as software, videos, sewing patterns, and art work in addition to the usual books and journals, depending on the circumstances of their local communities. Public libraries provide services such as literacy programs, homework assistance, and public meeting rooms. Particularly in the United States, they serve as the "on ramp to the information superhighway," providing free public access to Internet resources.

People often turn to their public libraries when academic, school, and corporate libraries lack resources, and when they lack Internet access. When child care services are lacking, parents may send their children to the public library after school. Because these institutions provide such a diverse array of services to their local communities, they rarely attempt to build comprehensive research collections, leaving that responsibility to academic libraries and national deposit collections. Instead, they rely on interlibrary lending to provide materials beyond the scope of their collections.

The use of public libraries is growing, and budgets rarely keep pace with increasing demands. In the United States, and to some degree in Europe, libraries are supplementing their budgets with funds from a variety of sources, including private gifts, corporate partnerships, and fees for services. They also rely on volunteers from the community to supplement their paid staffs. Declining materials budgets increase the reliance on interlibrary lending and other cooperative agreements.

Public libraries serve diverse constituencies in their communities, provide a wide array of services, and make difficult choices among competing goals. They balance print collections with online services, physical places with virtual spaces, space for people with space for books, and buying with borrowing. Many of the choices they make occur behind the scenes and thus are not readily apparent to their user communities. Other decisions, such as the choice of certain books, put them in the spotlight. In their attempt to serve diverse communities and to provide open access to information from a variety of perspectives, public libraries sometimes offend the political, religious, moral, or cultural sensibilities of some of their constituents. These concerns are being transferred from print to digital materials with the advent of computer networks.

Most of the people, most of the time, are happy with the resources and services provided by their public libraries (Benton Foundation 1996). Some people, at some times, are not. An obvious example is the controversy over filtering or limiting access to Internet sites, a problem for which no easy answer exists. Filtering is ineffective, due both to the rapid appearance and disappearance of individual sites and to problems with information retrieval. Filtering on "bad words" restricts access to good sites. For example, simply blocking access to sites containing the word "breast" denies access to legitimate sites on breast cancer. Conversely, objectionable sites

will slip through filters if they do not contain any of the specified "bad words." Purveyors of pornography are likely to be as technically sophisticated as those who wish to filter it. Some libraries are controlling access via policies such as requiring identification to use terminals. Others use screen filters that make the display visible only to the person seated in front of it. Yet other libraries refuse to filter or control beyond requiring parental permission for minors, applying the same policies as for print materials. A greater concern for many public libraries is people tying up library terminals for email and games, which diminishes the availability of the terminals for information access.

A less predictable controversy over public library policies is one that erupted at the San Francisco Public Library over the replacement of the card catalog with an online catalog and the "weeding" of the collection. That particular storm was preceded by an article in the *New Yorker* (Baker 1994) directed at these practices in public and academic libraries. Objections to online catalogs were based on the assertion that much information is lost in the conversion of catalogs to electronic form, such as notes on the back of cards. In fact, the reverse is true, as card catalog records usually are upgraded to modern cataloging standards as part of the conversion process (Schortlaender 1992). More subtle are objections to the loss of card catalogs as icons of a print culture and as familiar tools whose well-worn oak drawers and paper cards offer tactile pleasures unmatched by keyboards and computer screens. The card catalog controversy can be attributed partially to communication problems between librarians and their user communities. Librarians do not always include their user communities in decision-making processes for such momentous changes as catalog conversion, nor do they often communicate the reasons for their choices until a controversy arises.

Communication issues regarding the "weeding" of collections cut more deeply into principles and practices of library management. Baker's first *New Yorker* article stated that "the function of a great library is to store obscure books," and that "this is above all the task we want libraries to perform: to hold on to books that we don't want enough to own, books of very limited appeal" (Baker 1994, pp. 78-79). His later article focused directly on weeding policies of the San Francisco Public Library, expressing the view that it is that library's responsibility to preserve these materials for

posterity (Baker 1996). Few American public libraries attempt to maintain research collections, leaving that responsibility to academic libraries and to deposit collections such as the Library of Congress. Storage space for books is expensive, especially in the centers of major cities. Today, public spaces for people to gather, to browse current publications, to read, and to use resources on site are usually viewed as higher priorities. Older materials that are in minimal demand often are stored off site or are "weeded" or "de-acquisitioned"—library euphemisms for "discarded." At such later time as a user requests an item not on site, it is retrieved from storage or requested through interlibrary lending. Libraries usually cooperate to ensure that at least one copy of each discarded item remains in the region and is available for lending. Much as digital documents are moved electronically to the point of demand, libraries are replacing "just in case" storage of physical documents with "just in time" delivery to the point of demand. By sharing this responsibility for low-demand items, they save money that can be diverted to other services.

These principles and practices are not widely known outside the library community. People understandably view the local public library as "their library." The fact that their library functions as part of a tightly coupled infrastructure of systems and services is not apparent unless explained. And even if explained, members of the community may not necessarily agree with the choices made. Such was the case with the San Francisco Public Library, one of the large American public libraries that had, in the past, attempted to maintain a research collection. Baker, members of the staff, and prominent members of the public disagreed (loudly and publicly) with the director's choices in placing other priorities over maintaining the research collection. These decisions were made in consultation with the city council and the library board some years earlier, in the process of designing a new building, but were not evident to the public until the building was opened.

