Distributed Information Management

William M. Pottenger[1]
Lehigh University

Miranda R. Callahan and Michael A. Padgett
National Center for Supercomputing Applications

INTRODUCTION

A review of this sort commonly begins with the observation that current developments in computer technology are radically changing the nature of library science and information management. Widespread digitization of information and the ubiquity of networking have created fundamentally new possibilities for collecting, distributing, and preserving information; just as important, however, as the changing technological and organizational systems themselves, are the repercussions that these powerful world-scale information networks will have on the social and cultural structures that they have been developed to serve. Similarly, the formation and development of these new technologies will, to no small extent, depend on the cultural forces that brought them into existence in the first place, as the shape of information technology and the institutions it serves are in many ways interdependent. To capture the complexity of the interwoven technological and societal forces that guide the growth of information management, then, we will we will need to cast a wide net over the fields of

[1] William M. Pottenger acknowledges with deep gratitude the help provided by his Lord and Savior Jesus Christ in accomplishing this work.

information, computer, and library science to gather topics and themes in all those areas that are shaping and being shaped by the development of distributed information systems.

A picture of such a dynamic field, encompassing so many different areas of social and technological significance, must of necessity be broadly painted. This chapter delineates the scope and effects of distributed information management, touching on current developments, experiments, and cultural implications of this rapidly changing area of research. As any user of the Internet might guess, the large number of distributed information management projects makes a truly comprehensive review of the field impossible; here we will attempt to cover at least the most important and influential work being done in the area.

Because the technologies of networking and computing are now in a state of intense, expansive growth, it is also difficult to single out any trend or thrust in the development of distributed information systems as being especially noteworthy or important—too much research is going on in too many different areas, and the capabilities of computers, distribution systems, and search engines are continually increasing. To restrict a survey of this field to any given subset of developments, then, might risk becoming excessively narrow, or even arbitrary. Instead, basing this chapter on a generous interpretation of distributed information and digital libraries, we shall review some of the ways in which technology, social systems, and inherited knowledge structures intermesh to form, and be formed by, technologies of distributed information management. Another gray area in this review is the problem of the definition of boundaries in a distributed information system. In their discussion of the boundary problem, ACKERMAN & FIELDING distinguish the "broadly-construed" from the "narrowly-construed" library system (pp. 3-5). Because of the interconnection of most networked systems, and the variety of projects developed to manage distributed information, it is often difficult to determine the difference between being "inside" and "outside" a system like a distributed digital library. This survey, therefore, deals with themes relating both to closely-knit systems under organized control, based on traditional ideas of the library, and to more open systems, in which control is distributed—the most "bottom-up" of these being, of course, the Internet (WALLACE, HARTER).

Most difficult, perhaps, is the problem of defining exactly what constitutes a distributed information management system (often termed a digital library). Given the fast-growing, constantly changing nature of this field, the only sort of definition possible at this time is an open and informal one, such as that given by ARMS, who defines a digital library as "a managed collection of information, with associated services," (ARMS, p. 2) such that the material is digitized and accessible over a network; or even more succinctly by LESK, who states that a digital library is "a collection of information that is both digitized and organized." (LESK, p. 1). Following this definition, this chapter can be seen most generally as an exploration of the various ways in which people are approaching these double problems of digitization and organization, and how their solutions are gradually creating the new form of the digital library.

This chapter is divided into five general areas; first, an introductory section deals with the cultural and social aspects of digital libraries. The following sections treat technological issues: searching in a loosely-coupled distributed system; organization of a distributed collection; indexing, search, and retrieval in tightly-coupled libraries, with specific examples; and, finally, problems with archiving in a distributed environment. A concluding section discusses future work.

CULTURAL AND SOCIAL INFLUENCES

The traditional provider of information in modern society has been, up to the advent of computerized information services, the library. The most general function of the library, in both its public and private incarnations, is as a gathering center for information[2]—which, until recently, has always been artifactual. The sheer concreteness of books, journals, and other collected materials has necessitated that the basic form of the library be centered around a repository; the implication of substituting (or even just adding) digitized distributed information in this system is that, at the very least, the shape of the library and its services will

[2] Although the role of the library in society has always been much more than that of a repository!

evolve—but, as many authors note, the social and cultural changes wrought by this shift will be at least as dramatic as the structural ones.

Library and Internet Culture

The joining of the traditional library with the distributed information network has prompted some writers to examine the merging of "library" and "net" values, which they see as combining dialectically to create the culture of the digital library. The traditional library, claim LEVY & MARSHALL, is culturally associated with "notions of fixity and permanence" (p. 5, White Horse) that have almost automatically been carried over into the conception of the digital library, which is then expected to also exhibit these qualities. Contemporary authors and philosophers, however, almost invariably characterize current postmodern culture as fragmented, fluid, and ephemeral; YOUNG even argues that the impact of this new culture is most "pervasive" in the fields of library and information science. Indeed, many of the current controversies in digital library development can be seen as manifestations of the clash between traditional librarianship and a new, free-flowing "cyberculture," managed by technologists from a computer science background. The break made by computer-oriented managers with standard library practices is often radical enough to be the source of heated argument—see HENRY, who finds fault with the "new paradigm" of digital preservation, or CRAWFORD's polemic against the purely digital library, for example.

Digital material is naturally ephemeral and aspatial, and it is this tension between solidity and impermanence which is one of the core issues governing the development of the distributed digital library—how and to what extent should a digital collection emulate the virtuous "solidity" of a set of concrete objects?

It is not only the insubstantiality made possible by the distributed network that will influence the development of the library, however, but also the cultural ideas which have sprung up to accompany new technological capabilities. Not just the computer qua machine, but also the very idea of the computer as a "metaphor for personal identity" (YOUNG, p. 113) will profoundly affect the ways in which society—and the institution of the library—will develop. In this sense, it is notable that BUSH, in his seminal essay "As We May Think," describes a mechanized electronic library as "an enlarged intimate supplement" to memory, evoking already in 1945 a hint of cyborgian melding of reader and machine.

On the other hand, there are also points of similarity between what are traditionally seen as research library and Internet values. (ANTELMAN & LANGENBERG) A major goal of the university library has always been the free exchange of knowledge and information; on an even wider level, this goal has also been shared by the American public library system since its inception. In the age of the Internet, this ideal seems to have been elevated by some to an almost religious, unquestioned belief (LEVY, DLib 1/2000, p. 4); indeed, FOX & MARCHIONINI go so far as to describe information as "a basic human need." (p.

31) Furthermore, Internet and research library cultures, seeing information as a valuable good, are generally both opposed to censorship and commercialization, which could be seen as hindrances to the free flow of information. ANTELMAN & LANGENBERG attribute this similarity in values to the common development of both in the university community (p. 54)—though these values have often been seen as shared for the most part by the entire public library community. ATKINSON goes so far as to root the library's moral strength (its "ability to uphold social ethics" (p. 247)) in its nonprofit status; equally vehement pleas against commercialization of the Internet on moral grounds are too numerous to count.

