Overview of Digital Library Components and Developments

Edward A. Fox
Department of Computer Science, Virginia Tech
Blacksburg, VA 24061 USA – fox@vt.edu – http://fox.cs.vt.edu

Abstract

Digital libraries are being built upon a firm foundation of prior work as the high-end information systems of the future. A component architecture approach is becoming popular, with well established support for key components like the repository, especially through the Open Archives Initiative. We consider digital objects, metadata, harvesting, indexing, searching, browsing, rights management, linking, and powerful interfaces. Flexible interaction will be possible through a variety of architectures, using buses, agents, and other technologies. The field as a whole is undergoing rapid growth, supported by advances in storage, processing, networking, algorithms, and interaction. There are many initiatives and developments, including those supporting education, and these will certainly be of benefit in Latin America.

1. Introduction

Digital libraries extend and integrate approaches adopted in traditional libraries, as well as in distributed information systems, to yield high-end information systems, services, and institutions. Here we explore some of the parts or components of digital libraries and discuss several of the developments in this emerging field.

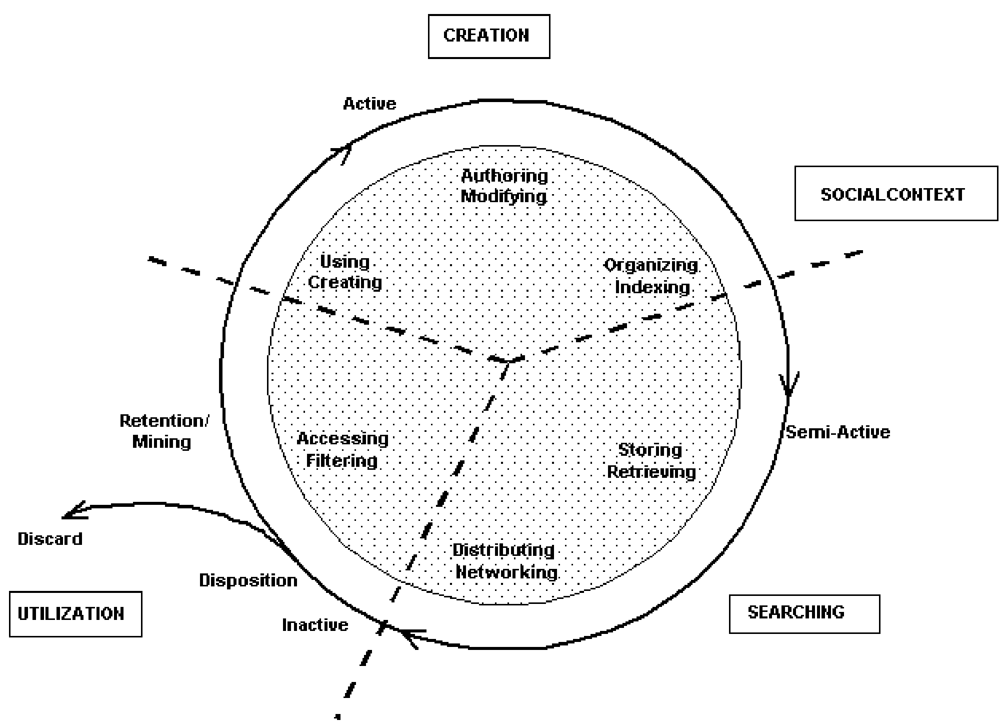
Figure 1 summarizes much of the discussion of a US National Science Foundation funded workshop on Social Aspects of Digital Libraries [1]. Comprehensive digital libraries will help users manage all phases of the information lifecycle. Of particular import is to simplify the authoring and creation processes so that wider populations can participate, adding all types of multimedia content directly into digital libraries. Downstream access allows readers to benefit from this type of computer-mediated communication, across time and space. Ultimately, it is hoped that knowledge will be shared and then lead to yet another cycle of discovery, authoring, and utilization that is facilitated by digital libraries.

Digital libraries are distinguished in that they afford services connected with each of the phases of the lifecycle. They integrate technologies from a variety of disciplines to help realize the designs articulated by early visionaries.

1.1 Foundations

Vannevar Bush was one of the first to clearly describe problems related to the modern explosion of information and to appeal to technology to help us meet our needs regarding scholarly communication [2]. Licklider painted a more complete picture, identifying the needs for better distributed processing, human-computer interaction, document management, and retrieval [3]. Salton helped launch the modern era of automatic indexing and search through some 30 years of laboratory research [4]. But real development of digital libraries per se only began in the early 1990s, drawing upon such visions as well as statements of needs and requirements from prospective users [5-7]. Work on early projects like TULIP [8], ongoing efforts to reach consensus and establish standards like the Dublin Core Metadata Initiative [9, 10], and several rounds of research funding [11], have all helped lay a firm foundation as a clearer understanding of its scope has emerged.

Information Life Cycle



NOTE: The outer ring indicates the life cycle stages (active, semi-active, and inactive) for a given type of information artifact (such as business records, artworks, documents, or scientific data). The stages are superimposed on six types of information uses or processes (shaded circle). The cycle has three major phases: information creation, searching, and utilization. The alignment of the cycle stages with the steps of information handling and process phases may vary according to the particular social or institutional context.

Figure L=1. Information Life Cycle: Diagram from Workshop Report on Social Aspects of Digital Libraries, http://www-lis.gseis.ucla.edu/DL/

1.2 Definitions

The scope and definition of the field of digital libraries has been the subject of intensive debate, which is well summarized in [12]. Here we simply remind the reader of the integrative nature of the field through definitions that show such combinations:

- Library++ = library + archive + museum + …
- Distributed information system + organization + effective interfaces
- User community + collection (content) + services

However, to help clarify our subsequent analysis, we add our own definition, drawing upon the 5S framework [13], with its 5 key constructs. Thus, digital libraries are complex systems that

1. help satisfy information needs of users (societies),
2. provide information services (scenarios),
3. locate and present information in usable ways (spaces),
4. organize information in usable ways (structures), and
5. communicate information with users and computers (streams).