There are no simple ways to resolve these debates, which are about how to achieve the social goals of public libraries in supporting learning, informing the citizenry, and making progress in the "sciences and useful arts." Managers in private companies regularly make tough choices such as "just in case" vs. "just in time" inventory management. Library managers make similar tough choices, but the subject is not spare auto parts; it is

information resources that have cultural significance for their local communities. These debates will continue, and well they should. The public has the right to participate in determining the future of its social institutions.

#### Academic Libraries

The mission of academic libraries is more focused than that of public libraries: They serve the curricular and research needs of students, faculty, and staff of their parent institutions. They collect materials in direct relation to academic programs, with increasing depth for undergraduate, masters, and doctoral courses. As new curricula and programs are established, and as existing programs expand into new areas, library collections are expected to expand proportionately. Among the fundamental criteria for accrediting academic programs is the adequacy of library collections to support the curricula. Major research universities with extensive graduate programs and research facilities necessarily have more substantial library collections and services than do small colleges that focus on undergraduate education.

Even the most extensive academic library collection cannot serve all the needs of a university community, so academic libraries rely on cooperative arrangements for interlibrary lending and document delivery. They also rely on each other for specialized collections. If one library in a region has an extensive collection in one subject area, such as musicology, Latin America, or the history of medicine, other regional academic libraries may maintain only basic collections for those topics and build depth in other areas instead. Libraries borrow from each other and offer access and borrowing privileges to members of affiliated institutions. In Southern California, for example, the University of California at Los Angeles and the University of Southern California may be opponents on the football field, but they are partners in sharing library access. Their library users can observe the advantages of cooperative ventures in borrowing resources and in using the services of neighboring libraries. Less apparent to users is that they gain access to a larger body of resources when partner libraries share costs of expensive materials and decide jointly on which institutions will acquire, preserve, and conserve unique materials.

Major research libraries also serve as national and international resources. They catalog their materials in shared bibliographic utilities,

making their holdings widely known. They collect materials in many, if not most, of the 400+ languages found in shared cataloging systems. Because they are building research collections, they do far less weeding than do public libraries. Storage costs remain an issue, however. Academic libraries often rely on off-site repositories, retrieving materials on demand. The nine campuses of the University of California, for example, share two large regional library facilities, one in Northern California and one in Southern California.

By now, the operations and services of most academic libraries in developed countries are highly automated. Integrated systems track materials from the time of ordering through receipt, payment, cataloging, physical processing (bar codes, spine labels, etc.), shelving, circulation, off-site storage, conservation, and any other handling or disposition. Metadata are exchanged online between library catalog systems and bibliographic utilities. Orders and payments may be exchanged between libraries and vendors through electronic data interchange (EDI), based on international standards. Integrated library systems are linked to other electronic resources, such as databases of abstracting and indexing services and digital libraries of text, images, and other content.

Academic libraries provide integrated access to hybrid collections of print and electronic resources. A common electronic gateway may enable access to a wide array of systems and services (Hyslop 1996; Olsen 1997). From computers in the library, in offices, in dormitories, in homes, in hotels, or elsewhere, members of the university community can identify what resources are owned by their institution, what is on order, and which physical materials currently are on loan to other borrowers. Some systems enable users to request items for delivery to their offices, place holds, extend the due dates for borrowed items, and request materials from elsewhere. Services often include digital libraries maintained on campus networks and on digital libraries located elsewhere. University libraries may pay for subscriptions to external services and provide access through a library or university gateway. Access privileges often are authenticated by domain name or other means. In this way, users are saved the trouble of maintaining passwords for individual systems and services. Users may be searching a rich array of digital collections without realizing that these are not publicly available resources or that their library is paying for their access rights. Many

academic libraries are expanding their services to include electronic publishing, tele-learning, distance-independent courses, and other information technology applications for higher education.

Users of academic libraries are profoundly affected by the shifts in scholarly publishing discussed in chapter 4. The prices of scholarly journals have increased at far higher rates than either library budgets or inflation, resulting in substantial decreases in libraries' purchasing power, canceling of journal subscriptions, and reductions in purchases of monographs. Academic disciplines are affected to varying degrees, leading to tensions between departments. Science and technology disciplines rely more heavily on scholarly journals than on monographs, and a growing proportion of their information resources are available in electronic form. In contrast, the humanities and the social sciences rely on a mix of serial and monographic resources and fewer of their resources are available in electronic form. Heavy use is made of the digital resources available, but these supplement rather than substitute for collections of materials in print, microform, audio, and video formats, collections of photographs and other images, and archives and other collections of unique materials.

Users of academic libraries are acutely aware of the tradeoffs between access to physical items held locally and access to those held elsewhere. In the best of all possible worlds, all would have local and immediate access to all the resources they want, when they want them. Given that this option is neither technically nor economically feasible, librarians try to minimize the turnaround time to acquire materials on demand. The improvements in interlibrary lending and document delivery discussed above are reducing delivery times from weeks to days for postal shipments, and from weeks to hours for online or fax delivery. Such delays, though not desirable, may be acceptable for long-term scholarly research projects, and are an improvement over the situation prior to automation. Scholars frequently had to travel to the location where resources of interest were held. Humanities scholars continue to travel in search of rare or unique materials, although some of their research needs can be satisfied by online access to digitized resources.