Information and Knowledge

The cultural valuing of information as a good per se leads in the extreme to a purely quantitative valuation of library service, where more is always better, in and of itself. (See COFFMAN, for example, for an enthusiastic discussion of the possibility of creating the largest library on earth.) This drive to deliver the most information possible reveals an important consequence of distributed digital information for the organization and structure of the library: When vast quantities of information are readily available electronically, what does the library become? Before the advent of the Internet, libraries were financially and spatially limited in what they could collect; the librarian acted as selector and sifter of information, choosing the artifacts, which would be collected and maintained. The implication of fully-networked world-scale information communities, however, seems, paradoxically, to presage a reduced role for the library: If all the information in the world is suddenly available, and nothing need be excluded because of spatial or budgetary restraints, what is the institutional function of the library?

As early as 1979, BOORSTIN pointed out the disadvantages of the information glut, in his reformulation of Gresham's law: "Information tends to drive knowledge out of circulation." (p. 3) A main point of BOORSTIN's speech is that a vital part of libraries is their intentionality—they exist not just as repositories, but as collections thoughtfully developed by people for a specific purpose. The danger of enthusiastically embracing wide-scale digital distribution at the expense of traditional libraries is that systems which collect vast amounts of information "just because it is available" will displace systems of knowledge—institutions which have been developed to further or to advance some human intention. Along the same lines, CRAWFORD asserts that communities want libraries to be mainly not a source of information, but a source of books. Seen this way, the expansion promised by a distributed library could in one sense create a sort of diminishment of services provided.

The Electronic Librarian

This problem of purpose, or creating knowledge out of information, relates to the distributed digital library on many levels, centering on the question of how a library can create a knowledge structure out of (or impose one on) an unstructured sea of information. Many authors (ARNOLD, HARTER, YOUNG, ODLYZKO) foresee a fundamental shift in the role of the librarian in a distributed digital environment. Instead of tending to physical collections as in the past, the librarian will become a "knowledge navigator"—a mediator between the patron or researcher and the trackless network. (Perhaps corresponding to the "trailblazers" first described by BUSH fifty years ago.) When the idea of the library is expanded to include distributed collections that have not been grown and cultivated by its own staff, the metaphor of librarianship changes from that of the gardener to that of the tracker: Instead of carefully planning and developing in a controlled environment, the librarian must carve paths through the wilderness of a network not of his or her own making, becoming an explorer as well as a guide (see GRIFFITHS for an excellent in-depth discussion of this transformation; WARD and SCHWARZWALDER are also relevant).

Similarly, LANHAM sees a metamorphosis "from [a] curatorial to [an] interpretive" role, describing the librarian's new function as the constructor of "human attention-structures". This function is a response to what he sees as the need for an "economics of attention," to counter the problem of overabundance of information in a networked environment. ARNOLD addresses this change head-on in his "The Electronic Librarian is a Verb …" lecture, describing the cyberlibrarian as heroically “creating a syntax of digital knowledge,” (p. 5) in a library which has become “more a state of mind than a location […] a set of neural connectors.” (p. 12)

In a more short-term sense, bringing distributed digital systems into a library’s purview will create very tangible problems for the librarian—budgeting, for example, will need to be adjusted, if shelf space and acquisition are to be replaced or supplemented by net-accessed information. (YOUNG, ARNOLD) The process of acquisition itself will also be changed, as libraries begin to link to resources not physically present, instead of actually bringing objects into the library. Finally, ideas about archival management will change, if the contents of a library shift from primarily concrete objects to mainly items that are only virtually accessible.

Social Aspects of the Library

Another aspect of the shift to distributed information systems is the change in the social dynamic of the library. If the library becomes purely (or even just mostly) distributed, disappearing as a concrete “place,” the social interaction facilitated therein will be lost, unless it is deliberately added back in, in electronic form. One role played by the traditional public library is that of the community center, offering programs and gathering possibilities for groups of patrons. Furthermore, in a university setting, the library can promote serendipitous encounters between students and among faculty members, thus serving a truly social purpose.

More to the point, perhaps, social interaction can be an integral part of library work, as researchers exchange information while searching. (LEVY & MARSHALL, ACKERMAN) LEVY also cites a case study of the behavior of information analysts to show the centrality of collaborative activity (marking of papers, discussion) to the enterprise of research.

Here, the structure of the digital library intersects with the field of computer supported cooperative work, which generates various projects to facilitate group communication over a network, both in real time and as the construction of a "collective memory" of annotations. We can go back to BUSH to find a first description of this idea of the collective memory, in the "memex" used to create trails of linked information, which can then be shared with other researchers. Generally, common tools such as e-mail, listservs, chat rooms, and remote conferencing can all be seen as electronically-supported means for group communication; these projects, however, tend to be general-purpose, not specifically developed for digital library use—and there are simply too many of them to begin to describe in this chapter.

SIMMONS gives an interesting description of the "ideal" electronic collaborative tool, characterized by ease of use, support for all kinds of media, and freedom from the restrictive keyboard and chair. This tool, furthermore, would be outfitted with artificially intelligent computational agents available to each user, and provide a large display—perhaps even a virtual reality. Basically, SIMMONS wants to fully use the capabilities of networked communication to promote the freest, most spontaneous and complete exchange of ideas possible.

MARSHALL also discusses the role of annotation in digital collections, noting that the traditional library's age-old admonitions against marking in books can now happily be reversed in a digital setting, as readers' annotations "may become important adjuncts to the primary text." (p. 131) MARSHALL's study of student behavior with respect to annotations in university textbooks demonstrates the value even of informal, disorganized reader-to-reader communication. Distributed digital collections, she feels, can facilitate this sort of collaborative reading and discussion of a work in a more open, well-organized format, by providing the reader with books specifically designed to be marked up—and thus new value is added to the library collection. Indeed, digitally annotated books could be seen as even more valuable than their untouched non-distributed counterparts.

One project in development which will incorporate interactive user-to-user communication is the National Engineering Education Delivery System (NEEDS) (NATIONAL ENGINEERING EDUCATION DELIVERY SYSTEM), which is being specifically designed to not only provide distributed digital information for a specific educational purpose, but also to generate and nurture a "community of learners." User-added reviews and discussion groups are currently part of the NEEDS system, but part of the plan for future work involves the addition of "pedagogical metadata," allowing educators to tag digital objects with information about its educational applicability. (MURAMATSU)

The advantages of the distributed research collection, then, extend beyond the wide and instantaneous dissemination of digitalized text, with advanced searching and visualization technology. Even as the digital library reduces research to a more solitary activity, with patrons no longer gathered together at a single location to access resources, it can also make possible a whole new level of collaborative work, as readers-cum-annotators are able to grow simple texts into discussions across time and space. (ROBERTSON ET AL)

THE LOOSELY-COUPLED DIGITAL COLLECTION

The question of what should be "contained" in a digital library can be answered in many ways. The strictest definition is perhaps that given in 1994 by MIKSA & DOTY, who insisted that the definition of a library presupposes the existence of a bounded collection. The freest definition is that of the "anarchic and individualistic" Internet (HARTER), which lacks control and organization.

