OCCASIONAL PAPER

8

DIGITAL LIBRARIES:
DEFINITIONS,
ISSUES AND CHALLENGES

Gary Cleveland
IFLA UDT Core Programme

IFLA

March, 1998

International Federation of Library Associations and Institutions
UNIVERSAL DATAFLOW AND TELECOMMUNICATIONS
CORE PROGRAMME

The IFLA Core Programme on Universal Dataflow and Telecommunications (UDT) seeks to facilitate the international and national exchange of electronic data by providing the library community with pragmatic approaches to resource sharing. The programme monitors and promotes the use of relevant standards, promotes the use of relevant technologies and monitors relevant policy issues in an effort to overcome barriers to the electronic transfer of data in library fields.

CONTACT INFORMATION

Mailing Address:

> *IFLA International Office for UDT*
> *c/o National Library of Canada*
> *395 Wellington Street*
> *Ottawa, CANADA*
> *K1A 0N4*

UDT Staff Contacts:

> Leigh Swain, Director
> Email: leigh.swain@udt.ifla.org
> Phone: (819) 994-6833

> *or*

> Louise Lantaigne, Administration Officer
> Email: UDT@udt.ifla.org
> Phone: (819) 994-6963

> Fax: (819) 994-6835

> Email: UDT@udt.ifla.org

> URL: http://www.ifla.org/udt/

Occasional papers are available electronically at: http://www.ifla.org/udt/op/

DIGITAL LIBRARIES:
Definitions, Issues and Challenges

Gary Cleveland
UDT Core Programme
gary.cleveland@udt.ifla.org   March, 1998

The idea of easy, finger-tip access to information—what we conceptualize as digital libraries today—began with Vannenar Bush's Memex machine (Bush, 1945) and has continued to evolve with each advance in information technology. With the arrival of computers, the concept centered on large bibliographic databases, the now familiar online retrieval and public access systems that are part of any contemporary library. When computers were connected into large networks forming the Internet, the concept evolved again, and research turned to creating libraries of digital information that could be accessed by anyone from anywhere in the world. Phrases like "virtual library," "electronic library," "library without walls" and, most recently, "digital library," all have been used interchangeably to describe this broad concept.

But what does this phrase mean? What is digital library? And what are the issues and challenges in creating them? Moreover, what are the issues involved in creating a coordinated scheme of digital libraries? It has been suggested that digital libraries will only be viable within such a scheme (Chapman and Kenny, 1996). This paper provides a very high-level overview of digital libraries and briefly outlines each of these questions in turn.

1. WHAT IS A DIGITAL LIBRARY?

What is a digital library? There is much confusion surrounding this phrase, stemming from three factors. First, the library community has used several different phrases over the years to denote this concept—electronic library, virtual library, library without walls—and it never was quite clear what

each of these different phrases meant. "Digital library" is simply the most current and most widely accepted term and is now used almost exclusively at conferences, online, and in the literature.

Another factor adding to the confusion is that digital libraries are at the focal point of many different areas of research, and what constitutes a digital library differs depending upon the research community that is describing it (Nurnberg, et al, 1995). For example:

- from an information retrieval point of view, it is a large database
- for people who work on hypertext technology, it is one particular application of hypertext methods
- for those working in wide-area information delivery, it is an application of the Web
- and for library science, it is another step in the continuing automation of libraries that began over 25 years ago

In fact, a digital library is all of these things. These different research approaches will all add to the development of digital libraries.

Third, confusion arises from the fact that there are many things on the Internet that people are calling "digital libraries," which—from a librarian's point of view—are not. For example:

- for computer scientists and software developers, collections of computer algorithms or software programs are digital libraries.
- for database vendors or commercial document

suppliers, their databases and electronic document delivery services and digital libraries.

- for large corporations, a digital library is the document management systems that control their business documents in electronic form.
- for a publisher, it may be an online version of a catalogue.
- and for at least one very large software company, a digital library is the collection of whatever it can buy the rights to, and then charge people for using.