The reaction of the academic community to the changes in their library services is mixed. Many items previously available only in print form, such as catalogs, journals, and indexing and abstracting services, are now avail-

able online and are more convenient to use. They appreciate the wide array of new resources available in electronic form, from data sets to digitized manuscripts. However, the majority of the academic community laments the decreasing availability of current journals and scholarly monographs, whether in print or in electronic form, because of escalating costs. Most proposals for new models of scholarly communication endeavor to redress the loss of access to materials of all types, in addition to extending access to new forms of publication.

### Special Libraries

Special libraries in businesses, law firms, hospitals, museums, and other institutions are often the most innovative, responding quickly to the changing needs of their parent organizations. Special libraries operate under names such as "technical information center" or "resource collection" as well as "library." In networked organizations, special libraries offer online access to their resources and services and deliver information products in digital form to their users, in addition to the usual print-based services.

Special libraries are finding a new niche as information-related activities that once were discrete become integrated. For example, product designs and plans that are created online flow directly into the production process. Data about patients or clients can be maintained and used throughout the organization's relationship with those individuals. Similarly, electronic products such as publications, videos, films, games, or broadcast programs can be maintained online, and portions of them can be re-used in other products. The challenge in creating and maintaining information throughout its life cycle lies in describing and organizing content so it can be identified and located later by people other than those who created it, who may be seeking that content for purposes other than that which it was originally intended to serve.

The changing economics of information is influencing the role of special libraries in many organizations. Corporate libraries were the first to be funded as cost centers, such that services are charged directly to organizational units or to clients. The alternative, which is employed in many special and most academic libraries, is to fund libraries as part of overhead and to provide basic services to the user community without direct cost. Cost accounting for information services is problematic, however. The inputs are

hard to measure, as collections are built and maintained over many years, making it difficult to apportion costs to individual transactions. Similarly, the outputs are hard to measure, since the value of the information provided may not become apparent for months or years after the transaction. Yet the costs are real, and libraries must demonstrate their value and justify their costs to their parent organizations no matter what financial models are applied. Internal costing is but one more challenge in developing an adequate economics of information.

The response of user communities to special libraries can be assessed only in general terms, since the services of special libraries are tailored to their unique communities and since studies of those services are usually proprietary. The job market for information professionals in private industry is expanding rapidly, and new job titles are appearing, such as "metadata specialist," "information analyst," "digital resources coordinator," and "digital archivist." At the same time, some special libraries are being closed as part of cost reductions in research and development, information services are being outsourced, and librarians' jobs are being eliminated, indicating that these units are not necessarily seen as essential to the future of their parent organizations.

### Rethinking Libraries in a Digital Age

The question for the new millennium is not what to do about libraries but, more generally, what to do about access to information. What resources and services are needed, and by whom? Which are essential and which are desirable? What infrastructure is needed to support information resources and services? Who should provide them? These questions encompass a range of political, economic, and social issues. Historical democratic premises for providing open access to information through libraries are being challenged as information resources become a form of capital and as more aspects of information infrastructure are privatized.

Libraries are only one of many institutions in the midst of substantial structural changes. In North America, Europe, and elsewhere, fewer government funds are being devoted to public services such as libraries than in past decades. Professionals in the public and the private sector alike are expected to increase their productivity, which usually means more supervi-

sion of paraprofessional and clerical staff and a lower proportion of professional staff. Libraries are being asked to do more, often with fewer resources. Few of their traditional responsibilities can be abandoned, although many can be fulfilled more efficiently through automation and resource sharing, some can be deferred, and some can be relinquished to other agencies.

The present discussion of this vast range of issues will be limited to the evolution of libraries and to their present and future roles in providing access to information in a digital world. Four challenges faced by libraries are the invisibility of infrastructure, the changing nature of collections, the preservation of materials and content, and the boundaries between information institutions.

### Invisible Infrastructure

Infrastructure tends to be visible only when it breaks down. The invisibility of information work in particular was identified many years ago (Paisley 1980). Infrastructure is embedded in other structures, is transparent, and is linked with conventions of practice of day-to-day work (Star and Ruhleder 1996). Much of what libraries and librarians do is not apparent to their users. Considerable professional time and vast amounts of paraprofessional and clerical time are devoted to the processes of selecting, collecting, organizing, preserving, and conserving materials so they are available for access. The selection process requires a continuing dialog with the user community to determine current needs, continuous scanning of available resources, and judicious application of financial resources. Once selected, the items are collected, whether in physical form or by acquiring access rights. This process, which requires negotiation with publishers and others who hold the rights to desired items, sometimes takes months or years, depending on the materials and the rights. As new items are acquired, metadata are created to describe their form, content, and relationship to other items in the collection. Once in the collection, resources must be preserved and conserved to ensure continuous availability over time.

### Invisible Successes

If all these processes are done well, then people can find what they need, when they need it, in the form in which they want it. Users may encounter



a catalog that is easily searched, a web site with a clear organization of relevant materials, a library building with good signage that leads them to desired resources, and professional staff who know the collection and can guide them in their information-seeking activities. Conversely, when these processes are not done well, users are aware of inadequate resources, poor organization, complex user interfaces, and a lack of staff to assist them.