LAGOZE & FIELDING, taking into account the possibilities of unlimited information access as well as the need for the imposition of some sort of organization on a collection, define a digital library from two perspectives, the logical and the operational. In the logical sense, they see a collection as definable by a "set of criteria for selecting resources from the broader information space." Operationally, membership in a digital collection can be defined "in terms of resource discovery:" the digital library consists of all those resources which can be found using the library's resource discovery tools (excluding those objects found only in links from the directly discoverable set).

Often, the trade-off seems to be between the selection of a collection itself, and the power of the indexing machines that service it. If a collection is not well-indexed, or not easily searchable, it should be small, and works without value should be excluded, or it will become difficult to use successfully. If, on the other hand, we can build sophisticated search engines, and clever indexing algorithms, capable of teasing a value hierarchy out of the text contained in a large collection of information, it becomes less harmful to let as much information as possible flow into the system.

A Single Distributed Library: The Control Zone

An interesting compromise solution to the problem of scope is ATKINSON's idea of a "control zone" within an anarchic networked system. To create a control zone, a group would carve an organized, systematized space out of the uncontrolled Internet. This space would be a "single, distributed digital library…created and managed by the academic library community" (ATKINSON, p. 239), with information only coming into the zone after a process of review and selection. This idea, then, is a variation on the theme of the single universal library—an idea that seems to have always been with us. As with many such schemes, the possibility of its realization would in the end be dependent on patient diplomacy and committee work, to bring together the participants necessary to make the zone function on such a large scale.

In the structure of the zone, ATKINSON sees the librarian's function as adding value to information through his or her work. In this sense, then, librarians should distinguish one book from another by adding access value to books of high worth (putting them in the zone), and denying it to those of lower value. The librarian, then, becomes an evaluator of worth in some context, taking over part of the filtering role that is, in a print environment, to a large part fulfilled by the publishing community.

ATKINSON expands the idea of the value signifier in the control zone beyond the mere Boolean function of inclusion/exclusion by proposing the addition of two new types of metadata to accompany work in the zone. The first, "use level", would track the history of use of digital objects—in this way, it could be determined which items have been accessed by students, scholars, experts, and so on. Information about the accessor's status would be used to weight the importance or relevance of a given work. An object referenced often by expert users in a certain discipline, for example, could be automatically marked as highly important; whereas another object accessed just as often, but by student users rather than scholars, could then be marked accordingly. The second type of metadata, “work level”, would differentiate by level of difficulty and specialization. Although setting a work level would seem to require careful human deliberation, automatic use tracking is already being done today.

Use Tracking to Facilitate Searching

The web search engine Google (GOOGLE), for example, could be seen as implementing a basic kind of static “use level” algorithm to aid searching. Although the algorithm for this engine does not explicitly track user behavior in the web, it directly adds “access value” to pages depending on how many links on the net point to them: in other words, it tracks “net publisher” behavior. This means that the more web page creators decide that a specific site is valuable enough to be linked to, the more likely it is that this site will be returned by a search in the Google engine.

A dynamic prototype for user tracking to establish research-aiding metadata can be seen in KANTOR’s AntWorld project, available on the Internet (see KANTOR) (though this project does not go so far as ATKINSON suggests, in that it does not take the professional or academic status of the user into account when weighting their evaluations—concern for user privacy is understandably a major issue in user tracking projects). AntWorld is a collaborative system to facilitate Internet searching by allowing a networked group of users to augment pages in the web with information about their value. An Internet user can join the AntWorld by downloading software from the web site, and running an Antscape browser. This browser allows the user to enter a textual description of a "quest" (a search goal), and then, as he or she searches, to annotate links found with information as to how relevant they were to that quest. This information is then used to help other users evaluate search results. The basis for the AntWorld approach is the biological model of insect communication through pheromones; hence, the AntWorld term for meta-information about a link's value is Digital Information Pheromone. (KANTOR)

Similar work is being done by a group in France (BOUTHORS & DEDIEU), who have developed Pharos, a collaborative information sharing tool that allows users to contribute to a set of databases of annotations on web pages. These annotations are then accessible through a browser assistant, which tracks pages browsed and displays relevant information. Pharos uses weighting algorithms based, in part, upon the similarity of recommendations made, to automatically detect correlations between users; thus, a user is situated in a group of (theoretically) similarly-minded annotators, and can benefit from their evaluations without being flooded by the less useful information of an overall "average" evaluation.

Work funded as part of the second phase of the Digital Library Initiative is currently being done at the Oregon Graduate Institute of Science and Technology to create a system that can capture the document selections of experts and use them to aid subsequent problem solvers. This system to "track footprints" seems to be very similar to that which ATKINSON proposed for setting use levels in his control zone: The documents chosen by expert solvers are preserved in a trace, describing the "path" he or she took through a given collection. Navigation tools will then exploit this knowledge to help future solvers in their searches. Development of this system is still in its early phases. (GORMAN ET AL.)

The "Walden's Paths" system developed by SHIPMAN ET AL can be seen in one sense as the most highly directed path-based searching aid. This system allows educators to directly create paths for student use, as opposed to merely facilitating the passive gathering of information about sites as a side effect of regular browsing. By introducing an intermediary "Path Server" between a student's browser and the web, information about paths can be automatically provided when a student downloads an annotated page. SHIPMAN ET AL describe paths so created as "metadocuments"—here in the digital library, then, the functions of librarian and author meet, as the path organizer with his or her work lays down a new knowledge structure over the existing mass of digital information. This process of situating knowledge in the context of a path to aid students involved in a specific task is, then, at the most directed, organized end on the continuum of searching aids.

Hierarchical Distributed Dynamic Indexing (HDDI™)

A novel approach to organizing large quantities of loosely coupled material is a system in development jointly at Lehigh University and at the National Center for Supercomputing Applications at the University of Illinois which will automatically create hierarchical models in a distributed, dynamic environment (see POTTENGER, BOUSKILA & POTTENGER). This strategy is based on the algorithmic creation of subtopic regions of semantic locality in sets of distributed documents; this allows automatic discovery of similarities at a fine level of granularity amongst concepts within documents. In this way, hierarchical indexes (such as those created now "by hand" in many places on the web; YAHOO is probably the most well-known example) are generated for topics in documents in a volatile, distributed environment, providing the information seeker with an always up-to-date map of information spaces. The ability to generate large hierarchical indexes on the fly allows for a realistic, useful mapping of cyberspace without the need for time-consuming human intervention. This technique is most valuable when applied to documents within some institutional zone — to map out, for instance, large sets of corporate documents. Here, subjective issues relating to "importance" or "quality" can be sidestepped, and the power of the HDDI™ strategy can be fully leveraged—an unstructured set of documents lacking any sort of metadata can be bound to a hierarchical knowledge structure generated automatically based on word frequencies.

