The Invisible Library: Paradox of the Global Information Infrastructure

CHRISTINE L. BORGMAN

ABSTRACT
LIBRARIES ARE AN ESSENTIAL COMPONENT of a nation's information infrastructure, yet often they are invisible to their users and other stakeholders. In the context of this special issue, the paper presents four challenges faced by libraries and proposes research designs to address each of them. The four challenges involve: 1. invisible infrastructure, 2. content and collections, 3. preservation and access, and 4. institutional boundaries. I propose a mixture of research methods that includes surveys, case studies, documentary analyses, and policy analyses. Only with a better understanding of these challenges can libraries find their best fit in the information infrastructure of our networked world.

INTRODUCTION

Computer and communication networks now encircle the globe. Despite the oft-repeated claim that half the world's population has never made a telephone call, we receive daily television, radio, and newspaper reports filed via satellite from Afghanistan, one of the planet's least-developed countries. Many of these reports become available almost immediately on the Internet. Information technologies have become ubiquitous in the developed world and widely available elsewhere.

An increasing proportion of communication and commerce takes place via computer networks. Friends, family, colleagues, and strangers rely on e-mail to maintain relationships and to transact business. Most of the activities of writing, editing, and publishing involve computers and networks regardless of whether the final product appears online or on paper, mak-

ing "electronic publishing" a misnomer. Even in the "old economy," orders are placed, invoices are paid, and credit cards are verified and charged via computer networks. Individuals turn to the Internet as a primary source for all sorts of information—health, hobbies, homework, news, shopping, music, games, research, and general curiosity.

Libraries are but one of many institutions that could no longer function without computer networks, at least in the developed world. Libraries depend upon computer networks as a means to provide access to local and remote information resources. While physical materials continue to form the core of most library collections, fewer and fewer services require that users physically enter the library building. Even artifacts such as books can be ordered online for delivery to one's home or office.

A paradox of the networked world is that as libraries become more embedded in the information infrastructure of universities, communities, governments, corporations, and other entities, the less visible they may become to their users, funders, and policy-makers. Libraries must be integral components of the information infrastructure of their organizations if they are to provide the most effective, efficient, and appropriate services to their user communities. Independence and isolation are not suitable alternatives.

Historically, libraries have played key roles in information-oriented societies. Yet today, some of their roles are being duplicated by other public institutions such as archives and museums and by commercial providers of content and services. Individuals and organizations now have many information sources alternative to those provided by libraries, which would suggest that the role of libraries is shrinking. However, libraries are expanding to include a wider array of services, such as providing digital libraries and support for distance learning. Despite this broader scope, libraries exist in a competitive environment, facing greater demands for services and often with fewer resources to meet those demands.

Libraries can and should play key roles in the emerging global information infrastructure. To do so, they must address a number of complex challenges. Research on these challenges will assist libraries in identifying and accomplishing their roles in a global information infrastructure. The four challenges for libraries are introduced in a recent book (Borgman, 2000). Here I extend and update those issues, frame them as research questions, and suggest methods to explore them.

INFORMATION INFRASTRUCTURE

A first step in exploring the role of libraries in a global information infrastructure is to consider what is meant by "infrastructure." Familiar phrases such as "national information infrastructure" and "global information infrastructure" are rarely accompanied by clear definitions of the underlying concepts. Star and Ruhleder (1996) were among the first to de-

scribe infrastructure as a social and technical construct. Their eight dimensions can be paraphrased as follows: An infrastructure is *embedded* in other structures, social arrangements, and technologies. It is *transparent,* in that it invisibly supports tasks. Its *reach or scope* may be spatial or temporal, in that it reaches beyond a single event or a single site of practice. Infrastructure is *learned as part of membership* of an organization or group. It is linked with *conventions of practice* of day-to-day work. Infrastructure is the *embodiment of standards,* so that other tools and infrastructures can interconnect in a standardized way. It builds upon an *installed base,* inheriting both strengths and limitations from that base. And infrastructure becomes *visible upon breakdown,* in that we are most aware of it when it fails to work—when the server is down, the electrical power grid fails, or the highway bridge collapses.

Integrated library systems (i.e., automated systems that support core processing functions such as acquisitions, serials, cataloging, and circulation) offer a familiar example of an infrastructure within an organization. Following Star and Ruhleder's (1996) model, we see that integrated library systems are embedded in the work practices of libraries and depend upon certain jobs and relationships in addition to specific technologies. They support the processing of materials and resources at multiple sites and enable remote access to cataloging and other databases twenty-four hours a day. Upon joining the community, both staff and patrons learn to use the systems and to develop certain expectations of services. Integrated library systems embody national and international standards, both library-specific (e.g., MARC, Z39.50) and general technical standards (e.g., Unicode, TCP/IP). These systems build upon an installed base—usually consisting of cataloging records, holdings records, and other records in standard formats—and established practices. When the system breaks down—for example, when library catalogs cannot be searched, or when books cannot be renewed—then the infrastructure becomes very visible.

Information infrastructure is only one type of infrastructure, but one that has at least three definitions. Firstly, the term "information infrastructure" is often used as a public policy construct to include technical capabilities of the network, rights and guarantees of network services, and means for funding development and for regulating the network. Some examples are the (U.S.) National Information Infrastructure Act of 1993 (National Information Infrastructure: Agenda for Action, 1993), the European Union proposal for a unified European Information Infrastructure (Europe and the Global Information Society, 1994), and the Group of Seven (G-7) Ministerial Conference on the Information Society (1995). This last document established a framework for a global information infrastructure.

A second sense of the term "information infrastructure" is as a technical framework that incorporates the Internet and its services (National Research Council, 1994). The Internet is a network of networks, linking many layers of networks within organizations, within local geographic ar-

eas, within countries, and within larger geographical regions. The third sense of the term "information infrastructure" is as a general framework that encompasses a nation's networks, computers, software, information resources, developers, and producers (National Information Infrastructure: Agenda for Action 1993). In this article, the term "information infrastructure" is used in this last sense of an encompassing framework.