Libraries risk being victims of their own success. The more ubiquitous their presence in their organizations and communities, the less apparent their role may be. The more services that are provided electronically, and the less need to visit a physical place, the less their users may be aware of who provides those services.

#### *Invisible Content and Costs*

The claim that the Internet will replace libraries often is based on questionable assumptions. Three common misconceptions are that all useful information exists somewhere on the Internet, that information is available without cost, and that it can be found by anyone willing to spend enough time searching for it (Borgman 1997c; Miller 1997).

**online and offline resources** Only a small portion of the world's information exists in electronic form, much less on the Internet. Few journals or other print publications were created in electronic form until the late 1960s, and most of those early tapes and disks were discarded after the print products were published. Indexing and abstracting services began producing online databases in the 1970s, rarely converting print indexes that were created earlier than the 1960s. Vast portions of the world's knowledge resides in libraries, archives, museums, government agencies, and in private hands, in print or other hard-copy form, having been collected over decades or centuries. To use these materials, their existence first must be discovered, then access acquired. Catalogs, indexes, finding aids, and other forms of metadata, online and offline, serve the discovery function. Though some older materials such as rare and valuable manuscripts would gain a broader audience through digitization, relatively few older documents are in sufficient demand to justify digitization "just in case" someone wants them. Delivering physical copies or digitizing materials on a "just in time" basis are more feasible scenarios from an economic standpoint.

**fee vs. free** The assumption that everything on the Internet would be free was more prevalent before the growth of online bookstores, shopping malls, and stock brokerages (Borgman 1997c; Miller 1997). Naive statements that "information wants to be free" are gradually being replaced by the recognition that resources available without charge are being paid for by someone: advertisers, governments, universities (as part of grant or research projects), political or religious groups, or public or nonprofit organizations. With the exception of government information, such as the databases of congressional legislation and records maintained by the Library of Congress<sup>3</sup> and reports filed by publicly traded corporations with the Securities and Exchange Commission,<sup>4</sup> or university-supported full-text digital libraries of out-of-print materials, little of the content available without charge on the Internet duplicates content collected by libraries. Libraries deliberately collect materials that are difficult for individuals to acquire on their own. Many items, such as scholarly journals and legal and technical databases, are prohibitively expensive unless amortized over the cost of a large user community. Others, such as unusual items from foreign sources, may be available for a nominal fee but expensive to identify and acquire. Once materials are out of print or otherwise unavailable for purchase, libraries usually are the only place to find them. In assessing the cost of acquiring information, one must take into account the time spent by information seekers as well as the direct costs of the resources and the time spent by librarians and other personnel in making those resources available. Students may place little monetary value on their own time, but spending many hours to discover and acquire "free" resources is rarely cost effective in profit-making environments. Professional assistance in information retrieval is often more cost effective than self-service.

**added value** In assessing the tradeoffs between information resources available for purchase and those offered without fee, users need to consider other factors such as the integrity of the source, the reliability and validity of the content, and other intangibles. Commercial vendors of such

3. E.g., Thomas, named "in the spirit of Thomas Jefferson." The URL is [www.thomas.loc.gov](http://www.thomas.loc.gov).

4. E.g., EDGAR (Electronic Data Gathering, Analysis, and Retrieval system). The URL is [www.sec.gov/edgarhp.htm](http://www.sec.gov/edgarhp.htm).



products add value by ensuring the accuracy of content and quality of reproduction (or other characteristics specific to the format), offering multiple versions or formats of resources, and providing customer support. If users of digital information products have problems implementing them on local systems, questions about searching, or other difficulties, they usually can get assistance from the producer. Similarly, libraries provide support for their services and obtain additional assistance from the producers of the material they purchase. Such support is rarely available with free Internet resources, and thus substantial costs can be incurred in making use of them. The practice of law is especially dependent on information resources that are accurate, authoritative, and current. Cases can be won or lost based on the quality of the information on which legal arguments are founded. Lawyers, law professors, and law librarians are concerned about the integrity of new online resources in comparison to the trusted sources on which they have long relied (Ballard, Spahr, Andrews, and Ingersoll 1999; Edwards 1997; Haigh 1997; Newman 1999; White 1995). Many valuable sources of legal information are appearing on the World Wide Web, both on general subjects and on practice-specific areas such as tax, securities, intellectual property, and anti-trust (Johnson and Krzyminski 1999). Some free sources of case law and statutes are appearing online in direct competition with the established and expensive commercial services provided by Lexis-Nexis and Westlaw. Legal professionals are reluctant to rely on those free resources, however, because neither the accuracy that commercial publishers provide through proofreading and verification of final authoritative versions, nor the persistence of the source, are ensured (Edwards 1997; White 1995). Online sources for treatises and other materials may have comparable content, but some find the analytical capabilities inferior to that of paper (Haigh 1997). The concern for accuracy and authority in online legal resources has spawned a series of articles identifying criteria for evaluation and web sites that evaluate other online sources (Ballard, Spahr, Andrews, and Ingersoll 1999; Edwards 1997; Haigh 1997; Newman 1999; White 1995). All these authors conclude that the onus is on the researcher to evaluate the integrity of the source. To some extent, this always has been true. What is new is the proliferation of self-publishing and the impermanence of online sources. Determining the source and reliability of content is much more difficult in this environment.