A fairly spectacular example of what many people consider to be a digital library today is the World Wide Web. The Web is a gathering of thousands and thousands of documents. Many would call this huge collection a digital library because they can find information, just as they can do banking in a "digital bank" or buy compact discs in a "digital record store." Yet, is the Web a digital library? According to Clifford Lynch, once of the leading scholars in the area of digital library research, it is not. Lynch (1997:52) states:

> *One sometimes hears the Internet characterized as the world's library for the digital age. This description does not stand up under even casual examination. The Internet¾ and particularly its collection of multimedia resources known as the World Wide Web¾ was not designed to support the organized publication and retrieval of information as libraries are. It has evolved into what might be thought of as a chaotic repository for the collective output of the world's digital "printing presses.".... ...In short, the Net is not a digital library.*

Thus, in examining the various examples of what are called digital libraries, it appears that librarians have been confused about what a digital library is, that the word "library" has been appropriated by many different groups to describe either their areas of research or signify a simple collection of digital objects.

So what is a working definition of "digital library" that makes sense to librarians? As a starting point, we should assume that digital libraries are *libraries* with the same purposes, functions, and goals as traditional libraries—collection development and management, subject analysis, index creation, provision of access, reference work, and preservation. A narrow focus on digital formats alone hides the extensive behind-the-scenes work that libraries do to develop and organize collections and to help users find information.

The institutions involved in the American Digital Library Federation came up with a similar notion of "digital library." It also emphasizes the traditional underpinnings of libraries—selection, access, and preservation—as well as the fact that digital libraries will necessarily be constructed to serve particular communities (Waters, 1998):

> *Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.*

With the assumption that digital libraries are libraries first and foremost, we can list some characteristics. These characteristics have been gleaned from various discussions about digital libraries, both online and in print (See Arms, 1995; Graham, 1995a; Chepesuik, 1997; Lynch and Garcia-Molina, 1995):

- digital libraries are the digital face of traditional libraries that include both digital collections and traditional, fixed media collections. So they encompass both electronic and paper materials.

- digital libraries will also include digital materials that exist *outside* the physical and administrative bounds of any one digital library

- digital libraries will include all the processes and services that are the backbone and nervous system of libraries. However, such traditional processes, though forming the basis digital library work, will have to be revised and enhanced to accommodate the differences between new digital media and traditional fixed media.

- digital libraries ideally provide a coherent view of all of the information contained within a library, no matter its form or format

- digital libraries will serve particular communities or constituencies, as traditional libraries do now, though those communities may be widely dispersed throughout the network.

- digital libraries will require both the skills of librarians and well as those of computer scientists to be viable.

One thing digital libraries will *not* be is a single, completely digital system that provides instant access to *all* information, for *all* sectors of society, from *anywhere* in the world. This is simply unrealistic. This concept comes from the early days when people were unaware of the complexities of building digital libraries. Instead, they will most likely be a collection of disparate resources and disparate systems, catering to specific communities and user groups, created for specific purposes. They also will include, perhaps indefinitely, paper-based collections. Further, interoperability across digital libraries—of technical architectures, metadata, and document formats—will also only likely be possible within relatively bounded systems developed for those specific purposes and communities.

For librarians, this definition of a digital library, and these characteristics, are the most logical because it expands and extends the traditional library, preserves the valuable work that they do, while integrating new technologies, new processes, and new media.

2. WHAT ARE THE ISSUES AND CHALLENGES IN CREATING DIGITAL LIBRARIES?

The optimism and hype from the early 1990's has been replaced by a realization that building digital libraries will be a difficult, expensive, and long-term effort (Lynch and Garcia-Molina, 1995). Creating effective digital libraries poses serious challenges. The integration of digital media into traditional collections will not be straightforward, like previous new media (e.g., video and audio tapes), because of the unique nature of digital information—it is less fixed, easily copied, and remotely accessible by multiple users simultaneously. Some the more serious issues facing the development of digital libraries are outlined below.

2.1 Technical architecture

The first issue is that of the technical architecture that underlies any digital library system. Libraries will need to enhance and upgrade current technical architectures to accommodate digital materials. The architecture will include components such as:

- high-speed local networks and fast connections to the Internet
- relational databases that support a variety of digital formats
- full text search engines to index and provide access to resources
- a variety of servers, such as Web servers and FTP servers
- electronic document management functions that will aid in the overall management of digital resources

One important thing to point out about technical architectures for digital libraries is that they won't be monolithic systems like the turn-key, single box OPAC's with which librarians are most familiar. Instead, they will be a *collection* of disparate systems and resources connected through a network, and integrated within one interface, most likely a Web interface or one of its descendants. For example, the resources supported by the architecture could include:

- bibliographic databases that point to both paper and digital materials
- indexes and finding tools
- collections of pointers to Internet resources
- directories
- primary materials in various digital formats
- photographs
- numerical data sets
- and electronic journals

Though these resource may reside on different systems and in different databases, they would *appear* as though there were one single system to the users of a particular community.