DISTRIBUTED COLLECTION MAINTENANCE

The addition of distributed digital information to a collection presents unique opportunities and problems for the maintaining staff. As mentioned above, information stored in distributed systems tends to be both volatile and ephemeral. The information stored in Internet accounts, for example, is at the mercy of the owners of the computers where these accounts are stored—providers, who naturally tend to treat these accounts as short-term, nonpermanent space. (POCKELEY) More important, perhaps, is that digitally provided resources cannot be fixed (except in an artificial sense, as when a digital copy of a paper object is created—digital objects are not naturally fixed in one form in the same way that non-virtual objects are); they are continually open (and therefore also, often, subject) to change. Almost any digital medium is always a sort of slate, or palimpsest, capable of being erased and rewritten without much ado—highly unlike ink on paper in book form. For now, much that is published on the web behaves as if it were indeed concretized in a certain form, and is not changed; this situation, however, could be a temporary part of the movement to digital media, in which the old modes of work are carried over until new ways, which take advantage of the new possibilities presented by a digital format, come into being. The problem of constant reorganization and changing location of documents, however, is very much present even now, and on a distributed system, where information is provided in a set of links, this continual evolution can wreak havoc with non-dynamic organizational systems.

The issue of changing links reveals another source of problems in a distributed environment—the tension between local and global needs for information access. The addition of a distributed system to an information collection introduces a new layer of dependency, as maintainers must now depend upon other institutions for the upkeep of the non-local portion of their collection. Some data show, however, that currently fifty percent of URLs are not available after two years (PASKIN & STICKLEY). Proposed solutions to this problem deal with systems from the very loose to the tightly coupled—from systems in which the library provides access to unknown, uncatalogued resources (such as the Internet), to federations of libraries working closely together to achieve a common goal.

Persistent Object Identifiers

Perhaps the highest-level approach to the problem of continually changing information networks is the development of permanent link systems, so that local reorganization of sites will not affect global access to resources moved from one location to another. HODGE mentions several of these. First, the PURL project (PERSISTENT UNIFORM RESOURCE LOCATOR TEAM), supported by the Online Computer Library Center's Office of Research, is creating Persistent Uniform Resource Locators—URLs that point to an intermediate resolving service that returns the actual location of the desired resource. (The resolving URL itself, then, must be unchanging.) The PURL creators themselves see this solution as only temporary, until Uniform Resource Names (URNs) have become standardized and widely supported. (WEIBEL ET AL, SHAFER)

URNs, if implemented, would be persistent (they should last "longer even than the Internet," states ARMS (p. 235)), globally unique non-location-dependent resource names. (SOLLINS & MASINTER, LYNCH, ARMS ET AL) Like PURLs, however, URNs would need to be resolved into URLs in order to be used over a network. The basic problem, then, in developing persistent links, is the need to reconcile a static global name system with a dynamic location system—or, more to the point, it is the need to standardize and coordinate the adaptation of such a system across the Internet.

Another recent project to treat the problem of broken links is the Robust Hyperlink, proposed by PHELPS & WILENSKY. A Robust Hyperlink can be implemented using the semantics of a regular URL, augmented with a lexical signature computed from the reference document. In the example given, the signatures are created by taking the terms in the document with the highest "term frequency-inverse document frequency" (TF-IDF) values. This signature, then, can appear as a query appended to the URL, or as part of the HTML markup, etc., where it can then be used as a query submitted to search engines to find the keyed document. The authors claim that only a very small signature (about five words) is sufficient to facilitate quick location of individual documents, even in a space as vast as the web. The advantages of this scheme are that it is lightweight, simple, and can be immediately implemented. The authors describe it as an example of the web "bootstrapping" new features upon those already developed. The disadvantages seem mainly to be the same as for most of the permanent link schemes—they are not yet in general, widespread use.

One important project, described as the "ISBN for the 21st century" (PASKIN & STICKLEY), is the Digital Object Identifier, which would (as the name suggests) be a unique identifier for digital content. DOIs were developed specifically to facilitate the "management of copyrightable materials in an electronic environment" (PASKIN), by creating a system for managing permissions and facilitating transactions on digital objects. The DOI is an abstract specification of an identifier and the system to process such identifiers; a prototype has, however, been built, based on the CNRI Handle System, an identifier system capable of working with URNs (which, themselves, were also developed at CNRI). The Handle System is used, then, to translate DOI to URL format, giving a location to a non-location-specific identifier.

The CNRI Handle System is perhaps the most extensively developed implementation of a global naming service currently in use. Like the systems described above, the Handle System uses globally unique names, which are then mapped by an organizing authority into their physical locations. The system works by creating a confederation of name spaces; this has the advantage that individual name systems can join a handle system and still retain their local names. Resolution management is based on a hierarchical model, so that names can be resolved into locations either by a local handle service or by a global registry—or by both. This allows the resolution process to be replicated and distributed across a system. The development of this system, planned in part to provide a framework for digital library infrastructure (SUN & LANNOM, part 6), influenced to a large extent the evolution of the Networked Computer Science Technical Reference Library (NCSTRL) project, to be described later. The best introduction to the architecture of a handle system can be found in KAHN & WILENSKY's 1995 paper; the system itself is thoroughly described at CORPORATION FOR NATIONAL RESEARCH INITIATIVES, where a Handle Resolver can be downloaded to process Handle links even now.

Metadata

Metadata—cataloging or indexing information about an object—could easily take up a chapter of its own; here, we shall touch only on the newest aspects of metadata relating to distributed system development. In the distributed library, it seems that metadata has become more important than ever before, because it can facilitate the organization and management of networked information. On the other hand, some see the possibility that metadata for searching will become less significant in the future, as search engines are able to manipulate large amounts of text more rapidly, and thus become less dependent on finding aids (HARTER, p. 6).

Currently, the biggest obstacle to sophisticated wide-scale use of metadata in distributed systems seems to be not technological insufficiency, but lack of common standards. Imposing a single metadata protocol is difficult enough in a centrally-controlled digital library; in a confederated system, where each local node has its own user base and requirements, it can become extremely complex. The problem of integrating local and distributed cataloguing information is known as the problem of establishing interoperability—of creating systems that can get information from one another in a useful fashion. One method of allowing for information exchange is to make sure that "crosswalking" (automatic translation) between metadata formats is possible.

There are many projects aimed at developing systems of interoperable metadata. One tendency is to use a leaner, more abbreviated format such as the Dublin Core, to store only the most essential information about a work. HODGE mentions that it is also hoped that using a compact format will mean that publishers can provide metadata directly, so that the need for independent cataloging falls away (p. 8).

Resource Description Framework

An important project which is attempting to address the problem of incompatible metadata standards is the Resource Description Framework (RDF), developed by the World-Wide Web Consortium, along with others. RDF is basically an extension of XML (eXtensible Markup Language), and can be used to describe any resource that is uniquely identifiable by a Uniform Resource Identifier (URI). RDF has a simple data model, in which resources are associated with property-types, which can in turn point to other resources, or simple values (such as strings).

Such a simple data model, combined with structural and semantic rules, can be used to encode information from widely varying metadata in a single format. This heterogeneous encoding is achieved by using the XML namespace mechanism at the beginning of a record to give a pointer to a resource that has all the information about the metadata fields used in the record itself. Once a list of references has been given, providing a format for each metadata scheme to be used in the record, the XML tags in the data model are set up to include both the name of the tag, and the metadata model to which this tag belongs. Basically, instead of forcing all records to fit into a common scheme, a Resource Description Framework augments each tag with information about the metadata scheme to which it belongs. (MILLER gives an excellent introduction to this technology.) Thus, the RDF infrastructure can be used for the exchange of metadata among widely varying information-gathering communities; this sort of mechanism is pivotal for the creation of federated digital libraries on a large scale.