The imprimatur of an established publisher who ensures accuracy and assumes responsibility is an intangible value for which it is difficult to assign a price.

#### Content and Collections

In the eyes of the public, libraries are often synonymous with their collections. What may not be apparent is that collections are much more than the sum of their parts, bringing together disparate items and identifying relationships between them. The world's great collections are known for their depth in particular areas of study, rather than for breadth alone. Digital collections and hybrid collections of materials in many formats raise questions of what it means to "collect," especially when no physical artifact may be associated with a purchase.

#### Purposes of Collections

Historically, library collections have served four basic purposes: preservation (keeping materials for the future, as they may be unavailable if not collected at the time of their creation), dispensing (providing access to their contents), bibliographic (identifying what exists on a topic), and symbolic (conferring status and prestige on the institution) (Buckland 1992). All these purposes change in character with digital and hybrid collections. When a library acquires access to remote digital libraries on behalf of its user community, is that digital library part of "the collection"? Who is responsible for preserving digital content in distributed environments? What are the boundaries of a library's collection when it dispenses resources that it does not physically house and may not own? When libraries rely on cooperatively maintained digital libraries of metadata to determine what exists, where it exists, and how to acquire access to it, who is responsible for bibliographic control? Does having a large collection of electronic resources confer the same status on an institution as having a large collection of print materials?

#### Digital Collections

Libraries always have collected materials in a variety of formats, but digital documents are the first to transcend time and place. Digital documents can exist in numerous identical copies, or one copy can be accessed by

numerous users, accentuating the tradeoffs between assets and access (Higginbotham and Bowdoin 1993). Collection-development librarians try to spend their funds on the materials judged most likely to be needed by their user communities, and then rely on cooperative agreements to borrow other items on demand. However, in the case of electronic resources it may be more reasonable to acquire items only on demand. For example, journals may cease to exist as volumes of multiple issues, each with multiple articles; they may be replaced by unitary articles that are acquired individually.

#### Uses of Collections

As Buckland explains in his classic text on library services (1988), many books have been written on how to collect library materials in support of a user community, but little attention has been paid to why libraries collect materials or to how people actually use those materials. In adapting information institutions to new environments, more needs to be known about how people use both printed and digital materials. Information studies research is expanding in scope to encompass the full information life cycle: capturing information at the time of creation, making it accessible, and maintaining and preserving it in forms useful to the user community. Concerns also are expanding to include group processes, rather than viewing individuals as sole actors in information-related activities. The concept of a collection is changing accordingly. Collections are hybrid in form and in scope, linked with other collections internal and external to the organizations that support them.

#### Preservation and Access

Until recently, concern for the preservation of information resources was relegated to a small cadre of archivists, special-collections librarians, and museum professionals. Public awareness of the fragility of software and of the impermanence of digital data increased substantially because of the "Y2K" problem of converting software to handle dates in the year 2000 and beyond (Chen et al. 1999). Although preservation is a high priority, it must be balanced with access considerations. The best methods for long-term preservation are often poor for access, and vice versa. For example, microfilm is superior for preservation purposes but inferior for access, and digitization is superior for access but inferior for preservation.

A number of factors have made preservation and access pressing matters. One is the concern for preservation of existing print collections, many of which are physically deteriorating because of the fragility of the paper, film, microform, or other materials on which they were created and because of inadequate storage conditions. Another is the lack of means to maintain digital content in a usable form through changes in standards and technology. A third is the difficulty of maintaining persistent access to digital collections. A fourth is the failure to capture and preserve digital content at the point of creation, resulting in the deterioration of public and private records systems.

#### Preservation of Physical Documents

In a print environment, preserving materials means acquiring them while they are available and then storing them safely for use by present and future generations. Similarly, conservation of print means repairing physical materials to keep them usable, such as patching torn pages and treating brittle paper with chemicals. In most countries, national libraries and archives are charged with preserving and conserving the cultural heritage of their nations. Public libraries and local government archives often have similar responsibilities for their local communities. Special libraries and corporate archives have parallel responsibilities for the records of their parent organizations. These organizations follow accepted practices and standards, which are supported by national and international cooperative relationships, all of which are part of the (often invisible) information infrastructure.

"Brittle books," the crumbling of materials printed on acid paper between the middle of the nineteenth century and the middle of the twentieth, is the most widely known preservation problem. Large portions of the world's collections of print materials from this period are deteriorating, especially those not stored in modern buildings with adequate heating, air conditioning, and humidity controls (Kislovskaya 1996; Lesk 1990; Porck 1996; Sanders 1987). The Commission on Preservation and Access was formed in 1986 "to develop and encourage collaborative strategies for preserving and providing access to the accumulated human record." *Slow Fires*, a video about the deterioration of paper, film, and other physical storage media, was shown widely on television and at public events. The campaign raised awareness inside and outside the library and archives fields,

involving historians and other scholars, governments, funding agencies, and the public at large.<sup>5</sup>

International cooperative efforts are underway with the long-term goal of preserving at least one copy of each item deemed of historical significance and providing access to its content (Cook 1997; McClung 1996). No single institution can afford to preserve and conserve all its deteriorating materials; however, by dividing the responsibility, each institution can preserve its share of the world's collection and improve the long-term availability of content for all. Unfortunately, this solution is less simple and complete than it sounds. Preservation and conservation are too expensive to save everything that anyone deems significant, and thus the selection process is often contentious. Underlying the selection process is the difficulty of setting objective criteria for significance, in view of the many social, cultural, and political factors involved. A critical policy challenge lies in determining who decides what is significant.