Within a coordinated digital library scheme, some common standards will be needed to allow digital libraries to interoperate and share resources. The problem, however, is that across multiple digital libraries, there is a wide diversity of different data structures, search engines, interfaces, controlled vocabularies, document formats, and so on. Because of this diversity, federating *all* digital libraries nationally or internationally would an impossible effort. Thus, the first task would be to find sound reasons for federating particular digital libraries into one system. Narrowing the field in such a manner would reduce the technical and political hurdles required to establish common practices. Further, because of the often uncertain futures of both de jure

and defacto standards over time, what those standards are is unclear.

2.2 Building digital collections

One of the largest issues in creating digital libraries will be the building of digital collections. Obviously, for any digital library to be viable, it must eventually have a digital collection with the critical mass to make it truly useful. There are essentially three methods of building digital collections:

1. **digitization**, converting paper and other media in existing collections to digital form (discussed in more detail below).
2. **acquisition of original digital works** created by publishers and scholars. Example items would be electronic books, journals, and datasets.
3. **access to external materials** not held in-house by providing pointers to Web sites, other library collections, or publishers' servers.

While the third method may not exactly constitute part of a local collection, it is still a method of increasing the materials available to local users. One of main issues here is the degree to which libraries will digitize existing materials and acquire original digital works, as opposed to simply pointing to them externally. This a reprise of the old access versus ownership issue—but in the digital realm—with many of the same concerns such as:

- local control of collections
- long-term access and preservation

What about digital collection building in a coordinated scheme? There are many reasons why building digital collections is a good candidate for coordinated activity. First, acquiring digital works and doing in-house digitization are expensive, especially to undertake alone. By working together, institutions with common goals can gain greater efficiencies and reduce the overall costs involved in these activities, as was the case with retrospective conversion of bibliographic records. Second, it also reduces the redundancy and waste of acquiring or converting materials more than once. Third, coordinated digital collection building enhances resource sharing and increases the richness of collections to which users have access.

How can specific materials to be processed by a given institution be identified? Who collects and/or digitizes what materials could be based on factors such as:

- **collection strengths**. A particular library with a strong collection focus could be responsible for digitizing selected portions of it and adding new digital works to it.

- **unique collections**. If a library has the only copies of something, they are obviously the ones to digitize it

- **the priorities of user communities**. Such priorities will justify holding the materials locally, for example, because of the demands of a curriculum

- **manageable portions of collections**. When there is no other overriding criteria, then material can be divided up among institutions simply according to what is reasonable for any one institution to collect or digitize

- **technical architecture**. The state of a library's technical architecture will also be factor in selecting who digitizes what. A library must have a technical architecture up to the task of support a particular digital collection.

- **skills of staff**. Institutions whose staff don't have the necessary skills can't become a major node in a national scheme.

Yet, no matter how a collection is built—of materials digitized in-house, of original digital works, or of providing access to materials by pointing to other external resources—libraries in a collective must ensure it is preserved and made available in perpetuity. For example, if the only copies of digital works reside on a particular publisher's server, then what happens if the publisher goes bankrupt? Or if the market value of a particular work approaches zero? What if all of part of a digital collection of a library were lost, such as through some catastrophic event? Ensuring long-term preservation and access will require policies and a scheme by which redundant permanent copies are stored at designated institutions. Preservation issues will be discussed further later in the paper.

2.3 Digitization

Recall that one of the primary methods of digital collection building is digitization. What does this term mean exactly? Simply put, it is the conversion of any fixed or analogue media—such as books, journal articles, photos, paintings, microforms—into electronic form through scanning, sampling, or in fact even re-keying. An obvious obstacle to digitization is that it is very expensive. One estimate from the University of Michigan at Ann Arbor, the organization responsible for the JSTOR project, puts the cost of digitizing a single page at $2 to $6 dollars US (Chepesuik, 1997:48).

How do you go about deciding what parts of a collection to digitize? There are several approaches available, at least theoretically:

- **retrospective conversion of collections**— essentially, starting at A and ending up a Z. However ideal such complete conversion would be, it is impractical or impossible technically, legally, and economically. This approach can arguably be dispensed with as a pipe dream.