The Open Archives Initiative

Another interesting current project that is exploring metadata interoperability from another perspective is the Open Archives (formerly the Universal Preprint Service) initiative (OAi). The OAi makes an excellent case study for this chapter—not only because of its currency, but because it relates to many of the issues of loosely coupled distributed systems, on several levels. First, the OAi is concerned with developing infrastructure to support interoperability between digital collections. Also important, however, is that the nodes in the distributed system are author self-archiving (also known as e-print) systems. This initiative, then, aims to transform scholarly communication (VAN DE SOMPEL & LAGOZE) by taking full advantage of the capabilities of thoroughly networked communication systems. In this new paradigm, scholars can themselves disseminate information and results quickly on a wide scale, avoid giving up the rights to their work, and bypass the rigidity of peer-review, and expensive journal costs (VAN DE SOMPEL & LAGOZE).

Author self-archiving systems have been in existence for at least ten years, and their use is growing. One of the major archives is arXiv.org, at Los Alamos, established in 1991 by Paul Ginsparg for physics papers, but now also including other technical areas. Another is the NCSTRL (Networked Computer Science Technical Reference Library) collection of computer science reports, which is itself based on a distributed model, with services exchanging information using the Dienst protocol. There are countless others—too many to mention here.

The aim of OAi, then, is to facilitate search and retrieval services that span these archives, in part by establishing protocols for interoperability. The OAi's Santa Fe Convention, a set of specifications created at an October meeting in Santa Fe, lays out "a technical and organizational framework designed to facilitate the discovery of content stored in distributed e-print archives." (SANTA FE) This convention, then, dealt with interoperability in metadata harvesting—gathering of information about documents stored in the archives. First, a core set of metadata elements was established—the Open Archives Metadata Set (OAMS). This set contains only nine elements, for maximum interoperability (and searchability at a coarse level of granularity) between dissimilar archives. Second, the convention set forth that XML would be used for representing the OAMS, as well as local metadata sets. Again, XML is an excellent choice for advancing interoperability, as it is highly flexible, and growing in popularity—not an unimportant factor in a field which has not yet settled into a group of common standards. Finally, it was agreed to use the Open Archives Dienst Subset protocol to exchange information about OAMS, as well as archive-specific metadata (VAN DE SOMPEL & LAGOZE).

Following steps outlined at the Open Archives web site, e-print providers can bring an archive into compliance with the Santa Fe convention and register it with the OAi, thus making the data available as a node in a distributed system of archives. Service providers can then establish search engines to run over the available archives, accessing all the information through the interoperating protocols. At Virginia Tech, for example, FOX and others have done extensive work building on the Open Archives system, including the development of an "OAi Repository Explorer," which allows browsing of OAi-compliant archives

Digital Object Models

At a different level of abstraction from that of interoperability between metadata, libraries must deal with the problem of the format in which the contents of the library itself can be stored and presented. All materials stored in digital libraries must obviously be digitized in some standard format. This means that an issue which must be resolved to enable large-scale distributed libraries is the construction and standardization of effective digital object models. At the highest level, digital objects stored in libraries should exhibit enough uniformity to be stored, accessed, and presented using the same protocols. On the other hand, a wide variety of formats (text, sound, video) should be available across the system, each with its own methods of use and display. The most influential abstract digital object model seems to be that which was proposed by KAHN & WILENSKY, in which a digital object is divided into three components: a handle (unique identifier), a metadata container, and a data container—most current ideas and implementations of digital objects expand in some way upon this pattern. One project based on these theories which delivers the multi-level functionality desired from digital objects is the Flexible and Extensible Digital Object and Repository Architecture (FEDORA), developed at Cornell University, developed by LAGOZE ET AL.

FEDORA

The basic FEDORA digital object model is that of interoperating components. Currently, documents are generally provided in a variety of formats (such as HTML, PDF, image files, etc.), each of which requires a specialized program for viewing or manipulation; a goal of FEDORA is to make this specialization transparent, so that the user can access heterogeneous media over a single interface system.

A FEDORA object has, at the lowest level, packages called DataStreams, which contain the content of the object itself in some form. (These DataStreams can be either physically associated with the object, or distributed themselves, appearing in the object only as a link). Above these content packages are interface components called Disseminators; these components provide functions that allow a user to actually access and use the content of the object's data streams. The power of the Disseminator concept is its flexibility: each Disseminator can be associated with a different sort of interface to the object—one can provide a text view, another could give metadata information, etc. By creating different Disseminator components for different access methods, the behavior of an object can be made context-dependent.

On top of the FEDORA object model is the repository model. In this scheme, a Repository component manages storage of and access to the digital objects, which are handled at a high level, as interchangeable black boxes. This encapsulation of various formats into a generically manipulable object type greatly simplifies collection maintenance. More importantly, perhaps, for a developing system, the model is designed to be easily extensible, and to facilitate the addition of rights management schemes. Pivotal to this extensibility is FEDORA's data packaging system, which, as we have seen, wraps content and presentation in a standardized package for manipulation by storage programs; thus new item and presentation types can be easily folded into a pre-existing repository. DUSHAY & PAYETTE provide a good short overview of FEDORA; PAYETTE & LAGOZE give a more thorough description.

Multivalent Document Model

Another digital object model, developed at Berkeley as part of the Digital Library Initiative project, is the "multivalent document model," designed specifically for openness and extensibility across a distributed system. (WILENSKY & PHELPS) In this model, objects are separated into layers of content, outfitted with functional modules called "behaviors"—small reusable programs that can be loaded dynamically. An interesting behavior which has been added to this model is a distributed annotation scheme, which allows users to mark up the objects themselves; a feature of this system is that the markings appear directly on the documents, as if they were part of the object itself. A prototype of this model has been implemented in Java; details can be found at the MULTIVALENT DOCUMENT HOME PAGE.

SEARCH AND RETRIEVAL IN A DISTRIBUTED ENVIRONMENT

Distributed digital objects are maintained and outfitted as described above with an appropriate identification system (and perhaps metadata) to facilitate the main business of the library—search and retrieval. The creation of interoperable metadata and unique, global object identifiers gives a foundation upon which the superstructure of a search system must be constructed, to do the work that gives the collection value above and beyond that of a mere repository. Indeed, when the supply of knowledge objects grows limitless, as more and more repositories are linked through networks, search algorithms and procedures may become the definitive core of a digital library. Collections, in the sense of a set of items discoverable through a certain interface, can become specialized and individually configurable—dynamic and responsive organizational systems in and of themselves. Here, we briefly discuss the main issues involved, and give a case study of the globally distributed NCSTRL digital library of computer science papers.

Distributed Catalog Systems

The most basic, straightforward approach to cataloging distributed digital objects for retrieval is probably the centralized union catalog, where information about all available resources is gathered and pulled into a single location. Searches are then run quickly on this single machine, and need not pull information down from a network with each request. Most Internet search engines, for example, currently use a single merged catalog. The University of California's online MELVYL catalog for the California Digital Library system, which pulls information from twenty-nine separate facilities into a single database, is an example of a digital library system, which uses this approach. (COYLE) The largest union catalog currently in existence is the OCLC Online Union Catalog, which includes 35 million bibliographic records.