#### *Preservation of Content*

The "brittle books" problem was a catalyst for more general discussion of preservation, conservation, and access to information resources. The first level of preservation decisions is often whether the physical object is sufficiently significant to save and conserve, or whether preserving the content is sufficient. Sometimes both are required.

If items are significant as artifacts, such as manuscripts or first editions of important works, then physical conservation is required. If items are worth saving primarily because of their content, such as classic works that were printed in many copies, then photographing or scanning them may be sufficient. If the content is to be captured and the artifact saved, then the document is handled carefully, turning pages individually and laying the document flat for photography or scanning. This process and some methods of chemical treatment can cause further physical deterioration, so materials are best handled only once (Lesk 1990; Lynn et al. 1990; Porck 1996; Sanders 1997). If only the content is to be preserved, then the object may

5. The Commission on Preservation and Access is now a program of the Council on Library and Information Resources (1755 Massachusetts Ave. NW, suite 500, Washington, DC 20036; <http://clir.stanford.edu>). Slow Fires (Sanders 1987) is available on video from the Council, as is *Into the Future* (Sanders 1997).

be dismantled so that individual pages can be stacked in high-speed automatic devices. The latter approach is the least expensive method of preserving content, but it destroys the object in the process. Similar tradeoffs exist in preservation and access for materials other than paper. Film, videotape, and sound recordings are deteriorating too. Sometimes the physical artifact must be conserved and saved, and at other times it is sufficient to transfer the content to another format.

Once the content is captured, it can be stored as a digital document or on microfilm, depending on form and content, and on access and preservation requirements. Microfilm is the most common form for preserving text and still images, as the film is readable for as long as 300 years under optimal conditions. Microfilm is poor for access, however. Pages must be located on microfiche or long reels of microfilm, and most people find the machines unpleasant to use. Digitization is often proposed as an alternative, because it is excellent for access. At present, however, digitization is poor for preservation (Ester 1996; Lynn et al. 1990; Smith 1999; Weber and Dorr 1997).

These decisions about the means by which content is preserved bring issues of form and function into sharp relief. The most cost-effective method of creating a global digital library would be to scan the collections of the world's libraries and archives with fast automatic scanners, destroying the materials in the process. This is not a socially acceptable solution, of course. Not even the most heartless technocrat would destroy the Gutenberg Bible or Shakespeare's First Folio for the sake of cheap digitizing. But where should the line be drawn? There is neither enough time nor enough money to preserve in physical form all the printed materials, films, videotapes, and sound recordings that already exist, much less all the letters, memoranda, reports, and other documentation. Among the most significant—and delicate—decisions that librarians, archivists, museum curators, and other information professionals make is what to select, preserve, and conserve. Lacking adequate funds for preservation, they are faced with difficult choices between saving the content of a large number of objects through inexpensive scanning and saving much smaller number of objects through physical preservation. The fact that these institutions destroy some materials in the process of saving them is neither widely known nor widely advertised. The tradeoffs in preservation and access and the associated economics

are difficult to explain to a book-loving public. A major research library recently found itself at the center of public controversy over the destruction of books in the preservation process. Despite following accepted professional practices, the library ceased dismantling books in response to public opinion. As a result, it will be able to preserve the content of far less of its crumbling collection.

#### **Preservation of Digital Documents**

Whether materials are "born digital" or digitized from other formats, digital documents must be maintained in a form that continues to be useful and usable to those who may later seek it. In view of the rate of advances in information technology, maintaining content in a continuously viable form is a major challenge. Magnetic media (computer disks, audio, video, and data tapes, etc.) must be copied every few years to maintain the readability of content, and must be stored properly to ensure long-term readability (Hedstrom 1998; Van Bogart 1995). Even if the medium remains viable, it can be difficult to find devices that will read older formats. Already it is difficult to locate operational devices to read media that were widely distributed only a few years ago, such as 5¼-inch floppy disks and 3½-rpm phonograph records. Devices to read 8-inch floppy disks, 78-rpm records, Betamax videotapes, and reel-to-reel film are even harder to find. Drives for 3½-inch disks have ceased to be a standard feature of new computers; thus, these disks soon will become obsolete.

Even if the media are readable and devices are available, a user must locate hardware with the necessary operating systems and application software to read older files. Many people have cabinets full of disks containing word processing files created on versions of Microsoft Word or WordPerfect for pre-Windows DOS. Others have files created on Wordstar under the CP/M operating system, which competed with early versions of DOS. Similarly, many people keep files containing spreadsheets, images, sounds, and other content long after they have disposed of the hardware and software on which the files were created. Unless these files are transferred to subsequent generations of hardware and software quickly, they are not likely to be read again.