THE ROLE OF LIBRARIES IN INFORMATION INFRASTRUCTURE

Libraries are inherently information institutions. They are part of a nation's information infrastructure in the third sense of the term (above). Libraries rely heavily on computers and computer networks, at least in developed countries. They select, collect, organize, preserve, conserve, and provide access to information resources. They provide an array of information services, and may also develop and produce content. Although these characteristics suggest that libraries would be considered central to the development of information infrastructure in most countries, few policy documents about information infrastructure mention the role of institutions such as libraries, museums, or archives in providing content or services. Clearly, it is up to the library community to identify and articulate its goals in information infrastructure and to act upon them.

This article addresses several of the challenges facing libraries in determining their present and future roles in their nation's information infrastructure and in a global information infrastructure. These challenges involve the following issues:

1. Invisible infrastructure
2. Content and collections
3. Preservation and access
4. Institutional boundaries

These four topics were first proposed in Borgman (2000, chapter 7). Here I extend the scope of each topic, identify associated research questions, and suggest methods by which the questions could be addressed. The conceptualization and literature reviews are by no means exhaustive. Each of the four challenge topics is deserving of article-length, if not book-length, treatment. For ease of explanation, the research designs are described as individual studies at individual institutions. In practice, research should be replicated at many institutions. Better yet, research to address these challenges should be coordinated to provide broad insights on a regional, national, and international basis.

For this special issue we were asked to identify research questions that are important for the field to address in the next five to ten years. Predicting the future is always risky. Issues and trends are hard to spot, and even if on target, the timing is unlikely to be accurate. The best one can do is to offer "the view from here." Thus, implicit in these research designs is

the admonition to reassess the relevance of these questions, and to do so continuously. As scholars and practitioners, we should endeavor to scan the environment and to be aware of issues at least one day sooner than our stakeholders.

Challenge 1: Invisible Infrastructure

Despite the expanding scope of library services, more people seem to claim that they never go to the library anymore because everything they need is online. Even more disturbing are statements by managers who expect to build new campuses or new offices with minimal library collections, because they see a diminishing role for libraries. Why are libraries so invisible?

The invisibility is partly due to the successes of the institution. Good library design means that people can find what they need, when they need it, in a form they want. Good design is less obvious than bad design, and thus libraries risk being victims of their own success. Another component is the invisible content and costs of libraries. Many users are simply unaware of the expense of acquiring and managing information resources or the amount of value added by libraries and librarians. 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 that they are available for access. The selection process requires a continuing dialog with the user community to determine current needs, continuous scanning of available information resources, and judicious application of financial resources. Once selected, the items are collected, whether physically 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 resources 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. The invisibility of information work was identified long ago (Paisley, 1980), but the implications of this invisibility are only now becoming widely apparent.

Library and information services should be tightly coupled with other aspects of the information infrastructure of an organization (university, school, city government, corporation, etc.). But how do libraries provide a seamless infrastructure while maintaining visibility? How do they continuously respond to the evolution of their communities, or better yet, anticipate the evolution of the community's infrastructure as a means to provide the best resources and services? Libraries have a variety of stakeholders, including their users, their parent organizations (which are usually their primary funding source), other funding sources (foundations, donors, paying customers), and employees. Some stakeholders of libraries are par-

ticularly difficult to identify or characterize, such as the future users of their collections, many of whom have not yet been born.

Research Questions. The Invisible Infrastructure issues are summarized in the following questions:

- How visible are libraries to their stakeholders?
- How are the goals of stakeholders reflected in the library's goals?
- What are the consequences of visibility or lack of it?
- How can libraries be more visible to their stakeholders?

Research Design. This is a complementary set of research questions. We want to know how aware users are of the library and how embedded library services are in their practices. Similarly, it would be useful to know how aware other library stakeholders are of library services, and their expectations of those services. On the other hand, we also wish to know how aware the library is of user and stakeholder needs, plans, and strategic directions, and how well these goals are reflected in the library's plans.

The design presented here is tailored to university libraries, because many universities are currently developing their information infrastructures. However, most aspects of the design are stated in terms that could be adapted to other settings such as national, public, school, and special libraries. Some additional questions for public libraries are given at the end of this section.

Studies to address these research questions could be exploratory or descriptive in nature. Not enough is yet known about the problem to conduct an explanatory study such as an experiment (Babbie, 2001). Surveys and case studies are good starting points to address the challenge of libraries' visibility. Qualitative studies of users in their own environments also will be fruitful (Lofland & Lofland, 1995).

Sample. We could interview a representative sample of students, faculty, and staff of the university. Staff would have a range of perspectives, varying from administrative assistants to vice presidents (or vice chancellors or vice provosts, depending upon the organizational structure). Students' perspectives are likely to vary by disciplinary interests and degree objectives. Thus, stratified samples would be appropriate. Some questions will be adapted to different strata, such as asking questions of faculty about teaching and research, and of students about coursework and noncoursework needs.

Case studies of small groups or individuals could be fruitful. The sample could include a department in each of physical sciences, life sciences, technology, social sciences, humanities, and professional schools. Faculty from each group could be interviewed in their offices, looking closely at their information-related practices.

Selected Topics and Questions.   Services: When do you use library services? Please describe the most recent time you used any library services. What

prompted you to use them? What other types of information resources did you use? In what order did you use them? (Add probes to determine whether the library is used as "one-stop shopping," is the first stop, the stop of last resort, etc.) What services or sources of information in the library are most valuable to you? Least valuable? What do you most use that the library does not provide? What else should the library provide?

Infrastructure: Where does the library fit in the university? What do you think is the most important service it can provide or role it can play? If the library budget were to be cut by 20 percent, where would *you* make the cuts? If the budget were increased by 20 percent, what would you add? What role should the library play in teaching? In research? In supporting administrative activities?

Strategic planning: Data about the university's process of strategic planning for information infrastructure should be gathered and analyzed. Where does the library fit? How will resources be delivered to offices, classrooms, and off-site for distance-independent learning? What are the priorities for the university? Who is involved in strategic planning?

Additional questions for public libraries: What kind of strategic planning is the city, county, state, or other parent government doing for information technology and where do library services fit? What role does the community see for the library?

Challenge 2: Content and Collections

Until very recently, libraries were judged by their collections rather than by their services. Scholars sought out, and traveled to, the great collections of the world. The collections of major libraries are much more than the sum of their parts; disparate items are brought together, and relationships between items are identified. But what does it mean "to collect" in today's environment, when libraries provide access to content for which no physical artifact is acquired? The question is further complicated by the fact that access may be temporary for the term of a contract, rather than (relatively) permanent, as for purchased materials.