2. Components

To build digital libraries we must ensure that each of the "S" constructs is addressed, and so can use 5S as a checklist or guideline. In operational terms, however, many digital libraries are built out of components that are integrated into a production quality system. Figure 2 highlights some of the most important such components.

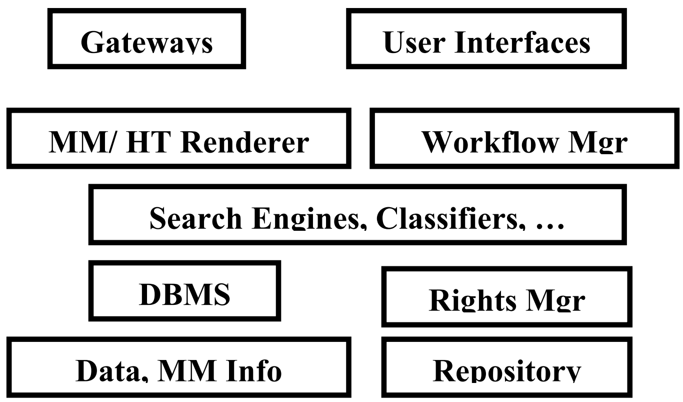


Figure C=2. Components of a Digital Library

In the following subsections we explore issues and subcomponents related to this figure.

2.1 Digital objects

The actual content of digital libraries is made up of a number of digital objects. In some cases these may be thought of as data sets (e.g., a table of results, the genomic information for an individual). In others they will be multimedia information, such as an image, graphic, animation, sound, musical performance, or video. Many can be thought of as documents, which carry content in some structure or structures, perhaps made up of logical or physical divisions such as sections or pages. Some of the objects will be "born digital", such as this paper, while others may be representations of some physical object (such as a painting that is shown through a digital image) that result from some type of digitization process. Into the foreseeable future, digital libraries will be hybrid constructs, where paper, microforms, and other media carry much of the content that is of interest and only the metadata is in digital form.

2.2 Metadata

Digital objects are described, structured, summarized, managed, and otherwise manipulated in surrogate form through the use of "metadata", which literally means data about data. Three types are often distinguished: descriptive, structural, and administrative. Metadata is usually produced through a process called “cataloging” that is often carried out by trained librarians. Collections of such information are commonly stored in “catalogs”. In computerized environments, metadata may be automatically or semi-automatically extracted or derived from the original content, or the “full-text” may simply be indexed and searched without involving metadata (as happens on the WWW when search engines are employed). Nevertheless, if metadata is available and can be used along with content terms derived from full-text documents, the result is even better [14]. Similarly, if only metadata is available in computer form to describe a digital object, it must be used in digital libraries. Hence, metadata should be used when available in a digital library, and is an important aspect in many such systems.

2.3 Repositories and Harvesting

As can be seen in Figure 3, we can think of digital libraries as containing a collection of digital objects (DOs), each of which has one or more sets of metadata objects (MDOs) associated. This “repository” part of a digital library may, as is the case in the Open Archives Initiative [15], follow certain conventions [16]. In particular, according to the latest specifications, an “Open Archive” (OA) is a computer system with a WWW server that behaves according to an OA protocol to allow other computers to harvest metadata from it. That protocol supports requests to, for example:

- list what types of metadata format are present,
- list what structure of sets and subsets are used to organize or partition the content,
- disseminate or return a particular MDO, or
- list URIs (unique identifiers) for all MDOs added during a particular date range.

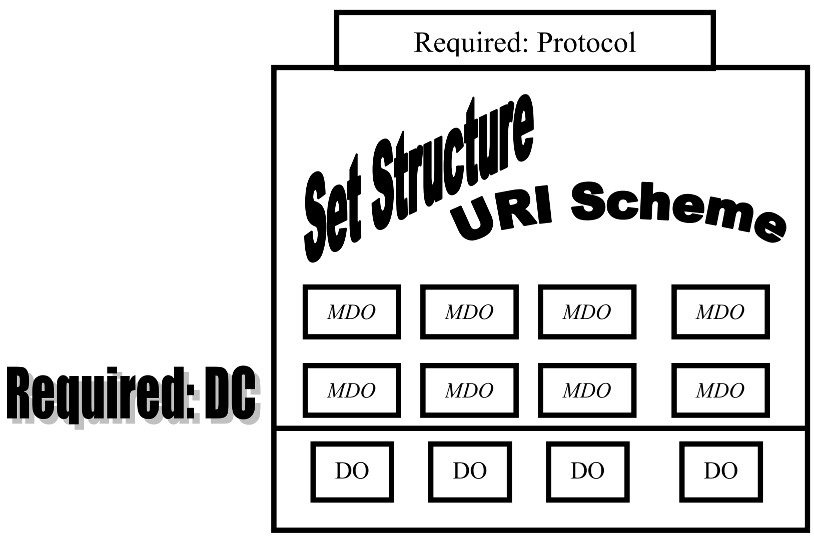


Figure O=3. Open Archives Repository

In particular, every OA must be able to return MDOs that conform to the Dublin Core and are coded in XML. They must have documented policies about what DOs are included, what “archiving” is in place, and what terms and conditions apply to use, if any. Having a repository component that is an OA means that access to digital library content is open, ensuring widespread interoperability at one important level.