Cooperative Online Resource Catalog

One project currently under development which extends the idea of the large union catalog to material available on the web is the Cooperative Online Resource Catalog (CORC). This project, under the auspices of OCLC, is constructing a cooperative web catalog through the distributed efforts of member libraries. In a cooperative process reminiscent of ATKINSON's control zone, staff at individual institutions create records of web-based information, which are then merged into a central database to be shared with the rest of the participating community. Although the work of adding resource records must still be done by hand, CORC's record editing tools help speed data creation by automating much of the data collection process, filling in fields with information that can be machine-harvested. The CORC project also provides libraries with resource access tools that can be dynamically added to local web portal pages; in this way, library service can be melded with web searching services to provide an integrated point of access for a vast body of networked information—much of which has been selected and screened for quality by library personnel. Indeed, the real value of the CORC project lies perhaps in the cataloging standards which will need to be developed for the system, because it is in the creation and maintenance of such standards that library staff will guard the quality of the material recorded in the catalog; the combined personal time and effort of selectors distributed across a wide system sets CORC apart from a purely automatic cataloging project. The CORC project, currently still in a developmental phase, is accessible on the Internet from ONLINE COMPUTER LIBRARY CENTER.

While a centralized union catalog ensures speedy search times, it has disadvantages in large-scale systems. First, there is the obvious redundancy involved; each record at a remote site must be copied and stored locally. Perhaps more importantly, metadata records need to be molded to fit into a unified pattern for the single catalog, without undue homogenization and loss of information. Melding all records into one catalog makes it difficult for individual collections to maintain formats and metadata specific to their community. Furthermore, the dynamic nature of networked information space presents a real problem for a master index approach, as new information is continually being added and updated across a large library federation: the catalog would have to be in a state of perpetual change to reflect the volatility of the system.

The Virtual Union Catalog

One approach to these problems is the concept of the virtual union catalog, currently being tested in the University of California library system. In a virtual catalog, records need not be merged together and stored in a redundant collection; instead, the grouping is generated by searching through distributed catalogs in real time. COYLE's description of the testing done on this system reveals the sort of problems that plague distributed systems on every level: lack of common standards makes query formulation that will work on all systems next to impossible; system downtime at local nodes compromises overall searching comprehensiveness; and the sorting and merging of a large set of retrieved records is highly computation-intensive.

NCSTRL and Dienst

Probably the largest distributed library system currently on the Internet is the Networked Computer Science Technical Reference Library (NCSTRL). NCSTRL provides access to over 30,000 documents from over 100 educational and research institutions around the world. Key to the NCSTRL approach to combining heterogeneous libraries is an open architecture system: that is, protocols are given to specify interfaces to a set of digital library services, but each local organization is free to implement these interfaces in any way it wants. In this way, libraries can choose technologies appropriate for local needs, while still satisfying the requirements for membership in a distributed confederation.

Another benefit of the open architecture approach is its extensibility; the general system can be augmented with services as required—an important feature of a growing digital library where standards are not yet set in stone. This modularity also allows institutions to plug into the library system as mediators, offering nonstandard, customized services. (See LEINER for a more detailed description of NCSTRL's open architecture system).

The technical infrastructure for this federation of libraries is provided by a distributed library protocol called Dienst (Distributed Interactive Extensible Network Server for Techreports), which was developed by Jim Davis of Xerox and Carl Lagoze at Cornell University. This protocol was created specifically to allow distributed searching of locally managed collections. The Dienst protocol provides for four areas of library service: user interfaces, indexes, collections, and repositories. A user interface provides a front end for people to access the library system; these interfaces then communicate with index servers, which maintain metadata about information stored in repositories—separate servers which can then feed document content to the user interface itself.

When Dienst was first developed, each institution had its own index, and any query had to traverse the entire networked system, gathering information from every server so as not to miss the material from any collection. This naïve form of distributed searching, in which quick response time was bounded by the slowest server or connection, failed to work well for international-scale systems, so a new regionally-based system was devised as NCSTRL expanded.

In the current version of Dienst, the network is broken up into several regions of well-connected servers, called connectivity regions. Each of these regions has a single collection server and several index servers, which maintain replicated metainformation about the entire NCSTRL collection. Thus, a query from one connectivity region is directed to its regional collection server, which need only send queries through its own area, avoiding long waits for data from slow or badly-maintained interregional connections. A single central collection server, then, maintains information about the regional servers and the index servers which feed into them; this central manager communicates with the regional managing servers, who then in turn serve as local controllers for each region of connectivity, forming a hierarchy of distributed control.

Smart Object, Dumb Archive (SODA)

One new approach to facilitating interoperability in distributed digital libraries is to endow digital objects themselves with functionality, thus removing the whole need to deal with different object formats at the collection level. In this Smart Object, Dumb Archive approach, the collecting mechanism becomes a simple tool for gathering and disseminating "intelligent" objects ("buckets"), which are themselves capable of "enforcing their own terms and conditions, negotiating access, and displaying their contents." (MALY ET AL, 1999) NELSON ET AL (1999) point out that this extreme modularization of digital library services, archives, and objects will allow each of these areas to develop independently, without dependence on the other two. Another benefit of moving functionality into the object is that it breaks the strong connection between an archive and the objects it contains. In most digital libraries, the structure of a digital object is dependent upon the archive in which it appears—but in a distributed, dynamic environment, where objects are accessible across varying systems, this tight coupling tends to be an annoying source of interoperability problems. (MALY ET AL, 1999) Finally, buckets are capable of aggregating many different data types into a single package, so that video, text, images, etc. can be presented in one unit, each packages with their own display methods.

The SODA model was specifically designed to improve on the NCSTRL/Dienst distributed library system. One objection the developers raise to Dienst (as well as other digital library frameworks) is that archive access protocols "have become unnecessarily complex." (MALY ET AL, 1999, p. 5) Indeed, one reason brought for the simplification of archives with regard to object interoperability is that this can free archives to become "smart" with regard to functions that don't just duplicate mechanisms which are better associated with digital objects—thus evolving into a future SOSA (Smart Object, Smart Archive) format. NELSON ET AL (1999) note that another drawback to the Dienst system is that it must explicitly define the definition and structure of a document, so that information about all possible media formats in use must be effectively hard-wired into the protocol. More logical and straightforward would be to decouple such information from the archival system, and let the object take care of its own format.

Old Dominion University and NASA Langley Research Center have developed a prototype testbed implementing this concept based on the existing NCSTRL system, called NCSTRL+. NCSTRL+ uses the Dienst protocol, simplified to no longer control the presentation of documents to the user, and modified to handle bucket objects. An idea for dividing collections into partitions, called "clustering," has also been introduced into the system, thus giving Dienst the ability to subdivide a collection along something other than an institutional boundary. NCSTRL+ provides clusters such as subject category, source language, and publishing institution. NELSON ET AL (1998) give a detailed description of the early phases of the NCSTRL+ implementation.