To explore the definition of "collection" in the networked information infrastructure, it is useful to return to Buckland's (1992) typology of the purposes for collections. These are 1. preservation (keeping materials for the future, as they may be unavailable if not collected at the time of their creation); 2. dispensing (providing access to their contents); 3. bibliographic (identifying what exists on a topic); and 4. symbolic (conferring status and prestige on the institution). The mapping of Buckland's typology to digital collections is not immediately obvious, and gives rise to several research questions in this area.

In recent years, much of the discussion of digital collections has come under the rubric of digital libraries (Lynch, 1999). "Digital libraries" is itself a contested term, as discussed in depth elsewhere (Borgman, 1999,

2000). In this article, the two-part definition established in Borgman et al. (1996) is assumed:

> Digital libraries are a set of electronic resources and associated techni-
> cal capabilities for creating, searching, and using information. In this
> sense they are an extension and enhancement of information storage
> and retrieval systems that manipulate digital data in any medium (text,
> images, sounds; static or dynamic images) and exist in distributed net-
> works. The content of digital libraries includes data, metadata that
> describe various aspects of the data (e.g., representation, creator, owner,
> reproduction rights), and metadata that consist of links or relationships
> to other data or metadata, whether internal or external to the digital
> library.

Digital libraries are constructed—collected and organized—by [and for] a community of users, and their functional capabilities support the information needs and uses of that community. They are a component of communities in which individuals and groups interact with each other, using data, information, and knowledge resources and systems. In this sense they are an extension, enhancement, and integration of a variety of information institutions as physical places where resources are selected, collected, organized, preserved, and accessed in support of a user community. These information institutions include, among others, libraries, museums, archives, and schools, but digital libraries also extend and serve other community settings, including classrooms, offices, laboratories, homes, and public spaces.

Implicit in this definition of digital libraries is a broad conceptualization of library "collections." One theme is that digital libraries encompass the full information life cycle: capturing information at the time of creation, making it accessible, maintaining and preserving it in forms useful to the user community, and sometimes disposing of information. With physical collections, users discover and retrieve content of interest; their use of that material is independent of library systems and services. With digital collections, users may retrieve, manipulate, and contribute content. Thus users are dependent upon the functions and services provided by digital libraries; work practices may become more tightly coupled to system capabilities.

A second theme implicit in the definition of digital libraries is the expanding scope of content that is available. Content now readily available in digital form includes primary sources such as remote sensing data, census data, and archival documents. Use of scientific data sets is computationally intensive, raising questions about the role the library should play in providing access to the resources and to the tools to use them (Lynch, 1999). Nor are scientific data the only challenge. As more archives and special collections are digitized, many primary sources in the humanities are becoming more widely available online than are secondary sources such as books and journals. Distinctions between "primary and secondary sources"

are problematic, however, as they vary considerably by discipline and by context. Some sources may be primary for some purposes and secondary for others. Here I oversimplify the terms by referring to raw data and to unique or original documents as primary sources and to analyzed or compiled data and to reports of research as secondary sources.

A third theme is the need to maintain coherence of library collections (Lynch 1999). Descriptions (and sometimes content) of journal articles, for example, can be found in catalogs, indexing and abstracting databases, and digital libraries. Users want to identify articles of interest and to move seamlessly from bibliographic references to the full text, and from references in those texts directly to the full content of the cited articles. Sometimes they also wish to link directly to primary sources on which the articles are based. Supporting these uses of journal-related information requires various forms of links within and between many independent catalogs, databases, and digital libraries.

Efforts at improving the coherence of collections include the Cross-Ref initiative (http://www.crossref.org) developed by a consortium of major scholarly publishers to link citations using Digital Object Identifiers (http://www.doi.org), and the Open Archives Initiative (OAI). CrossRef allows users to follow citations across the boundaries of individual publishers, while the OAI enables libraries to make their digital collections more widely available in a standard form (Lagoze & Van de Sompel, 2001; http://www.openarchives.org). Coherence always has been a problem in the print world, however. Catalogs of a library's collections typically contain entries only for about 2 percent of the individual items a user might seek, based on Tyckoson's clever assessment of some years back (Tyckoson, 1989). For the rest, library users are dependent upon indexing and abstracting databases, finding aids, various locally developed tools and arrangements (such as shelves for new books, or shelves organized by genre, as are common in public libraries), and the knowledge of librarians. However, even the concept of catalogs is changing as libraries merge records on their own holdings with records from indexing and abstracting databases and with records for online resources external to the collection. The use of Web-based portals or gateways is another step toward coherence. A portal can bring together in one place the many types of resources and finding aids offered by the library—a goal that was difficult to accomplish in the print environment.

The Content and Collections challenge outlined here is a subset of a larger set of concerns about how to evaluate digital libraries. Research, planning, and deployment of digital libraries all can benefit from evaluation—whether formative, summative, iterative, or comparative. Evaluation efforts can have substantial benefits to digital library development by focusing designers on measurable goals, by providing data on which to reassess those goals, and by assessing outcomes. An array of methods and measure-

ment issues was identified in a recent workshop on digital library evalua-
tion that was jointly sponsored by the National Science Foundation and the
European Union (Borgman, 2002).

Research Questions. The goal is to determine the nature of collections
and their role in the information infrastructure of parent organizations
(universities, governments, corporations, etc.), nations, and the world. We
can address the Content and Collections challenge via the following re-
search questions:

- What are "collections"?
- How are collections used?
- How can communities and collections best be matched?
- How can the coherence of collections be established and maintained?

Research Design. Multiple methods will be required to approach this
array of research questions. One approach is to conduct interviews, surveys,
and case studies of users and librarians to determine their views on these
questions and to study actual uses of collections. A particularly effective ap-
proach is to interview faculty in their offices, looking closely at their infor-
mation-related practices. This is one of a number of approaches we are
taking in the ADEPT project,[1] which is studying the use of digital libraries
for teaching undergraduate courses in geography (Borgman, et al., 2000).