At a slightly deeper level, interoperability among digital libraries requires that digital objects be accessible through some scheme for universal identifier names (e.g., URNs). MDOs thus have a URN that facilitates access (if authorized) to the corresponding DO. A likely policy for an OA is to use some specific type of URN (e.g., OCLC's PURL or CNRI's handle – also employed for Digital Object Identifiers, DOIs). In today's ubiquitous WWW environment, it is presumed that users and computers will be able to retrieve a desired DO if its identifier is known.

2.4 Rights Management

Considering again Figure 2, we note that in the layer about data, multimedia information, and repositories is the "rights manager" which must protect intellectual property rights. In the trivial case, which fortunately is common, content is freely available so nothing is needed here. In some cases too, where content is encrypted, content management is outside the scope of the digital library, since secure objects are stored and retrieved and the steps of encryption and decryption occur remotely. Similarly, some content may have "watermarks" added in a way that makes removal difficult, so that subsequent access can be monitored or controlled.

However, as e-commerce, e-government, and other movements spread, it will be crucial for many digital libraries to manage rights. This typically involves a number of steps.

1. The digital library should include policies and rules specifying the management required.
2. The users of the digital library should be authenticated in some way so they are known.
3. The content of the digital library should be shown to be authentic.
4. Payment should be made if access requires that in a particular case.
5. Users who are authorized to access a DO are allowed to do so.
6. Subsequent access with the DO may take place after retrieval to a user's site.

IBM has developed sophisticated technology and digital library systems with support for rights management [17]. Xerox staff have developed a Digital Property Rights Language that relates to step 1 above. Password schemes and systems like Kerberos relate to step 2. Hashing with MD5 or use of digital signatures can support step 3. E-commerce mechanisms, especially those enabling micro-payments or subscriptions, relate to step 4. Rules processing relates to step 5, as was done at Case Western Reserve University – employing Prolog to work with users, user classes, documents, and document classes – in order to find the lowest cost solution. In step 6, there usually is little control, unless a scheme like IBM's Cryptolope mechanism is involved wherein the DO is encapsulated with code that limits access (e.g., prohibits printing).

2.5 Indexing, Resource Discovery, Searching, and Retrieving

Considering Figure 2 further, we clearly must support finding DOs, directly or through MDOs, so that they can be identified, retrieved, and used. Often, DOs and/or MDOs are automatically indexed so that some index structure is built to speed up search. Such indexing may build upon any manual indexing carried out by authors, other creators, or indexers. Automatic indexing also may involve first classifying DOs, such as when OCLC's CORC project suggest Dewey Decimal Classification entries for a WWW page that is being cataloged.

Indexes may be centralized or distributed. They may be two-level, allowing a resource discovery phase to proceed to find what source(s) should be included in the second (lower) level search. Indexes also may have multiple parts, such as when a document has a text, image, audio, or video part. Content-based indexing of multimedia information generally involves identifying and assessing features that characterize the DOs, whether they involve concepts, n-grams, words, keywords, descriptors, phonemes, textures, color histograms, eigenvalues, links, or user ratings.

Most commonly, searching in a digital library involves an information retrieval (IR) system or search engine. In some cases a database management system is used instead or underlies the IR system. In any case, retrieval will be more effective if a suitable scheme is used to combine the various types of evidence available [18], to indicate if a DO might be relevant with respect to the query that is used to express the user's information need.

2.6 Linking, Annotating, and Browsing

Once a DO is found, it often is appropriate to follow links from it to cited works. Further, notes can be recorded as annotations and linked back to the works, so they can be recalled later or shared with colleagues as part of collaborative activities. If suitable clustering is in place, other DOs that are "near" a given work may be examined. Or, using a classification system appropriate for the content domain, users may browse around in "concept space" and link at any point between concepts and related DOs. Browsing also can proceed based on any of the elements in the MDO. Thus, dates, locations, publishers, contributing artists, language, and other aspects may be considered to explore the collection or refine a search.

2.7 Interfaces and Interaction

Ultimately, users will connect through a human-computer interface and interact with the digital library, though in some cases the digital library may be an embedded system that is seen only indirectly (e.g., through a word processor that allows one to search for a quotation). Most commonly, a digital library has an interface for users to search, browse, follow links, retrieve, and read documents. As can be seen in Figure 4, that interface may be specialized according to what roles the user will play.

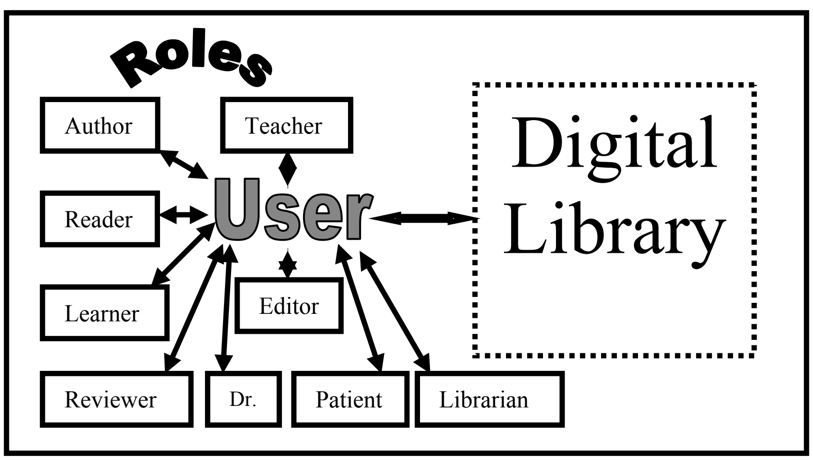


Figure U=4. Users Direct

For example, the Computer Science Teaching Center, a digital library of courseware about computing, requests that all users, except those just browsing or searching, login to identify themselves. Then it knows and tailors the interface so that suitable authorized users can submit, review, or edit works. Further, CSTC encourages users to submit courseware they have developed that others might use, and supports their entering in suitable metadata as well as uploading their applets, demonstrations, laboratory exercises, interactive multimedia training resources, etc. After the work has been entered by the creator, it can be improved through peer review and "certified" for public use, or even accepted for publication in the ACM Journal of Educational Resources in Computing (JERIC) [19]. Thus, instead of requiring current complex and expensive chains of processing for journal submissions, handling with digital libraries may be implemented through this "users direct" model.