- **digitization of a particular special collection or a portion of one**. A small collection of manageable size, and which is highly valued, is a prime candidate.

- **highlight a diverse collection** by digitizing particularly good examples of some collection strength

- **high-use materials**, making those materials that are in most demand more accessible.

- **an *ad hoc* approach,** where one digitizes and stores materials as they are requested. This is, however, a haphazard method of digital collection building.

These approaches can be used alone or in combination depending upon a particular institution's goals for digitization.

Nested within these approaches are several criteria for selecting individual items. These include:

- their potential for long-term use
- their intellectual or cultural value
- whether they provide greater access than

possible with original materials (e.g., fragile, rare materials)
- and whether copyright restrictions or licensing will permit conversion.

2.4 Metadata

Metadata is another issue central to the development of digital libraries. Metadata is the data the describes the content and attributes of any particular item in a digital library. It is a concept familiar to librarians because it is one of the primary things that librarians do—they create cataloguing records that describe documents. Metadata is important in digital libraries because it is the key to resource discovery and use of any document. Anyone who has used Alta Vista, Excite, or any of the other search engines on the Internet knows that simple full-text searches don't scale in a large network. One can get thousands of hits, but most of them will be irrelevant. While there are formal library standards for metadata, namely AACR, such records are very time-consuming to create and require specially trained personnel. Human cataloguing, though superior, is just too labour extensive for the already large and rapidly expanding information environment. Thus, simpler schemes for metadata are being proposed as solutions.

While they are still in their infancy, a number of schemes have emerged, the most prominent of which is the Dublin Core, an effort to try and determine the "core" elements needed to describe materials. The first workshop took place at OCLC headquarters in Dublin, Ohio, hence the name "Dublin Core." The Dublin Core workshops defined a set of fifteen metadata elements—much simpler than those used in traditional library cataloguing. They were designed to be simple enough to be used authors, but at the same time, descriptive enough to be useful in resource discovery.

The lack of common metadata standards—ideally, defined for use in some specified context—is yet another a barrier to information access and use in a digital library, or in a coordinated digital library scheme.

2.5 Naming, identifiers, and persistence

The fifth issue is related to metadata. It is the problem of *naming* in a digital library. Names are strings that uniquely identify digital objects and are part of any document's metadata. Names are as

important in a digital library as an ISBN number is in a traditional library. They are needed to uniquely identify digital objects for purposes such as:

- citations
- information retrieval
- to make links among objects
- and for the purposes of managing copyright

Any system of naming that is developed must be permanent, lasting indefinitely. This means, among other things, that the name can't be bound up with a specific location. The unique name and its location must be separate. This is very much unlike URLs, the current method for identifying objects on the Internet. URL's confound in one string several items that should be separate. They include the method by which a document is accessed (e.g., HTTP), a machine name and document path (its location), and a document file name which may or may not be unique (e.g., how many index.html files do you have on your Web site?). URLs are very bad names because whenever a file is moved, the document is often lost entirely.

A global scheme of unique identifiers is required, one that has persistence beyond the life of the originating organization and that is not tied to specific locations or processes. These names must remain valid whenever documents are moved from one location to another, or are migrated from one storage medium to another.

Three examples of schemes proposed to get around the problem of persistent naming are PURLs, URNs, and Digital Object Identifiers.

- **PURLS**. PURLs are *persistent* URLs. They are a scheme developed by OCLC in an attempt to separate a document name from its location and therefore increase the probability that it will always be found. PURLs work through a mapping of a unique, never-changing PURL to an actual URL. If a document moves, the URL is updated, but the PURL stays the same. In operation, a user requests a document through a PURL, a PURL server looks up the corresponding URL in a database, and then the URL is used to pass the document to the user.[1] Because PURLs also confound a name with an access method, like URLs, they are not true

names (Lynch, 1998).

- **Uniform Resource Name** (URN). URNs are a development of the Internet Engineering Task Force (IETF). A URN is not a naming scheme in itself, but a *framework* for defining identifiers (Lynch, 1998). They contain a naming authority identifier (a central authority given the task of assigning identifiers) and an object identifier (assigned by the central authority). Like PURLs, URNs must be resolved, through a database or other such system, into actual URLs. Unlike PURLs, however, a URN can be resolved into more than one URL, such as one for each of several different formats. There is currently no working URN system.