A complementary approach is to document the nature of extant col-
lections (physical and digital), the metadata that exists for them, and the
functions and services available to support them.

Sample. For the behavioral and policy questions, we could interview a
representative sample of students, faculty, and staff of one or more univer-
sities, as proposed for the first challenge. A similar stratified sample that
reflects disciplinary interests and degree objectives would be appropriate.
However, smaller samples for more in-depth interviews would be needed
for this set of studies. Some questions are best addressed to library staff,
although in many cases it would be beneficial to address similar questions
about collections to librarians and to users.

Case studies of small groups or individuals may be especially fruitful,
as behavioral studies of information use tend to be detailed and labor-in-
tensive to conduct. Content, collections, uses, and users vary considerably
by discipline, so multiple studies with different samples would be required.

Collection studies could be approached in several ways. Samples could
be drawn from the obvious collections, such as books, journals, and on-
line databases to which the library subscribes. The surveys and interviews
should yield some definitions of what a library's users view as collections.
These are likely to include locally developed resources outside the purview
of the library (e.g., survey data, scientific data, collections of models assem-
bled for research projects). Any or all of these collections could be sam-
pled for study, with the goal of determining how well the data, metadata,

functionality, and services match the expressed needs of the user communities they are intended to serve.

   Selected Topics and Questions. Buckland's (1992) typology of collection purposes generates some framing questions for both the behavioral and documentary approaches to researching this challenge: 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 printed materials?

   Research that asks the respondents to define basic concepts is particularly difficult, for it risks leading the respondent to a desired outcome. An approach that Caidi (2001) found effective in getting respondents to define "information infrastructure" was to offer them a list of distinct definitions (she used four different definitions of "infrastructure") and to ask them to explain which of them best reflects their own understanding of the concept. The respondents were able to expand upon one or more of the definitions to arrive at their own conception. A similar approach might be particularly effective in eliciting definitions of "collection" from information seekers and from library staff.

   Several questions should address the "information life cycle": What do people do with information resources once they have them? How do they use them? Do they write new documents (articles, books, music, art, performances, etc.)? Do they publish online and/or offline? Do they use the resources to read, research, prepare for exams, get a job, invest, or make health decisions?

   Multiple groups should be asked about their definitions of collections, the value of collections, criteria for selection, and how they use collections. Some additional questions can be tailored to individual groups, such as the following:

   Faculty: Ask questions about the use of collections for teaching and research. How do they collect and organize resources for their courses? Where do they get new materials? How do they make them available to students (e.g., as texts, course readers, library reserves, electronic reserves, Web sites)? Who assists them in collecting and organizing resources now? Who should do so in the future? What balance of primary and secondary sources do they use? How does their use of collections vary between teaching and research?[2]

   Faculty should also be asked about their engagement in research projects to construct collections of digital resources for their fields. Increas-

ingly, research groups are assembling portals that aggregate a range of re-
sources for a research problem. Digital library projects within individual
disciplines of the sciences, social sciences, and humanities are producing a
wealth of new and innovative resources for teaching and research (http://
www.dli2.nsf.gov). However, these projects tend to address technical aspects
of constructing digital libraries, rather than their use for teaching and re-
search. Much more needs to be known about the uses and users of such
collections.

   Students: Ask questions to determine the collections they use and for
what purposes. When do they go to Google and other Internet search en-
gines? When do they use library or other university sources? How does
online availability enhance or constrain their information seeking? How
much of their collection use is for course-related vs. non–course related
purposes?

   Library staff: Librarians and other library staff may make fine-grained
distinctions between types of collections and uses of them, given their pro-
fessional education and experience. How do they define collections? What
are their criteria for selection, preservation, authority, authenticity, etc.?

Challenge 3: Preservation and Access

   While little agreement may exist on the definition of a library "collec-
tion," most librarians would agree that the collections must be preserved
so that they remain accessible. Portions of physical collections are crum-
bling, and libraries are undertaking cooperative efforts to preserve the
content, physical artifacts, or both. Preservation of digital collections is yet
more complex and potentially even more expensive than preserving printed
resources. Most printed volumes will survive via "benign neglect," provid-
ed they are shelved under adequate climate controls. Digital resources must
be continually migrated to new software and new technologies, thus active
management is required for preservation (Smith, 1999). When a library
owns the rights to the digital content, the library presumably is responsi-
ble for maintaining continual access, absent other cooperative agreements.
When a library is leasing access to digital content, responsibility for preser-
vation may be diffuse. Authors are unlikely to take responsibility and, even
if they might wish to do so, may not have the legal authority if they have
assigned copyright to the publisher. Publishers wish to maintain control,
but few are willing to assure long-term continuous access. Even if they were
willing, the rate of acquisitions and mergers in the publishing industry sug-
gests that long-term commitments may be difficult to enforce. Recently,
publishers have expressed more interest in allowing libraries to maintain
digital content, but the economic model under which libraries might ac-
cept such responsibility is not clear (National Research Council, 2000; Yakel,
2001). Third parties such as OCLC are now stepping to the fore as reposi-
tories, which is a promising model (http://www.oclc.org).

"Access" is a term that is widely used in our field but rarely defined. It incorporates aspects of freedom, ability, connectivity, usability, and rights. Elsewhere (Borgman, 2000, p. 57), I defined "access to information" as "connectivity to a computer network and to available content, such that the technology is usable, the user has the requisite skills and knowledge, and the content itself is in a usable and useful form."

In my initial framing of the challenge of preservation and access (Borgman, 2000, chapter 7), I focused primarily on the library's role in preserving digital resources. Preservation and access are critical public policy concerns in which libraries should have a voice, as social institutions with substantial responsibility for maintaining access to their institutions' and nations' informational and cultural heritage. Deanna B. Marcum (in this issue of *Library Trends*) ably addresses the challenges faced by libraries in this arena. Thus, I turn my attention to the challenge of long-term access to online content and the concerns for persistence of content in national and international information infrastructures.