Following retrieval, users examine a list of results or else work with more sophisticated schemes for visualizing and managing results [7, 20-23]. Many other types of rich interaction are possible through innovative digital library interfaces [24].

In particular, as can be seen in Figure 2, digital libraries may use special software to present or render multimedia, hypertext, and hypermedia content. Real-time requirements must be satisfied to ensure adequate quality of service when streaming content (recall 5S) is involved, especially if multiple streams are involved. Such software may be launched from a WWW browser, which typically only supports a limited range of hypermedia.

Also related to interaction with the digital library is the matter of workflow management. Especially with the "users direct" model, the various people interacting with a particular DO may leave some record of approval or review, which triggers others to continue with processing. For example, universities that are part of the Networked Digital Library of Theses and Dissertations [25-27] allow students to upload their works, make changes, secure approval by the graduate school, have cataloging information added by librarians, and ultimately release the suitably enhanced information for widespread access.

Finally, from Figure 2 we see that many digital libraries will have their own user interfaces. In addition, they may support gateway connections, using protocols like Dienst [28] or Z39.50 [29]. As we work toward world digital libraries, we must ensure that our interfaces support multimedia and multilingual access, as well as various schemes for interconnection with other digital libraries and resources.

2.8 Architectures and Interconnecting

Since the field of digital libraries is young, there still is active investigation regarding architecture, interconnection, and interoperability [30]. Figure 2 shows one, rather high-level, decomposition of a digital library into components. Given the range of legacy systems that are used today as parts of digital libraries, the actual situation often is more complex.

To simplify matters, several interconnection strategies have been explored. At Stanford, a bus approach has been used [31-33]. Mediation code "wrap" around various collections or resources to make suitable conversions to representations supported by the bus and the other services connected to it.

Agents provide another interconnection mechanism [34-39]. Many agent-based systems use KQML as the language for transporting knowledge constructs [40]. Yet another approach is the distributed scheme supporting federated search that underlies the Dienst system [28, 41, 42].

Nevertheless, currently it is not possible to identify the best architecture(s) for digital libraries. We must look to future technological developments and actual deployments to resolve such questions.

3. Developments

Since the early 1990s, the digital library field has emerged as an important area of research and development. There are hundreds of projects and thousands of reports/publications describing them. One of the active research programs is in the USA, funded by the National Science Foundation. A summary of that Digital Libraries Initiative makes clear the broad scope of work underway [43]. NSF has deliberately selected a diverse set of content areas, genre, media, and user communities in an effort to rapidly develop the field [11]. In the following sections we explore the overall process of such development.

3.1 Technology

Work on digital libraries has been facilitated by technical advances in a number of areas. For the first time, storage systems are readily affordable that can handle enormous text collections, very large image collections, and large audio or video collections. Fast processors, supercomputers, cluster-computers, networks of workstations, and other computational aids have provided ample processing capacity to handle user communities operating on a global scale. Increases in network speeds and bandwidth have made it possible to build distributed systems that perform well and have high reliability. Computationally expensive algorithms have been refined so that useful techniques such as LSI can help with multilingual retrieval and other applications [44]. High-end graphics systems and virtual environments also have evolved to be usable for information visualization as well as interfacing with digital libraries [45]. Representation schemes like PDF and XML have made digital documents easy to produce and share, facilitating information interchange and encouraging further digitization.

An example of the effects of technology can be seen with regard to the MARIAN digital library system developed at Virginia Tech [46-49]. Over the last decade of work, coding has switched from C to C++ to Java. Hardware has switched from mainframes to minis to PCs. Current work to enhance performance includes development of new algorithms to manipulate inverted files [50] on the Virginia Tech PetaPlex system, which has 100 processors and 2.5 terabytes of disk storage capacity. Earlier studies of performance [49] demonstrated that the architecture is scalable, and showed that changing some of the internal communication from TCP to UDP would lead to substantive improvement.

3.2 Economic, Social, and Legal Issues

According to the 5S framework, the topmost level deals with “societies”. Though technology has made possible many advances in digital libraries, all such efforts are situated in a social context, as can be seen in Figure 1. Many of the key social issues were identified in an NSF-funded workshop on this topic [1]. An excellent explanation of social issues in constructing and evaluating digital libraries appears in [51].

Underlying work on the Networked Digital Library of Theses and Dissertations is a strong social and educational rationale, to prepare the next generation of scholars for the Information Age [25, 52-55]. Its aims – of encouraging discussion about intellectual property rights among students and faculty, of building awareness and infrastructure about digital libraries on campuses, and of developing a new genre for communication among graduate students and researchers – are largely being met. The impact of economics is remarkable, in that making works available for free leads to hundreds or thousands of downloads per work per year. This contrasts with interlibrary loan or buying copies for roughly $50, which very rarely led to more than 5 accesses per year.