- **Digital Object Identifier (DOI) System**. DOI is an initiative by the Association of American Publishers and the (American) Corporation for National Research Initiatives designed to provide a method by which digital objects can be reliably identified and accessed. The CNRI Handle system, which underlies DOI, is a system that resolves digital identifiers into the information required to locate and access a digital object. The main impetus of the DOI system is to provide publishers with a method by which the intellectual property right issues associated with their materials can be managed.[2]

The issue of persistent naming raises it head in a coordinated scheme, as well. Persistent names is an *organizational* problem, rather than an engineering problem. Technically, a system to handle names is possible, however, unique identifiers will only persist if some institution takes responsibility for their management and migration from a current technology to succeeding generations of technologies. Thus, one goal of a coordinated digital library scheme would be to identify an institution or institutions that would take charge of issuing, resolving, and migrating a system of unique names.

2.6 Copyright / rights management

Copyright has been called the "single most vexing barrier to digital library development" (Chepesuik, 1997:49). The current paper-based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects

---

[1] For more information, see www.purl.org.

[2] See www.doi.org.

are less fixed, easily copied, and remotely accessible by multiple users simultaneously. The problem for libraries is that, unlike private businesses or publishers that own their information, libraries are, for the most part, simply caretakers of information—they don't own the copyright of the material they hold. It is unlikely that libraries will ever be able to freely digitize and provide access to the copyrighted materials in their collections. Instead, they will have to develop mechanisms for managing copyright, mechanisms that allow them to provide information without violating copyright, called rights management.

Some rights management functions could include, for example:

- usage tracking
- identifying and authenticating users
- providing the copyright status of each digital object, and the restrictions on its use or the fees associated with it
- handling transactions with users by allowing only so many copies to be accessed, or by charging them for a copy, or by passing the request on to a publisher

2.7 Preservation

Another important issue is preservation—keeping digital information available in perpetuity. In the preservation of digital materials, the real issue is technical obsolescence. Technical obsolescence in the digital age is like the deterioration of paper in the paper age. Libraries in the pre-digital era had to worry about climate control and the de-acidification of books, but the preservation of digital information will mean constantly coming up with new technical solutions.

When considering digital materials, there are three types of "preservation" one can refer to:

- **the preservation of the storage medium**. Tapes, hard drives, and floppy discs have a very short life span when considered in terms of obsolescence.   The data on them can be refreshed, keeping the bits valid, but refreshing is only effective as long as the media are still current.   The media used to store digital materials become obsolete in anywhere from two to five years before they are replaced by better technology.   Over the long term, materials

stored on older media could be lost because there will no longer have the hardware or software to read them. Thus, libraries will have to keep moving digital information from storage medium to storage medium.

- **the preservation of access to content.** This form of preservation involves preserving access to the *content* of documents, regardless of their format. While files can be moved from one physical storage medium to another, what happens when the formats (e.g., Adobe Acrobat PDF) containing the information become obsolete? This is a problem perhaps bigger than that of obsolete storage technologies.   One solution is to do data migration—that is, translate data from one format to another preserving the ability of users to retrieve and display the information content. However, there are difficulties here too—data migration is costly, there are as yet no standards for data migration, and distortion or information loss is inevitably introduced every time data is migrated from format to format.

The bottom line is that no one really knows how yet how to best migrate digital information. *Preserving digital information: The Report of the Task Force on Archiving of Digital Information* (RLG, 1995) by the US Commission on Preservation and Access and RLG states, "the preservation community is only beginning to address migration of complex digital objects" and such migration remains "largely experimental." Even if there were adequate technology available today, information will have to be migrated from format to format over many generations, passing a huge and costly responsibility to those who come after.

- **the preservation of fixed-media materials through digital technology.** This slant on the issue involves the use of digital technology as a replacement for current preservation media, such as microforms. Again, there are, as yet, no common standards for the use of digital media as a preservation medium and it is unclear whether digital media are as yet up to the task of long-term preservation.   Digital preservation standards will be required to consistently store and share materials preserved digitally (Chepesuik, 1997).