Online resources are most commonly identified by URLs (Uniform Resource Locators), (Berners-Lee, Masinter, & McCahill, 1994). URLs identify a location, rather than a document, and thus are far less stable than bibliographic references. Persistence issues associated with URLs are best explained by example. My home page currently resides at this URL: http://is.gseis.ucla.edu/cborgman/. This is the fourth URL for my home page in the last five years. The URL has varied due to changes in the department name ("dlis" to "is") and to variations in local conventions such as the use of computer names in URLs (e.g., "skipper") and internal hierarchy (e.g., "/faculty/~cborgman"). The content of my home page is updated periodically, with new entries added and new documents posted. The links to those documents sometimes change, due to location changes or to changes in the status of the document (e.g., from draft to published). Documents are sometimes superseded by more current versions with different names and locations. The software in which the documents are written and posted includes various versions of Corel WordPerfect, Microsoft Word, and PDF. This simple and common example typifies the array of persistence problems related to the use of URLs:

- *Location changes:* the home page is at a new URL; documents linked from the homepage move to different URLs.
- *Content changes:* the home page address is the same but the content has changed; documents are updated without changing name or location.
- *Format changes:* the document is migrated to a new software format; the intellectual content may be the same, but the documents are no longer "bit for bit" identical.
- *Status changes:* the document content is no longer current; it may have been superseded by another document at another location, and may or may not be linked to the subsequent document.

The lack of persistence of URLs becomes increasingly problematic as people rely more heavily on online documents. Yet, we know little about how individuals and organizations cope with these problems. What are users' expectations for stable access to online documents? They probably expect home pages to be updated, but they probably also expect to find the same individual document at the same URL the next time they visit. Some of these problems are being addressed by new forms of identifiers such as URIs and URNs (Berners-Lee et al., 1994; Berners-Lee et al., 1998), but none claims to be a universal solution, nor are they widely implemented. The proposed OpenURL standard (Van de Sompel & Beit-Arie, 2001; http://library.caltech.edu/openurl/) provides context-sensitive linking and supports the CrossRef/DOI (digital object identifier) initiative of major publishers. As of this writing, the OpenURL approach is being implemented in commercial software for library applications and appears promising for some aspects of the URL persistence problem.

Bibliographic references are far more stable than URLs, but still have some of these persistence issues. Catalogers control variations by establishing relationships between items, works, and manifestations, and by establishing cross references between related works or editions (Leazer, 1994; Svenonius, 2000; Tillett, 1991, 1992). The cataloging approach may work within a closed network of cooperating libraries, but Webmasters and writers and publishers of online documents are not bound by cataloging practices or other sets of consistent rules. The costs of creating cataloging records usually are deemed justified for printed documents that libraries will hold indefinitely. However, the cost of creating cataloging or metadata records for every electronic document may be prohibitive. The information science research community is revisiting the age-old question of when to invest in description at the time of record creation and when to invest in improved retrieval techniques for use at the time that information is sought (Liddy, et al., 2002). Automatic indexing may prove sufficient for retrieval by elements that exist in the record, but extrinsic metadata, such as intellectual property rights and the provenance of electronic records, also may be needed. All of these metadata choices will influence the persistence of electronic documents.

Research Questions. Preservation and access of online documents is a challenge being tackled aggressively by technical and policy organizations such as the Internet Engineering Task Force (http://www.ietf.org), and the World Wide Web Consortium (http://www.w3c.org), and by the library and information science community (Marcum, this issue; Yakel, 2001). The LIS community can contribute productively to these discussions by addressing the following research questions:

* To what degree does the lack of stability of online documents, and links to online documents, influence preservation of, and access to, library resources?

- To what degree are users, seekers, and producers of digital resources aware of online persistence issues?
- How do users, seekers, and producers of digital resources address persistence problems?
- How might library methods for organization of knowledge be employed to improve the stability of access to online resources? How might other organization of knowledge practices be employed, such as those from the archival and museum communities? What are the implications for persistence when little, if any, metadata are associated with documents?

Research Design. Studies to address these questions will require a combination of analyzing the use of library resources (research question #1); interviewing users, seekers, and producers of Internet resources about their activities and practices (questions #2 and #3); and theoretical and empirical studies of knowledge organization principles (question #4).

The first research question could be addressed by studies of a library's collections to identify the distribution of digital resources that are under the library's control (e.g., locally managed digital libraries), that are partially under the library's control (e.g., in commercial databases for which access is leased), and those over which the library has minimal control (e.g., on the World Wide Web or other Internet source). The studies should assess how much each of the resources depends upon URLs, URIs, or other identifiers such as Digital Object Identifiers, ISBN, ISSN, etc.

Research question #4 could follow the models of prior research on document relationships conducted by Gilliland-Swetland (2000), Leazer (1994), Svenonius (2000), and Tillett (1991, 1992). Research questions #2 and #3 require user studies similar to those outlined in the first two challenges.

Sample. The most comprehensive approach to addressing the persistence problem in preservation and access would be to study all four questions within one institution. In that way, the array of available resources could be compared with the practices of those who use them, and with the organizational methods applied. Alternative approaches are to address each of the four research questions across multiple institutions, or to address each question individually. Research question #4 is most easily separated from the other three, as libraries apply reasonably consistent knowledge organization practices—at least within a given country. Multinational comparisons of organizational practices also would be valuable.

Samples for the surveys and case studies (research questions #2 and #3) could be drawn in the same way as in the first two challenges. However, it may also be necessary to study the practices of Webmasters (inside and outside the institution), and writers and publishers of online resources who are outside the institution but whose resources are used by people within the institution under study.

Selected Topics and Questions. Preservation and access of library resources:
Conduct a "collection analysis" of digital resources to which the library
provides access. How is persistence maintained in each of these resources?
What identifiers are employed? What are the principles underlying each
type of identifier? How stable are the identifiers? What data are available
on the persistence of identifiers? These data might be provided by purvey-
ors of digital libraries, by search engines, by the Internet Archive (also
known as the *Wayback Machine*) (http://www.archive.org/) from other stud-
ies, and by collecting transaction data from university servers.

User awareness of online persistence problems and user coping mechanisms: Ask
users and seekers of digital resources about their experiences in locating
information online. How often do they encounter incorrect addresses for
resources? How often do they encounter links to new addresses where doc-
uments were moved? When they find incorrect addresses, what do they do?
Do they search for the resources using other tools? Do they try to find the
original source or pointer to the site? How do they identify sites of inter-
est? To what extent do they rely on bookmarks, search for known sites, or
rely on links provided by others? Do they download or otherwise capture
content of interest to preserve it locally, in anticipation of dead links? Do
they attempt to verify if the content is the same as was sought? If so, what
are their methods and criteria? How do they use the links once found? Do
they maintain a record of dates visited, for example?