Social and technical issues often relate. For example, the culture and social atmosphere determines how electronic theses and dissertations will be managed on a particular campus, or across broader boundaries like state or nation. Universities with advanced infrastructure, like Virginia Tech, MIT, and Caltech, have their own services. On the other hand, there are regional/national projects associated with NDLTD in Ohio, Catalunya, Australia, Germany, India, Portugal, and South Africa. Because of this arrangement, a federated search mechanism was implemented [56] though future plans call for use of Open Archives and regular harvesting. Groups that own content or have a tradition of managing it can continue to do so, while at the same time technical approaches can allow comprehensive search across such distributed collections.

Legal issues also become more visible with digital libraries. There often was little concern over copyright issues when preparing a dissertation that would largely go unread, sitting on a shelf in a local library. However, with the potential of thousands of downloads from around the world, authors must be very careful not to include a copyrighted image or other content without permission.

Other economic, social, and legal issues have come to the fore with digital libraries. In many states, like Virginia, a single university library can run a service for the whole state, usually at a rate that benefits from volume purchasing. On the other hand, many such services involve a contract with an information provider, that often has a complex set of terms and conditions, meaning that libraries now require more legal counsel.

Digital libraries allow new groups to assemble around new collections of interest. In many countries, digital libraries show promise regarding preserving cultural, historic, and linguistic records. In a number of situations, they have the potential of aiding economic development by supporting (distance) education. We explore such cases among those discussed in the next subsection.

3.3 Initiatives and Projects

There are hundreds of digital library efforts underway around the world. A good place to find out about many of these is in *D-Lib Magazine* [57], which by the end of 2000 has about 1000 entries.

As mentioned earlier, the National Science Foundation in USA has funded the Digital Libraries Initiative with awards granted in 1994, 1998, and 1999. Well over $50M has gone to a wide range of research projects. NSF support to Stanford University through this program has led in part to the appearance of Google, one of the most popular search services on WWW. Other technical advances have led to methods for processing and searching collections of images, music, and video. A great expansion in work on geographic information systems and other spatial data has also resulted. Microsoft's terra server demonstrates how very large data collections can serve huge user communities in spite of using relatively modest hardware resources.

The museum community has demonstrated methods of accessing distributed cultural heritage information [58]. Sacred and/or precious resources, including antiquities, have become available from sources such as the Vatican [59, 60]. Special collections, such as of butterflies [61] or floristic information [38, 39, 62, 63], have become accessible through digital libraries.

In some ways the broadest and most influential digital library efforts have to do with education. For example, in USA, after years of study, it was decided that NSF should support development of a digital library to support learning by (undergraduate) students [64]. This has led to over $50M committed to the National Science (and Mathematics, Engineering, and Technology Education) Digital Library, which should be launched by the end of 2002 [65]. Many thousands of teachers, and millions of students, will benefit from this resource as it aims to enhance learning through dynamic interaction, visualization, simulation, and other computer-related devices.

4. Conclusion

In conclusion, we try to put digital libraries in perspective by answering key questions.

Why? We build digital libraries for many reasons. They can help us preserve our linguistic, literary, historical, and cultural heritage. They make access simpler and cheaper. They lower the costs of disseminating information. They help us establish new communities around new collections that can now become available. They support teaching and learning, especially in the context of distance or lifelong learning. They allow rich media types to be included and managed effectively. They encourage authors to create and share, and others to collaborate and quickly build on newly discovered knowledge.

How? Digital libraries provide comparable or better services to traditional libraries at relatively low cost over existing networks and tools like those used for WWW. They are built in most cases upon prior technologies, such as library automation packages or records management systems. Research systems like MARIAN and Dienst may become more widely available and used. Certainly, light-weight software like that for Open Archives will be commonly used. Connecting a URN scheme, a repository mechanism, search engine(s), and WWW interface can lead to a relatively simple system in short order.

What? All types of content will be available through digital libraries. They will become the high-end of the information systems world, including not only bibliographic and full-text content but also images, music, and video. Included will be medical images, images of scanned pages, engineering drawings, satellite images, educational videos, courseware, oral histories, 3D renderings of museum objects, musical performances, etc.

Where? Digital libraries are appearing all around the world. They are at universities, publishers, government agencies, and public libraries. Some companies are shifting to use them for records and reports as well as for customer access. In the context of this booklet, we are assured that they will be widely used throughout Latin America!

5. Resource List

There is a great deal of information available regarding digital libraries. An easy way to get started is to consult our online courseware [66]. In particular, follow links to Resources and to References (and thence to Repositories for a set of other Web sites to consider).

Early events in the field are summarized in a 1993 *Sourcebook* [67]. A handy reference work for those building digital libraries is an extensive white paper prepared for Sun Microsystems [68]. The first two single-author books providing an overview appeared in 1997 [69] and 2000 [70]. *D-Lib Magazine* [57] is an online publication to which many in the field send news, early results, or helpful summaries. There is one journal [71], but numerous special issues on the topic have appeared in other journals [43, 72-76]. Numerous conferences sponsored by ACM [77, 78] or by other organizations. In 2001 the first ACM / IEEE-CS joint conference will take place [79]. There also are annual regional conferences in Europe and Asia, in additional to national events in countries like Japan and Russia. It is hoped that the reader will explore these various resources further, and participate in some of the many workshops and conferences.

6. References

[1] C. Borgman, “Social Aspects of Digital Libraries,” UCLA, Los Angeles, NSF Workshop Report, Feb. 16-17, 1996, 1996. http://www-lis.gseis.ucla.edu/DL/

[2] V. Bush, “As We May Think,” *Atlantic Monthly*, 1945.

[3] J. C. R. Licklider, *Libraries of the Future*. Cambridge, MA: MIT Press, 1965.

[4] G. Salton, *Automatic Information Organization and Retrieval*. New York: McGraw-Hill, 1968.