Various methods have been proposed for dealing with hardware and software obsolescence and with the deterioration of magnetic media. So far,

all the proposed methods have significant limitations. The least viable approach is to create a "cybernetic museum" (Cook 1995) containing working models of obsolete hardware and software. This approach has several weaknesses. It is not likely that sufficient spare parts could be acquired and sufficient expertise maintained to make this approach feasible over a long period of time. Obtaining and maintaining all possible combinations of software for each generation of hardware would be difficult. Another weakness of the museum approach is that data created on older devices (e.g., 8-inch floppy disks) will have been transferred to newer media (e.g., CD-ROMs) for which no device drivers exist on the original machine (Rothenberg 1999). A more promising approach for some applications is to emulate software applications and associated operating systems on temporary systems, thus enabling older software to be run on newer computers (Rothenberg 1995, 1996, 1997, 1999). The most general solution, and one currently in use at large scientific data centers, is to migrate digital content to new media with each new generation of technology (Baru et al. 1998a, 1998b; Foster and Kesselman 1999; Halem et al. 1999; Hedstrom 1991, 1993, 1998; Moore et al. 1999; Rajasekar et al. 1999). The migration approach requires continuous investment in the data to maintain access, but it takes advantage of newer, denser, faster, and less expensive storage technologies as they become available.

All the proposed data-preservation strategies require active efforts to maintain the data in readable form, rather than the passive strategies of putting a book on a shelf or putting paper and microfilm in a storage vault. Many different strategies will be required to address a wide range of circumstances, such as large scientific data centers, corporate archives, faculty research data and teaching materials, and families' digital photographs, genealogical, and tax records.

At this time, we have more questions than answers about how to ensure long-term preservation of digital materials. Awareness of the problem is growing rapidly, which should lead to more research and more policy aimed at solving it (Chen et al. 1999; Hedstrom 1998; Rothenberg 1999).

#### **Persistent Access**

In a print environment, libraries hold copies of materials long after their authors are dead and their publishers are out of business. In a digital

environment, when the creator of a digital library goes out of business, or decides that providing the service is no longer profitable, or stops supporting the system for any other reason, the content may cease to be available. Digital libraries can disappear without notice, such as when the computer running them crashes and the owners decide not to revive it. Libraries that provide access to commercial digital libraries may do so under lease arrangements that prohibit them from maintaining older materials or converting them to another format when the publisher stops supporting the product. Universities are concerned about the disposition of digital libraries created by their faculty once research funding ends, the course ceases to be taught by that instructor, or the instructor leaves the university. Back issues of electronic journals or other materials may not be available anywhere, to anyone, if those who hold the rights to them have not ensured their preservation.

#### *Capturing Digital Records*

Creating records in digital form is introducing a host of unanticipated preservation problems in addition to those outlined above (Cook 1995; Cook 1997; Ester 1996; Kahle 1997; Lesk 1997b; Lynn et al. 1990; Society of American Archivists 1995). When organizations created all their documentation on paper, they had systematic means of controlling it. Most documents were created in multiple copies, with at least one stored in a central place under the control of archivists, records managers, or librarians. When people later needed to find memoranda, purchase orders, contracts, design specifications, correspondence, or other essential documents, they knew where to look. Now most such documents are created on desktop computers, and the "originals" may reside only on local disk drives or on network servers. These documents are subject to purging when the computer user needs more disk space, when the computer system is upgraded, or when a new employee takes over the job. Documents may be password protected and thus under the exclusive control of the individual who created them. Although documents on personal computers may be copied as part of routine network backup procedures, the documents may not be readable by others who do not have the appropriate software or passwords. Further, documents captured in backup procedures may be

described only by the file name, lacking other metadata necessary to identify their contents. Few organizations have systematic and effective plans to ensure long-term preservation and access of documents created on personal computers.

#### *Institutional Boundaries*

Demands facing libraries, such as the risk of becoming an invisible part of the infrastructure, the changing nature of collections, and preservation and access of content, are also concerns of archives and museums. Rayward (1993, p. 230) notes that libraries, archives, and museums had "an undifferentiated past." Throughout much of the eighteenth and the nineteenth century, books, papers, works of art, specimens of plants and animals, fossils, minerals, coins, and other objects were gathered in common collections. These collections supported broad, multi-disciplinary intellectual interests, without the division between the sciences and the humanities that we take for granted today. The functional differentiation of these three institutions is relatively recent and has resulted in separate physical locations and specialized professional practices. Each takes primary but not exclusive responsibility for a type of material. Libraries mostly collect published materials. Archives mostly collect the records of individuals, organizations, and governments. Museums collect almost anything, organizing it around a general theme (such as art, history, or natural history), a specific theme (such as air and space or automobiles), or a highly specialized theme. Two extremes of specialization (both located in Los Angeles) are the Museum of Jurassic Technology (Weschler 1995), which collects oddities of the natural world, and the La Brea Tar Pits, an active scientific site that displays objects retrieved from the adjacent tar pits.

Partitioning intellectual content among these three sets of institutions is an artificial division of the natural world, and it does not necessarily serve the information seeker well. To study a painter, for example, a student or a scholar ideally should have access to published works about the artist and the artist's work (e.g., books and journals collected by a library), primary materials (e.g., letters and sketches collected by an archive), and the paintings themselves (e.g., the objects collected by a museum). Many information-bearing objects could reasonably be housed in any of these institutions,

and indeed ownership is often contentious. Does an important map once owned by a famous person belong with his or her other papers in an archive, or in a library map collection, or framed in a museum? Does a book containing original art works as illustrations belong in a library, in a museum, or in a museum library? Does a bust of a significant author belong in a library reading room, or in a museum? No clear answers to such questions exist for physical objects, and the questions become more complex with digital objects.