Similar sets of questions can be asked of those who maintain Web sites
and write for online publication. What are their practices for assuring per-
sistence of their content? What are their criteria for updating existing doc-
uments, for creating new documents, and for indicating when and what
type of changes have been made to a document? What address mecha-
nisms do they employ? How often do documents change address, and
under what circumstances? The answers to these questions are likely to
vary widely by genre, so multiple studies should be conducted. Web sites
maintained by libraries, archives, and museums are likely to have more
sophisticated practices than sites for Weblogs ("blogs"), political protests,
or fan clubs, for example. A useful approach would be to determine the
distribution of sites visited by the user community and then to segment
the study of sites accordingly.

Organizational methods to address persistence problems: These studies will be
informed by results from the prior studies on the distribution of digital
resources, on how online resources are used, on the types of problems
encountered, and on users' approaches to dealing with these problems.
Models for improving Web organization, such as "the semantic Web" (http:/
/www.w3c.org), should be analyzed from a persistence perspective. The
various representation models employed by libraries, archives, and muse-
ums should be examined for lessons about persistence that can be applied
to organization of online sources. While global solutions would require

coordinated, long-term approaches, libraries and other institutions can seek methods to improve the persistence of their own resources now.

Challenge 4: Institutional Boundaries

My original framing of issues associated with institutional boundaries focused on relationships between libraries, archives, and museums (Borgman, 2000). These three information institutions face similar concerns, such as the risk of becoming an invisible part of the infrastructure, the changing nature of collections, and preservation and access for content and artifacts. The distinction between these information institutions was not well established until the late twentieth century (Rayward, 1993). Until then, 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, multidisciplinary intellectual interests, without the division between the sciences and the humanities that we take for granted today.

Much of the distinction between these institutions is based on the type of material collected. 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 such as the history of a particular automobile. These distinctions by type of material become less useful as more content exists in a common form, namely digital. Furthermore, partitioning intellectual content among these three sets of institutions is an artificial division of the natural world that does not necessarily serve the information seeker well.

In a world of physical materials, access was determined by physical space: users had to decide which building to enter. Access mechanisms (catalogs, finding aids, museum directories) were located inside the buildings. Now the access mechanisms for many collections are available on-line; users can browse the holdings of libraries, archives, and museums, and even "visit" virtual museum collections. Search engines such as Google (http://www.google.com), AltaVista (http://www.altavista.com), Alexa Internet (http://www.alexa.com), and one of the newest, Teoma (http://www.teoma.com) do not distinguish between these institutions or between institutions and individuals, for that matter. Topical searches in these engines produce matches from across the spectrum of public and private, commercial and nonprofit, scholarly and personal opinion, published and unpublished, and formal and informal sources.

Paradoxically, the holdings of information institutions are often the least visible to Internet search engines. This is known as the "dark Web" problem (Lynch, 2001). Search engines generally can capture content only on static Web pages. The contents of library catalogs are stored in data-

bases. Web pages of search results are generated dynamically for each query; they do not exist in a static form that search engines can capture. Thus, a Google search on "Shakespeare" may retrieve sites that specialize in Shakespearean memorabilia (as described in their Web pages), sites of theaters that are currently performing Shakespearean plays, and Shakespeare fan clubs, but usually will not retrieve catalog records for books in libraries or for records in archives. Harvesting models, such as the Open Archives Initiative (Lagoze & Van de Sompel, 2001) will solve part of the dark Web problem. The dark Web encompasses not only the catalogs, finding aids, and directories of information institutions, but also the vast intranets that are hidden behind firewalls of many corporations, governments, and other organizations. The Internet consists of a mix of public and private sites, and search engines actually index only a small proportion of all extant Web pages.

While the broad retrieval by search engines such as Google offers many new opportunities for information seekers (and is extremely popular), in some respects it represents a step backwards from traditional approaches to knowledge organization. One of the most fundamental problems with Internet navigation is the lack of context for the search (Furner, 2002; Solomon, 2002). The Internet is being used to find sites, sources, services, documents, people, and activities that would be located by diverse offline mechanisms, if at all—library catalogs, phone directories, museums, archives, travel agents, government agencies, encyclopedias, directories of persons, etc. In most other information retrieval situations, context is provided by segmenting the database being searched or by constraining the meaning of terms within the database.

Although the context for a search may be obvious to the user, search engines can operate only with the terms they receive. A user who is planning a European trip may type "Paris" into a search engine. He or she probably expects to retrieve tourist information on the city of Paris, France, but how does a search engine know that? A student studying the *Iliad* more likely wants to know about the Greek hero after which the city of Paris was named. In other contexts, someone who enters "Paris" as a search term may be seeking a source for plaster of Paris, movies that contain the word "Paris" in the title, people with the first or family name of Paris, or historical, economic, or political perspectives on the city.

Thus, the challenge of institutional boundaries has several components. One component is the fuzzy lines between types of information institutions. A second is the fading of boundaries between institutional sources for discovering information resources. Searchers may make little distinction between searching the resources of libraries, archives, museums, corporate, or other organizations. Third is the difficulty of establishing context for searching. Coherence of collections, as discussed in the second challenge,

is difficult when framed in terms of the resources offered by an individual library. How does a library provide a "coherent user experience" to a community that has access to a vast array of resources beyond the library?

Research Questions. The blurring of boundaries between information institutions and between information institutions and other sources of collections and services raises new questions about the visibility and role of libraries. Many of these are policy questions, and all will be informed by the results of studies on the prior three challenges presented.

• What are the roles of information institutions in providing access to information?
• Where do institutions add value to information resources and services?
• What forms of cooperation and alliances between institutions are most beneficial, and for what purposes?
• How is context best provided in information seeking and use?