[5] E. A. Fox, D. Hix, L. Nowell, D. Brueni, W. Wake, L. Heath, and D. Rao, “Users, User Interfaces, and Objects: Envision, a Digital Library,” *J. American Society Information Science*, vol. 44, pp. 480-491, 1993.

[6] E. A. Fox, “Envision-ing a Computer Science Digital Library,” presented at Digital Libraries of the Future Panel, ACM Multimedia 93, Anaheim, CA, 1993.

[7] L. Heath, D. Hix, L. Nowell, W. Wake, G. Averboch, and E. A. Fox, “Envision: A User-Centered Database from the Computer Science Literature,” *Communications of the ACM*, vol. 38, pp. 52-53, 1995.

[8] W. C. Dougherty and E. A. Fox, “TULIP at Virginia Tech,” *Library Hi Tech*, vol. 13, pp. 54-60, 1995.

[9] Dublin-Core-Community, “Dublin Core Metadata Initiative”. WWW site. Dublin, Ohio: OCLC, 1999. http://purl.org/dc

[10] S. Weibel, “The State of the Dublin Core Metadata Initiative: April 1999,” *D-Lib Magazine*, vol. 5, 1999. http://www.dlib.org/dlib/april99/04weibel.html

[11] M. Lesk, “Perspectives on DLI-2 - Growing the Field,” *D-Lib Magazine*, vol. 5, 1999. http://www.dlib.org/dlib/july99/07lesk.html

[12] C. L. Borgman, “What are digital libraries? Competing visions,” *Information Processing and Management*, vol. 35, pp. 227-243, 1999.

[13] E. A. Fox, “The 5S Framework for Digital Libraries and Two Case Studies: NDLTD and CSTC,” in *Proceedings NIT'99*. Taipei, Taiwan, 1999. http://www.ndltd.org/pubs/nit99fox.doc

[14] E. A. Fox, “Some Considerations for Implementing the SMART Information Retrieval System under UNIX,” Cornell Univ. Dept. of Comp. Science, Ithaca, NY, Technical report, TR 83-560, Sept., 1983.

[15] H. Van_de_Sompel, “Open Archives Initiative”. WWW site. U. Ghent: OAI Group, 2000. http://www.openarchives.org

[16] H. Van_de_Sompel and C. Lagoze, “The Santa Fe Convention of the Open Archives Initiative,” *D-Lib Magazine*, vol. 6, 2000. http://www.dlib.org/dlib/february02vandesompel-oai/02vandesompel-oai.html

[17] H. M. Gladney, “Safeguarding Digital Library Contents and Users: Interim Retrospect and Prospects,” *D-Lib Magazine*, vol. 4, 1998. http://www.dlib.org/dlib/july98/gladney/07gladney.html

[18] N. J. Belkin, P. Kantor, E. A. Fox, and J. A. Shaw, “Combining the Evidence of Multiple Query Representations for Information Retrieval,” *Information Processing and Management*, vol. 31, pp. 431-448, 1995.

[19] L. Cassel and E. A. Fox, “ACM Journal of Education Resources in Computing”. WWW site. , 2000. http://purl.org/net/JERIC/

[20] L. Nowell and D. Hix, "User interface design for the project Envision database of computer science literature," in *Twenty-second Annual Virginia Computer Users Conference*. Blacksburg, VA, 1992, pp. 29-33.
[21] L. Nowell and D. Hix, "Visualizing search results: User interface development for the project Envision database of computer science literature," in *Advances in Human Factors/Ergonomics, Proceedings of HCI International '93, 5th International Conference on Human Computer Interaction*, vol. 19B, *Human-Computer Interaction: Software and Hardware Interfaces*: Elsevier, 1993, pp. 56-61.
[22] L. T. Nowell and E. A. Fox, "Envision: Information Visualization in a Digital Library". Demonstration. Seattle, WA: ACM SIGIR'95, July 10, 1995.
[23] L. T. Nowell, R. K. France, and E. A. Fox, "Visualizing search results with Envision". Demonstration. Zurich, Switzerland: ACM SIGIR'96, Aug. 19, 1996.
[24] R. Rao, J. O. Pedersen, M. A. Hearst, J. D. Mackinlay, S. K. Card, L. Masinter, P.-K. Halvorsen, and G. G. Robertson, "Rich Interaction in the Digital Library," *Communications of the ACM*, vol. 38, pp. 29-39, 1995.
[25] E. Fox, "Networked Digital Library of Theses and Dissertations: An International Collaboration Promoting Scholarship," *ICSTI Forum, Quarterly Newsletter of the International Council for Scientific and Technical Information*, vol. 26, pp. 8-9, 1997. http://www.icsti.org/icsti/forum/fo9711.html#ndltd
[26] E. A. Fox, "Networked Digital Library of Theses and Dissertations," in *Proceedings DLW15*. Japan: ULIS, 1999. http://www.ndltd.org/pubs/dlw15.doc
[27] E. Fox, "NDLTD: Networked Digital Library of Theses and Dissertations", 2000. http://www.ndltd.org
[28] C. Lagoze and J. R. Davis, "Dienst: An Architecture for Distributed Document Libraries," *Communications of the ACM*, vol. 38, pp. 47, 1995.
[29] International_Standard_Maintenance_Agency_Z39.50, "International Standard Maintenance Agency Z39.50", vol. 2000: The Library of Congress Network Development and MARC Standards Office, 2000. http://lcweb.loc.gov/z3950/agency/
[30] A. Paepcke, C.-C. K. Chang, H. Garcia-Molina, and T. Winograd, "Interoperability for Digital Libraries Worldwide," *Communications of the ACM*, vol. 41, pp. 33-43, 1998.
[31] M. Baldonado, C.-C. K. Chang, L. Gravano, and A. Paepcke, "The Stanford Digital Library Metadata Architecture," *International Journal on Digital Libraries*, vol. 1, pp. 108-121, 1997.
[32] S. Melnik, H. Garcia-Molina, and A. Paepcke, "A Mediation Infrastructure for Digital Library Services," presented at Proceedings of the Fifth ACM Conference on Digital Libraries, San Antonio, exas, 2000.
[33] A. Paepcke, "Using the InfoBus". Web site. Palo Alto: Stanford University Digital Libraries Project, 1999. http://www-diglib.stanford.edu/diglib/pub/userinfo.html
[34] W. Birmingham, "University of Michigan Digital Library Project". homepage. , 1995. http://http2.sils.umich.edu/UMDL/
[35] C. Nicholas, G. Crowder, and I. Soboroff, "CARROT: An Agent-Based Architecture for Large-Scale Document Information Systems," UMBC, Baltimore TR CS-2000-01, 2000.
[36] J. A. Sánchez and J. L. Leggett, "Agent services for users of digital libraries," *Journal of Network and Computer Applications*, vol. 20, pp. 45-58, 1997.
[37] J. A. Sánchez, J. J. Leggett, and J. L. Schnase, "AGS: Introducing agents as services provided by digital libraries," in *Proceedings of the Second ACM International Conference on Digital Libraries (DL'97)*. Philadelphia: ACM, 1997, pp. 75-82.
[38] J. A. Sánchez, C. A. Lopez, and J. L. Schnase, "An agent-based approach to the construction of floristic digital libraries," in *Proceedings of the Third ACM International Conference in Digital Libraries (DL´98)*. Pittsburgh: ACM, 1998, pp. 210-216.