PRESERVATION IN A DISTRIBUTED DIGITAL ENVIRONMENT

If one important function of the library is storing and facilitating access to documents, another is preserving the documents in a collection for long-term storage. The problem of preservation is especially complex in a distributed information system like the Internet: If a network of informational material is constantly evolving, at what point should one take a "snapshot" of the available material? And what about hyperlinks—what are the boundaries of a document, which is tightly woven into an information network? Again, the basic tension arises from the attempt to force virtual objects to behave as if they were concrete—from the desire to archive part of a virtual system using methods derived from traditional practices of object storage. This insistence on forcing new technologies to adapt to an older paradigm seems to mean that there will necessarily be a certain amount of arbitrariness involved in any decision about what parts of a hypertext document should be archived. HODGE writes that some organizations preserving hypertext documents store only links, but not the information therein—which has obvious disadvantages in an evolving system. Others store link content only from certain trusted sites, which is better, but still leaves opportunity for information to be lost. (The idea of the permanent digital object identifier, described above, could be a partial solution to this problem.)

Cultural Aspects of Distributed Preservation

Sociologically, too, the web is currently seen as a basically superficial medium, through which one "browses" or "surfs," in which no conventions have yet been established for typical library-related information, such as provenance (POCKELEY, p. 15). Furthermore, LYNCH (1999) points out that published web material lacks a common structure (such as that of the book, with its table of contents, bibliography, etc.), meaning that pages have an unsettled, informal look, and their content is harder to quickly grasp; it becomes difficult to formulate archival policies for material which seems to exist in such an anarchic context. LYMAN & KAHLE (1998) put it succinctly: "Like oral culture, digital information has been allowed to become a medium for the present, neither a record of the past nor a message to the future." (p. 2)

Other preservation problems arise from the impermanence of material in a distributed system—when objects have no real substance, and exist only to be distributed (like e-mail), who is responsible for their preservation? The importance of informal communication in the study of history and culture is witnessed by the large number of books of correspondence, which have been published and found useful through the centuries; LUKESH warns that our current neglect of electronic correspondence will result in significant historical losses for future scholars. Likewise, MARCUM asserts that, in some disciplines, the traditional print records do not adequately capture the intellectual development, which is preserved instead in "on-line databases, on-line exchanges of preprints, listservs, and the like" (p. 357).

Another obstacle in the path of developing distributed digital preservation practice is the current immaturity of digital publishing (in this sense, making material available on the Internet) itself, which is still—and perhaps will always be, to a certain extent—a basically informal, unstructured medium. When publishing has become easy enough that ten-year-old children regularly create publicly available web pages, what should the policy for legal deposit be? One problem with net publishing, from this perspective, is that it removes the screening effect that necessarily takes place in a print environment, where an intermediary publishing organization must decide what texts are worthy of becoming books. (NLC)

Technologies and Approaches

Traditional preservation approaches mainly deal with problems of material preservation—how can paper, tapes, or photographs be preserved so that the content they carry will be available to future scholars? In a digital context, content can now be separated from a particular physical manifestation; indeed, long-term preservation often necessitates the wholesale migration of data from one form (or format) to another, as software formats grow obsolete and are replaced by more advanced models. In the context of long-term archival, distribution of information occurs not only geographically, across networked systems, but also temporally—across time spans in which technologies and the formats they support vary much more than they do across distances in merely three dimensions.

One approach, then, is to preserve digital information in a simplest possible format, to make it as software-independent as possible (HEDSTROM). Another format-oriented approach is that of LYNCH (1999), who proposes canonical formats for preservation, which are capable of maintaining object authenticity across migrations, and create standard definitions as to what is "important" in a document form. As usual, the difficulties with such an approach are not mainly technological, but organizational.

ROTHENBERG takes a different tack; instead of changing data so that it can be read by future machines (through standard formats or migration), he would provide for the specification of hardware emulators, so that current (and past) programs can be run in the future on virtual versions of obsolete machines. This approach seems to transfer something of traditional preservation practices to the digital realm; content becomes bound not to a physical object, but to a specific bitstream, the interpretation of which is now always assured, by encapsulating information about display mechanisms along with the data itself. Intellectual property issues relating to the copying of operating system and software technology must be solved before this approach can be viable on a wide basis.

CONCLUSION

This chapter has presented some of the major issues affecting the development of digital libraries at the beginning of the 21st century. There is as yet little consensus about the best way to organize a distributed digital library — indeed, there is not even agreement on a single definition. Networking and digitization are facilitating the creation of an entirely new paradigm of information management, and standards and practices are still in a state of disorderly development. The course of the future, it seems, could be set by practitioners and organizers — those who not only devise intelligent, workable standards, but who can put together working implementations, and muster the cooperation between institutions needed to make any solution a true "standard."

Future Work

It is difficult to separate out "future work" from current projects in the field of digital libraries and distributed information management — the vast majority of such projects in existence are themselves essentially works in progress; if they are in operation today, it is often only as one stage of a trajectory aimed at some future functionality which is even faster and more powerful.

One area of research in which no general, scalable solutions have been found is indexing and search and retrieval techniques for non-textual data. Among many others, the Computer Vision group at UC Berkeley is currently working on a system for content-based image retrieval; "Blobworld," a system that separates images into coherent regions, is described in detail in UNIVERSITY OF CALIFORNIA AT BERKELEY DIGITAL LIBRARY PROJECT. The SIMPLIcity system (Semantics-sensitive Integrated Matching for Picture LIbraries), developed at Stanford, is another project to automatically semantically classify images, based on recognition and categorization of regions in the image itself. For sound files, MELDEX, an audio-based system for indexing and retrieval of melodies, has been created for the New Zealand digital library. This system will retrieve melodies from a database of almost 10,000 folk songs based on a few notes sung into a microphone. A demonstration page on the web, at NEW ZEALAND DIGITAL LIBRARY, allows browsers to try a prototype out for themselves.

Another area in which great changes are foreseeable in the future is the field of scholarly publishing. This is not so much a technological problem to be resolved as a (coming) paradigm shift whose implications will have to be absorbed by the academic community: as the boundaries between author and publisher blur or dissolve, and centrally controlled systems give way to distributed, non-monolithic models, peer-review type systems to protect quality will need to evolve, and the concept and usages of academic publishing will have to adapt. (See especially HARNAD for an impassioned defense of this new publishing model). The Open Archives Initiative, described above, is one foray into this new mode of information dissemination; the growing number of electronic journals (such as several quoted in this bibliography) also bear witness to the developing change in paradigm.

Technologically, a general trend seems to be towards more specialization and personalization of the searching process; indeed, one current project aims specifically to create Personalized Information Environments (PIEs), which brings the idea of the collection down to the level of the user, who is able with this concept to create his or her own personal digital library out of a vast distributed system (FRENCH & VILES). But as the developers of the PIE themselves suggest, it is Vannevar Bush's idea of the Memex, from 1945, which still drives much conceptual work in distributed information management; in a sense, much of what is being done now, at the beginning of the 21st century, can be seen as an attempt to fulfill Bush's seminal vision of fifty years ago.[3]

[3] Acknowledgements: We gratefully acknowledge the assistance of the staff in the Emerging Technologies Group at the National Center for Supercomputing Applications at the University of Illinois.