Research Design. The first three of these research questions are addressed most directly by policy research, and the fourth question also has policy components. Studies in response to the earlier challenges may provide baseline data and may identify some of the criteria for assessing roles and value. We can conduct documentary studies of the role of libraries and other information institutions in various local, regional, national, international, and cultural contexts. We can interview stakeholders with policy responsibilities, such as senior managers in government funding agencies, in universities, and in corporations. The third question can be addressed by studying the history of cooperation within and between these institutions, and by looking more broadly at other types of cooperation models. The fourth question is a mix of policy, technology, and behavioral studies. Context might be provided via institutional, technical, or business models.

Sample. As noted in the research design, we would examine various literature and policy documents about the roles of these institutions, and would interview a wide variety of stakeholders. Libraries, museums, and archives that have overlapping user communities should be studied together to address some of these questions.

Selected Topics and Questions. Policy questions: We tend to assume that libraries, museums, and archives serve overlapping communities. But what degree of overlap does exist, and in what areas? What roles do the stakeholders of each institution think are most important? What priorities do they ascribe to these roles? Many of the functions provided by these institutions require large amounts of invisible work, such as selecting, collecting, organizing, preserving, and conserving resources so that they are accessible. What priorities should be set for the invisible work of libraries? Who should do this work? Which parts are essential? Which parts are expendable? Which could be accomplished by more cost-effective means? Which require greater investments? Which functions could be disaggregated and

divided between institutions and which are most effective when aggregated (Fuller, 2002)?

   Context: Some of these questions are technical and are being addressed by the information studies community already (Furner, 2002; Solomon, 2002). Others can be addressed as behavioral or policy questions. When is it effective to segment user needs by institution? Will search engine models that allow users to categorize questions be effective? What if the categorization is source based (e.g., telephone number, restaurant review, medical dictionary, library catalog)? What are other models that might be effective?

SUMMARY AND CONCLUSIONS

   Libraries are an essential component of a nation's information infrastructure, yet they are rarely mentioned in the public-policy documents that define and frame such infrastructures. They often are invisible to their users and to their stakeholders. The library community is responsible for identifying its goals for local, national, and global information infrastructures and to act upon them. In this paper, I have presented four challenges faced by libraries and have proposed research designs to address each of them. The four challenges involve 1. invisible infrastructure, 2. content and collections, 3. preservation and access, and 4. institutional boundaries. While these were first identified in an earlier publication (Borgman, 2000), here I have expanded and updated them, proposed research designs to explore the challenges, and sought to complement other articles in this special issue.

   The challenge involving invisible infrastructure is the broadest of the four, and is a theme that runs through the other three. Libraries risk being victims of their own success, as good design and good service tend to be unobtrusive. The research questions posed for this challenge address how visible libraries are to their users and other stakeholders—but also how well stakeholders' goals are represented in library plans and policies. Visibility cuts both ways.

   The second challenge, of content and collections, addresses the problem of defining the concept of a "collection" in an environment where libraries provide access to a wide array of content that they may or may not possess. Research questions in this arena ask users and stakeholders to define what they mean by "collection," and ask about how they use various forms of collections and content. The coherence of collections that include diverse resources and serve diverse audiences is of particular concern.

   Preservation and access, the third challenge, is the most expanded from its earlier incarnation. I took that liberty because the challenge for library collections is being addressed in another article in this issue by Deanna B. Marcum, one of the most knowledgeable experts on the topic. Instead, I focus on the stability of access to online resources that are of value to a library's users, but over which the library may have little control. Research

questions in this area address user behavior with regard to persistence; the relationship between persistence, preservation, and access; and knowledge organization methods that might improve persistence.

The fourth challenge, involving institutional boundaries, is also expanded from its original framing, coming full circle to the challenge of invisibility. Not only are the boundaries blurring between three preeminent types of information institutions—libraries, museums, and archives—but the boundaries are blurring between the collections and services provided by these institutions and other entities. Search engines are both a blessing and a curse in this regard. They provide global searching capabilities while stripping those same searches of their context. Research questions in this arena focus on identifying roles of each institution, relationships between them, and ways to aggregate and disaggregate various functions.

The four challenges are intertwined and research on each of them will inform the others. I have proposed a mixture of research methods that includes surveys, case studies, documentary analyses, and policy analyses. Participation in these studies would be sought from users of information services, writers and publishers of content, stakeholders in parent organizations, and policy-makers far removed from libraries. While most of the studies are framed in terms of individual libraries, universities, or geographical regions, the designs are intended to be adaptable to larger and smaller units. I hope the guidance provided will encourage a wide range of information studies scholars and librarians to pursue research in these areas, for it is much needed. Only with a better understanding of these challenges can libraries find their best fit in the information infrastructure of our networked world.

ACKNOWLEDGEMENTS

NOTES
1. The ADEPT Web sites at UCLA (http://is.gseis.ucla.edu/adept/) and UCSB (http://www.alexandria.ucsb.edu/adept/) provide links to continuing research reports. The project is funded from 1999 to 2004 by the National Science Foundation's Digital Libraries Initiative (http://www.dli2.nsf.gov), grant no. IIS-9817432.
2. We are currently addressing these issues with geography faculty as part of the ADEPT project. See forthcoming work by Borgman, C. L.; Smart, L. J.; Millwood, K.; and Finley, J.

REFERENCES

Babbie, E. (2001). *The practice of social research* (9th ed.). Belmont, CA: Wadsworth.

Berners-Lee, T., Fielding, L., & McCahill, M. (1998, August). RFC 2396 Uniform Resource Identifiers (URI): Generic Syntax. Available at http://www.rfc-editor.org/.

Berners-Lee, T., Masinter, L., & McCahill, M. (1994, December). RFC 1738: Uniform Resource Locators (URL). (See also RFC1808, RFC2368, and RFC2396 at http://www.rfc-editor.org/).

Borgman, C. L. (1999). What are digital libraries? Competing visions. In G. Marchionini & E. Fox (Eds.), [Special issue] Progress toward digital libraries. *Information Processing & Management, 35*(3), 227–243.

Borgman, C. L. (2000). *From Gutenberg to the global information infrastructure: Access to information in the networked world.* Cambridge, MA: The MIT Press.

Borgman, C. L. (2002). Final report to the National Science Foundation. Fourth DELOS Workshop. Evaluation of Digital Libraries: Testbeds, Measurements, and Metrics. Hungarian Academy of Sciences, Computer and Automation Research Institute (MTA SZTAKI), Budapest, Hungary, 6–7 June 2002. Grant IIS-0225626. Retrieved on March 20, 2003, from http://www.dli2.nsf.gov/internationalprojects/working_group_reports/evaluation.html.