[39] J. A. Sánchez, C. A. Flores, and J. L. Schnase, “Mutant: Agents as guides for multiple taxonomies in the Floristic Digital Library,” in *Proceedings of the Fourth ACM Conference on Digital Libraries (DL'99).* Berkeley: ACM, 1999, pp. 244-245.

[40] A. Barceinas, J. A. Sánchez, and J. L. Schnase, “MICK: A KQML inter-agent communication framework in a digital library,” in *Memorias del Simposium Internacional de Computación (CIC'98)*. Mexico City, 1998, pp. 66-79.

[41] C. Lagoze and S. Payette, “An Infrastructure for Open-Architecture Digital Libraries,” Cornell University, Computer Science TR98-1690, 1998.

[42] S. Payette, C. Blanchi, C. Lagoze, and E. A. Overly, “Interoperability for Digital Objects and Repositories: The Cornell/CNRI Experiments,” *D-Lib Magazine*, vol. 5, 1999. http://www.dlib.org/dlib/may99/payette/05payette.html

[43] E. Fox, “The Digital Libraries Initiative: Update and Discussion: Guest editor's introduction to Special Section,” *Bulletin of the American Society of Information Science*, vol. 26, pp. 7-11, 1999.

[44] S. T. Dumais, “References to LSI Papers”, in *LSI WWW Site*, 1998. http://superbook.bellcore.com/~std/lsiPapers.html

[45] F. A. Das_Neves and E. A. Fox, “A study of user behavior in an immersive Virtual Environment for digital libraries,” in *Proceedings of the Fifth ACM Conference on Digital Libraries: DL '00, June 2-7, 2000, San Antonio, TX*. New York: ACM Press, 2000, pp. 103-111.

[46] F. Can, E. Fox, C. Snavely, and R. France, “Incremental Clustering for Very Large Document Databases: Initial MARIAN Experience,” *Information Systems*, vol. 84, pp. 101-114, 1995.

[47] E. Fox, R. France, E. Sahle, A. Daoud, and B. Cline, “Development of a Modern OPAC: From REVTOLC to MARIAN,” in *Proc. 16th Annual Int'l ACM SIGIR Conf. on R&D in Information Retrieval, SIGIR '93*. Pittsburgh: ACM Press, 1993, pp. 248-259.

[48] R. K. France, “MARIAN Digital Library Information System (home page)”, 2000. http://www.dlib.vt.edu/products/marian.html

[49] J. Zhao, “Making Digital Libraries Flexible, Scalable, and Reliable: Reengineering the MARIAN System in JAVA,” Virginia Tech Department of Computer Science, Blacksburg, VA, Master of Science Thesis, 1999. http://scholar.lib.vt.edu/theses/available/etd-070499-204531/unrestricted/SGML-etd/

[50] O. Sornil, “A Distributed Inverted Index for a Large-Scale, Dynamic Digital Library,” Virginia Tech Computer Science, Blacksburg, Ph. D. Dissertation Draft, 2000.

[51] J. Kilker and G. Gay, “The Social Construction of a Digital Library: A Case Study Examining Implications for Evaluation,” *Information Technology and Libraries*, vol. 17, pp. 60-70, 1998. http://www.lita.org/ital/ital1702.htm

[52] E. A. Fox, J. Eaton, G. McMillan, N. Kipp, L. Weiss, E. Arce, and S. Guyer, “National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources,” *D-Lib Magazine*, vol. 2, 1996. http://www.dlib.org/dlib/september96/theses/09fox.html

[53] E. A. Fox, R. Hall, N. A. Kipp, J. L. Eaton, G. McMillan, and P. Mather, “NDLTD: Encouraging International Collaboration in the Academy,” *Special Issue on Digital Libraries of DESIDOC Bulletin of Information Technology (DBIT)*, vol. 17, pp. 45-56, 1997. http://www.ndltd.org/pubs/dbit.pdf