What can libraries jointly do in a coordinated scheme? They can:

- create policies for long-term preservation
- ensure that redundant permanent copies are stored at designated institutions
- help establish preservation standards to consistently store and share materials preserved digitally

3.0 CONCLUSION

Libraries around the world have been working on this daunting set of challenges for several years now. They have created many digital library initiatives and projects, and have formed various national schemes for jointly exploring key issues. With several years accumulated experience, the initial enthusiasm surrounding the development of the digital library has been replaced by sober second thought. Librarians have discovered that, with a few exceptions, making a business case for digitization and investments in digital technology is more difficult than first envisioned, especially given the technical and legal constraints that must first be overcome. As with most other technical developments in libraries over the years, we will have to move forward in small, manageable, evolutionary steps, rather than in an rapid revolutionary manner.

SELECTED SOURCES

Arms, W.Y. (1995). Key concepts in the architecture of the digital library. *D-lib Magazine*, July, 1995. URL: http://www.dlib.org/dlib/July95/07arms.html

Bush, V., "As We May Think", *Atlantic Monthly*, July 1945, pp. 101-108.

Chapman, S. and Kenny, A.R. (1996). Digital conversion of research library materials: a case for full informational capture. *D-lib Magazine*, October, 1996. URL: http://www.dlib.org/dlib/october96/cornell/10chapman.html

Chepesuik, R. (1997). The future is here: America's libraries go digital. *American Libraries*, 2(1), 47-49.

Erway, R.L. (1996). Digital initiatives of the Research Libraries Group. *D-Lib Magazine*, December, 1996. URL: http://www.dlib.org/dlib/december96/rlg/12erway.html

Graham, P.S. (1995a). *Requirements for the digital research library*. URL: http://aultnis.rutgers.edu/texts/DRC.html

Graham, P.S. (1995b). *Long-term intellectual preservation*. URL: http://aultnis.rutgers.edu/texts/dps.html

Lesk, M. (1996). Going digital. *Scientific American*. March, 1996, 58-60. Also available at: URL: http://www.sciam.com/0397issue/0397lesk.html

Lynch, CA (1995). The Tulip project: context, history, and perspective. *Library Hi Tech*, 52(13), 8-24.

Lynch, C.A. (1997). Searching the Internet. *Scientific American*, March, 1997, 52-56. Also available at: URL: http://www.sciam.com/0397issue/0397lynch.html

Lynch, CA. and Garcia-Molina, H. (1995). *Interoperability, scaling, and the digital libraries research agenda: a report on the May 18-19, 1995 IITA Digital Libraries Workshop*. URL: http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html

Lynch, C.A. (1998). Identifiers and their role in networked information applications. Feliciter, January, 1998, pp. 31-35.

Masinter, L. (1995). *Document management, digital libraries, and the Web*. URL: http://www.cernet.edu.cn/HMP/PAPER/243/html/paper.htm

Miller, J.S. (1996). W3C and digital libraries. *D-Lib Magazine*, November, 1996. URL: http://www.dlib.org/dlib/november96/11miller.html

Nurnberg, P.J., Furuta, R., Leggett, J.J., Marshall, C., and Shipman III, F.M. (1995). Digital libraries: issues and architectures. In *Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries*. Austin, Texas, June 11-13, 1995, pp. 147-153.

Schatz, B. and Chen, H. (1996) Building large-scale digital libraries. *Computer*, May, 1996. Also available at: URL: http://www.computer.org/pubs/computer/dli/

Shreeves, E. (1997). Is here a future for cooperative collection development in the digital age? *Library Trends*, 4(3), 373-390.

Steele, Colin. (1995). The digital library: do's, don'ts and developments. *The Electronic Library*, 13(5), 435-437.

Stefik, M. (1997). Trusted systems. *Scientific American*, March, 1997, 78-81. Also available at: URL: http://www.sciam.com/0397issue/0397stefik.html

RLG. (1995) *Preserving digital information: The Report of the Task Force on Archiving of Digital Information*. Commissioned by the Commission on Preservation and Access and the Research Libraries Group. URL: http://www.rlg.org/ArchTF/tfadi.index.htm

Waters, D.J. (1998). What are digital libraries? *CLIR Issues,* July/August. URL: http://www.clir.org/pubs/issues/issues04.HTML

Weibel, S. (1995). Metadata: The Foundations of Resource Description. *D-Lib Magazine*, July 1995. URL: http://www.dlib.org/dlib/July95/07weibel.html