Borgman, C. L., Bates, M. J., Cloonan, M. V., Efthimiadis, E. N., Gilliland-Swetland, A., Kafai, Y., Leazer, G. L., & Maddox, A. (1996). *Social Aspects Of Digital Libraries.* Final Report to the National Science Foundation; Computer, Information Science, and Engineering Directorate; Division of Information, Robotics, and Intelligent Systems; Information Technology and Organizations Program. Award number 95–28808. Retrieved on March 20, 2003, from http://is.gseis.ucla.edu/research/dl/index.html.

Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. L., Mayer, R., Gwynn, D., Gazan, R., & Mautone, P. (2000). Evaluating digital libraries for teaching and learning in undergraduate education: A case study of the Alexandria Digital Earth Prototype (ADEPT). *Library Trends, 49*(2), 228–250.

Buckland, M. K. (1992). *Redesigning library services: A manifesto.* Chicago: American Library Association.

Caidi, N. (2001). The information infrastructure as a discursive space: A case study of the library community in Central and Eastern Europe. Ph.D. Dissertation, Information Studies, University of California, Los Angeles, 2001. Dissertation Abstracts International, 62, 3609.

Europe and the global information society: Recommendations to the European Council. (1994). Brussels: European Council. (Known as "The Bangemann Report.")

Fuller, S. (2002). *Knowledge management foundations.* Boston: Butterworth-Heinemann.

Furner, J. (2002). On recommending. *Journal of the American Society for Information Science and Technology, 53*(9), 747–763.

Gilliland-Swetland, A. J. (2000). *Enduring paradigms, new opportunities: The value of the archival perspective in the digital environment.* Washington, D.C.: Council on Library and Information Resources.

G-7 (Group of Seven) Ministerial Conference on the Information Society. (1995). *Chair's conclusions.* Retrieved on March 20, 2003, from http://web.archive.org/web/20000411003002/www.ispo.cec.be/g7/keydocs/G7en.html.

Lagoze, C., & Van de Sompel, H. (2001). The Open Archives Initiative: Building a low-barrier interoperability framework. In E.A Fox & C.L. Borgman (Eds.), *Proceedings of the Joint Conference on Digital Libraries, June 20–24, 2001, Roanoke, VA* (pp. 54–62). New York: ACM Press.

Leazer, G. H. (1994). A conceptual schema for the control of bibliographic works. In D. L. Andersen, T. J. Galvin, & M. D. Giguere (Eds.), *Navigating the networks: Proceedings of the ASIS Mid-Year Meeting, Portland, Oregon, May 21–25, 1994* (pp. 115–135). Medford, NJ: Learned Information.

Liddy, E. D., Gay, G., Harwell, S., & Finneran, T. (2002) A modest (metadata) proposal. Panel presentation, "How important is metadata?". In G. Marchionini & W. Hersh (Eds.), *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries,* p 369. July 14–18, 2002, Portland, OR. New York: ACM.

Lofland, J., & Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis* (3rd ed.). Belmont, CA: Wadsworth.

Lynch, C. A. (1999). Today and tomorrow: What the digital library really means for collections and services; Virtually yours: Models for managing electronic resources and services. In P. Johnson & B. MacEwan (Eds.), *Proceedings of the Joint Reference and User Services Association and Association for Library Collections and Technical Services Institute, Chicago, Illinois, October 23–25, 1997* (pp. 85–97). Chicago, IL: American Library Association, 1999.

Lynch, C. A. (2001). Metadata harvesting and the Open Archives Initiative. *ARL Bimonthly Report, 217,* 1–9.

Marcum, D. (2003). Research questions for the digital era library. *Library Trends 51*(4).

National information infrastructure: Agenda for action. (1993). Retrieved on March 20, 2003, from http://real.utk.edu/FINS/Information_Infrastructure/Fins-II-09.txt.

National Research Council; Commission on Physical Sciences, Mathematics, and Applications; Computer Science and Telecommunications Board; Committee on Intellectual Property Rights and the Emerging Information Infrastructure. (2000). *The digital dilemma: Intellectual property in the information age.* Washington, D.C.: National Academy Press.

National Research Council; Commission on Physical Sciences, Mathematics, and Applications; Computer Science and Telecommunications Board; NRENAISSANCE Committee (1994). *Realizing the information future: The Internet and beyond.* Washington, D.C.: National Academy Press.

Paisley, W. J. (1980). Information and work. In B. Dervin & M. J. Voigt (Eds.), *Progress in the communication sciences* (Vol. 2, pp. 114–165). Norwood, NJ: Ablex.

Rayward, W. B. (1993). Electronic information and the functional integration of libraries, museums, and archives. In S. Ross & E. Higgs (Eds.), *Electronic information resources and historians: European perspectives* (pp. 227–243).St. Katharinen: Scripta Mercaturae Verlag.

Smith, A. (1999). *Why digitize?* Washington, D.C.: Council on Library and Information Resources.

Solomon, P. (2002). Discovering information in context. In B. Cronin (Ed.), *Annual Review of Information Science and Technology, 36* (pp. 229–264). Medford, NJ: Information Today.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research, 7*(1), 111–134.

Svenonius, E. (2000). *The intellectual foundation of information organization.* Cambridge, MA: MIT Press.

Tillett, B. B. (1991). A taxonomy of bibliographic relationships. *Library Resources & Technical Services, 35,* 150–159.

Tillett, B. B. (1992). Bibliographic relationships: An empirical study of the LC machine-readable records. *Library Resources & Technical Services, 36,* 162–188.

Tyckoson, D. A. (1989). The 98% solution: The failure of the catalog and the role of electronic databases. *Technicalities, 9*(2), 8–12.

Van de Sompel, H., & Beit-Arie, O. (2001). Open linking in the scholarly information environment using the OpenURL framework. *D-Lib Magazine, 7*(3). Retrieved on March 20, 2003, from http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html.

Yakel, E. (2001). Digital preservation. In M. E. Williams (Ed.), *Annual review of information science and technology, vol. 35* (pp. 337–378). Medford, NJ: Information Today.