[54] E. A. Fox, R. Hall, and N. Kipp, “NDLTD: Preparing the Next Generation of Scholars for the Information Age,” *The New Review of Information Networking (NRIN)*, vol. 3, pp. 59-76, 1997. http://www.ndltd.org/pubs/nrin.pdf

[55] E. A. Fox, “Digital Libraries: Preparing the Next Generation of Scholars,” presented at NFAIS'98, Four Seasons Hotel, Philadelphia, PA, 1998. http://www.ndltd.org/talks/NFAIS98.ppt

[56] J. Powell and E. Fox, “Multilingual Federated Searching Across Heterogeneous Collections,” *D-Lib Magazine*, vol. 4, 1998. http://www.dlib.org/dlib/september98/powell/09powell.html
[57] W. Arms, “D-Lib Magazine”, vol. 2000, 2000. http://www.dlib.org
[58] W. E. Moen, “Accessing Distributed Cultural Heritage Information,” *Communications of the ACM*, vol. 41, pp. 45-48, 1998.
[59] H. M. Gladney, F. Mintzer, F. Schiattarella, J. Bescos, and M. Treu, “Digital Access to Antiquities,” *Communications of the ACM*, vol. 41, pp. 49-57, 1998.
[60] F. Mintzer, L. Boyle, A. Cazes, B. Christian, S. Cox, F. Giordano, H. Gladney, J. Lee, M. Kelmanson, A. Lirani, K. Magerlein, A. Pavani, and F. Schiattarella, “Towards online worldwide access to Vatican Library materials,” *IBM Journal of Research and Development*, vol. 40, pp. 139-162, 1996.
[61] J.-S. Hong, H.-Y. Chen, and J. Hsiang, “A Digital Museum of Taiwanese Butterflies,” in *Proceedings of the Fifth ACM Conference on Digital Libraries: DL '00, June 2-7, 2000, San Antonio, TX*. New York: ACM Press, 2000, pp. 260-261. http://digimuse.nmns.edu.tw
[62] M. Amavizca, J. A. Sánchez, and R. Abascal, “3DTree: Visualization of large and complex information spaces in the Floristic Digital Library,” in *Proceedings of Segundo Encuentro de Computación (ENC'99)*. Pachuca, Hidalgo, México, 1999.
[63] J. A. Sánchez, L. Fernández, and J. L. Schnase, “Agora: Enhancing awareness and collaboration in floristic digital libraries,” in *Proceedings of the Fourth CYTED-RITOS Workshop on Groupware (CRIWG'98)*. Rio de Janeiro, 1998.
[64] W. Arms, “Report of the NSF Science, Mathematics, Engineering, and Technology Education Library Workshop, July 21-23, 1998,” National Science Foundation, Division of Undergraduate Education. NSF 99-112, 1999. http://www.dlib.org/smete/public/report.html
[65] www.smete.org, “www.smete.org (home page): Information Portal: A Digital Library for Science, Mathematics, Engineering, and Technology Education”, NEEDS, Ed., 2000. http://www.smete.org
[66] E. A. Fox, “Digital Library Courseware”. WWW site. Blacksburg, VA: Virginia Tech Department of Computer Science, 1998. http://ei.cs.vt.edu/~dlib/
[67] E. A. Fox, “Sourcebook on Digital Libraries: Report for the National Science Foundation”. Technical Report TR-93-35. Blacksburg, VA: Dept. of Computer Science, Virginia Tech, December, 1993. Available by FTP from directory pub/DigitalLibrary on fox.cs.vt.edu
[68] P. Noerr, “The Digital Library Toolkit”, 2nd ed.Palo Alto, CA: Sun Microsystems, Inc., 2000, pp. 186.
[69] M. Lesk, *Practical Digital Libraries: Books, Bytes and Bucks*. San Francisco: Morgan Kaufmann Publishers, 1997.
[70] W. Y. Arms, *Digital Libraries*. Cambridge, MA: MIT Press, 2000.
[71] Springer-Verlag, “International Journal on Digital Libraries”: Springer-Verlag, 2000. http://link.springer-ny.com/link/service/journals/00799/
[72] S. Chen and E. A. Fox, “Guest Editors' Introduction to Special Issue on Digital Libraries,” *Journal of Visual Communication and Image Representation*, vol. 7, 1996.
[73] E. A. Fox and L. Lunin, “Introduction and Overview to Perspectives on Digital Libraries; guest editor's introduction to special issue,” *J. American Society Information Science*, vol. 44, pp. 441-443, 1993.
[74] E. A. Fox, R. Akscyn, R. Furuta, and J. Leggett, “Guest Editors' Introduction to Digital Libraries,” *Communications of the ACM*, vol. 38, pp. 22-28, 1995.

[75] E. A. Fox and G. Marchionini, “Toward a Worldwide Digital Library; Guest Editors' Introduction to Special Section on Digital Libraries: Global Scope, Unlimited Access,” *Comm. ACM*, vol. 41, pp. 28-32, 1998. http://purl.lib.vt.edu/dlib/pubs/CACM199804
[76] G. Marchionini and E. A. Fox, “Progress toward digital libraries: Augmentation through integration; Guest Editor's Introduction to Special Issue on Digital Libraries,” *Information Processing and Management*, vol. 35, pp. 219-225, 1999.
[77] E. A. Fox and G. Marchionini, “Proceedings of the First ACM International Conference on Digital Libraries, DL'96.”Bethesda, MD; New York: ACM, March 20-23, 1996.
[78] E. A. Fox and N. Rowe, *Proceedings of The Fourth ACM Conference on Digital Libraries, DL '99*. Berkeley: ACM, 1999.
[79] C. Borgman and E. A. Fox, *Proceedings of the First Joint Conference on Digital Libraries*. New York: ACM Press, 2001. http://www.jcdl.org