Only in recent years have these professional communities started to work together on common standards for description, representation, and organization. Although considerable progress is being made, the digital collections managed by these communities still are far from interoperable. Professional education for libraries and archives is beginning to converge, but libraries and archives maintain large bodies of professional practice distinct to their institutions.

Not are libraries, archives, and museums the only institutions with their own sets of professional practices related to information. Creators of geographic and spatial data have developed sophisticated metadata standards and are playing a key role in digital library research. Medical records management is a growing area, owing to the sometimes-competing concerns of health-care providers, insurers, patients, and regulators. Similarly, legal informatics is a distinct specialty, in view of the specialized structures and practices of the legal community. Scholarly and professional societies exist to serve these and many other specialties in information science, informatics, librarianship, archives, museums, records management, and related areas. Although multiple societies sometimes cooperate in organizing conferences, meetings, or publications, considerable duplication of effort exists.

As the content of collections and the metadata required to manage those collections are maintained in digital form, the distinctions between libraries, archives, museums, and other information institutions are blurring. In the long run, a holistic approach to information access holds promise for breaking down artificial barriers between disciplines and media, and thus doing a service to the user community. In the short run, these institutions and professions are seeking new niches in which to make unique contributions, sometimes cooperating with each other and sometimes competing.

## Summary and Conclusions

Computer networks offer a wealth of new opportunities for providing access to information, so much so that the continuing need for libraries is being questioned. The real question is not whether libraries are needed, but how best to provide access to information in a networked world and how best to support the marketplace for ideas. Libraries in democratic societies are part of a social strategy to promote learning and invention and to ensure an informed citizenry. As methods for creating, seeking, and using information are adapted for digital documents and distributed networks, the means for maintaining these social values are being reassessed. As knowledge becomes a form of capital, principles of open access to information in democratic societies also are being challenged.

Libraries, archives, museums, and other institutions have a long history of theories, principles, and practices for managing information in print. They have adapted their approaches to new media as they are invented. As a result, information institutions are distinguished more by their principles and practices than by the type of materials they manage. Many, if not most, of their approaches for managing physical artifacts can be adapted for digital documents. Libraries in particular play a central role in their countries' information infrastructures, selecting and acquiring resources when they are available, organizing them, preserving and conserving them, and providing access to them for their user communities. They work cooperatively to create "virtual collections," balancing "just in time" delivery with "just in case" collecting.

Relationships between libraries and their user communities also are changing. Public libraries have a broad service mandate and are often in the public eye. They make difficult choices between competing goals, such as print collections vs. online services, physical places vs. virtual spaces, space for people vs. space for books, and buying vs. borrowing. Tradeoffs among competing goals faced by academic libraries are similar to those faced by public libraries. In addition to their traditional responsibilities, many academic libraries are expanding into new areas, such as electronic publishing and tele-learning. Special libraries serve the most clearly defined user communities and provide the most tailored services. They rely heavily



on computer networks, digital libraries, and the collections of academic and public libraries to fill the information requirements of their users.

I identified four challenges in rethinking libraries in a digital age. The first is how to maintain visibility while being part of a well-functioning information infrastructure. When people are able to find the information they want, when they want it, and in a useful form, they are often not aware of the effort, expertise, and economic resources involved in providing it. The second is how to manage collections as they become more hybrid and distributed. The third is how to preserve physical and digital materials. Present collections of print, film, magnetic media, and other materials are deteriorating because of the instability of media on which they were recorded and because of poor storage conditions. Future collections are also at risk, because digital media deteriorate and the technology necessary to read and interpret them becomes obsolete. The fourth challenge is how to take advantage of the blurring boundaries between information institutions and information professions. In developing new approaches to managing distributed information resources, it should be possible to draw on the best theories, principles, and practices of libraries, archives, and museums. The fundamental goal is to balance cooperation and competition in implementing social strategies that continue to support cultural values for a digital age.

## 8

### Acting Locally, Thinking Globally

Ideally, a global digital library would provide access to information in all the world's languages, to all people, all the time. Anyone could create information in their native language, yet others could discover that information and have it translated into their preferred languages and formats. The ideal case may be well beyond our understanding of technology or human behavior, but it raises issues worthy of exploration. Generally speaking, the easiest information systems to construct are those with small, consistent collections that serve small and homogeneous user communities. In contrast, a global digital library is the hardest-case scenario: a vast and disparate set of collections intended to serve a vast and heterogeneous user community.

No single collection, user interface, or set of system capabilities will serve young and old, novice and expert, artist and physicist, in Peoria and Prague. Nor will any single system provide adequate access to books, music, movies, and numeric data, much less serve applications as diverse as electronic commerce, weather modeling, census tracking, library catalogs, and virtual classrooms. Yet people with varying backgrounds and skills, speaking different languages, have similar information needs. Content of interest may exist in a wide variety of forms and languages, and the same content and collections may be of interest to a wide range of people, for many different reasons. The prospect of a global digital library presents several opportunities. One is to make information resources accessible to particular user communities while at the same time making those same resources accessible to a broader, ill-defined, and perhaps unknown audience. Another is to enable users to bridge the many formats, representations, and languages of individual digital libraries in their quest for information resources. This