Bibliography

ACKERMAN, M. 1994. Providing Social Interaction in the Digital Library. In: Schnase, John L.; Leggett, John J.; Furuta, Richard K.; Metcalfe, Ted, eds. Proceedings of Digital Libraries '94: The First Annual Conference on the Theory and Practice of Digital Libraries; 1994 June 19-21; College Station, TX. Available WWW: http://www.ics.uci.edu/~ackerman/pub/94b11/dl94.final.html

ACKERMAN, M.; FIELDING, R. 1995. Collection Maintenance in the Digital Library. In: Proceedings of Digital Libraries '95: The Second Annual Conference on the Theory and Practice of Digital Libraries; 1995 June 11-13; Austin, TX. Available WWW: http://www.csdl.tamu.edu/DL95/papers/ackerman/ackerman.html

ANDERSON, MARTHA. 1999. A Tool for Building Digital Libraries. D-Lib Magazine. 1999 February; 5(2). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/february99/02journalreview.html

ANTELMAN, KRISTIN; LANGENBERG, DAVID. 1993. Collection Development in the Electronic Library. In: Proceedings of the 21st ACM SIGUCCS conference on User Services; 1993 November 7-10; San Diego, CA.

New York, NY: Association for Computing Machinery (ACM), 1993. 50-56. Available WWW: http://www.acm.org/pubs/articles/proceedings/userservices/263814/p50-antelman/p50-antelman.pdf

ARMS, WILLIAM Y. 2000. Digital Libraries. Cambridge, MA: The MIT Press; 2000. 287. ISBN: 0-262-01880-8.

ARMS, WILLIAM Y.; BLANCHI, CHRISTOPHE; OVERLY, EDWARD A. 1997. An Architecture for Information in Digital Libraries. D-Lib Magazine. 1997 February; 3(2). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/february97/cnri/02arms1.html

ARMS, WILLIAM; DAIGLE, LESLIE; DANIEL, RON; LALIBERTE, DAN; MEALLING, MICHAEL; MOORE, KEITH; WEIBEL, STUART. 1996. Uniform Resource Names: A Progress Report. D-Lib Magazine. 1996 February; 2(2). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/february96/02arms.html

ARNOLD, KENNETH. 1994. The Electronic Librarian is a Verb / The Electronic Library is not a Sentence. A Lecture Delivered at the New York Public Library. The Gilbert A. Cam Memorial Lecture Series. 1994 October 14. Available WWW: http://www.press.umich.edu/jep/works/arnold.eleclib.html

ATKINSON, ROSS. 1996. Library Functions, Scholarly Communication, and the Foundation of the Digital Library: Laying Claim to the Control Zone. Library Quarterly. 1996 July; 66(3): 239-265. ISSN: 0024-2519.

BACA, MURTHA, ed. 2000. Introduction to Metadata: Pathways to Digital Information. U.S.A.: Getty Information Institute; 1998 December 31. 47. ISBN: 0-89236-533-1. Available WWW: http://www.getty.edu/gri/standard/intrometadata/index.htm

BEARMAN, DAVID. 1999. Reality and Chimeras in the Preservation of Electronic Records. D-Lib Magazine. 1999 April; 5(4). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/april99/bearman/04bearman.html

BOORSTIN, DANIEL J. 1980. Gresham's Law: Knowledge or Information? Remarks at the White House Conference on Library and Information Services. Washington, D.C., Nov. 19, 1979. Washington, D.C.: Library of Congress; 1980. 6. (Center for the Book: Viewpoint series; no. 3) ISBN: 0-8444-0346-6.

BOUSKILA, FABIEN D.; POTTENGER, WILLIAM M. 2000. The Role of Semantic Locality in Hierarchical Distributed Dynamic Indexing. In: Proceedings of the 2000 International Conference on Artificial Intelligence (IC-AI 2000); 2000 June; Las Vegas, NV. Available WWW: http://www.ncsa.uiuc.edu/ET/HDDI/

BOUTHORS, VINCENT; DEDIEU, OLIVIER. 1999. Pharos, a Collaborative Infrastructure for Web Knowledge Sharing. In: Proceedings of the Third

European Conference on Research and Advanced Technology for Digital Libraries (ECDL) '99; 1999 September 22-24; Paris, France. New York, NY: Springer-Verlag, 1999. 215-233. (Lecture Notes in Computer Science, 1696) ISBN: 3-540-66558-7. Available WWW: http://link.springer.de/link/service/series/0558/tocs/t1696.htm

BUSH, VANNEVAR. 1945. As We May Think Atlantic Monthly. 1945 July; 176(1): 101-108. Available WWW: http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm

CORNELL DIGITAL LIBRARIES RESEARCH GROUP. 1999. Flexible and Extensible Digital Object and Repository Architecture (FEDORA). Available WWW: http://www.cs.cornell.edu/edlrg/FEDORA.html

CORNELL DIGITAL LIBRARIES RESEARCH GROUP. 2000. Dienst: Overview and Introduction. Available WWW: http://www.cs.cornell.edu/cdlrg/dienst/DienstOverview.htm

CORPORATION FOR NATIONAL RESEARCH INITIATIVES. Web site: http://www.handle.net/.

COX, RICHARD. 2000. Searching for Authority: Archivists and Electronic Records in the New World at the Fin-de-Siécle. First Monday. 2000 January; 5(1). Available WWW: www.firstmonday.dk/issues/issue5_1/cox/index.html

COFFMAN, STEVE. 1999. Building Earth's Largest Library: Driving into the Future. Searcher. 1999 March; 7(3). ISSN: 1070-4795.

COYLE, KAREN. 2000. The Virtual Union Catalog. D-Lib Magazine. 2000 March; 6(3). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/march00/coyle/03coyle.html

CRAWFORD, WALT. 1998. Paper Persists: Why Physical Library Collections Still Matter. ONLINE. 1998 January. Available WWW: http://www.onlineinc.com/onlinemag/OL1998/crawford1.html

CRAWFORD, WALT. 1999. Being Analog: Creating Tomorrow's Libraries. Chicago, IL: American Library Association; 1999. 245. ISBN: 0-8389-0754-7.

CRESPO, ARTURO; GARCIA-MOLINA, HECTOR. 1998. Archival Storage for Digital Libraries. In: Witten, Ian; Akscyn, Rob; Shipman, Frank M. III, ed. Proceedings of the Third ACM Conference on Digital Libraries; 1998 June 23-26; Pittsburgh, PA. New York, NY: The Association for Computing Machinery; 1998. 134-143. ISBN: 0-89791-965-3.

DANIEL, RON, JR.; LAGOZE, CARL; PAYETTE, Sandra D. 1998. A Metadata Architecture for Digital Libraries. In: Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries (IEEE ADL '98); 1998 April 22-24; Santa Barbara, CA. Available WWW: http://www.cs.cornell.edu/lagoze/papers/ADL98/dar-adl.html

DAVIS, JAMES R.; LAGOZE, CARL. 2000. NCSTRL: Design and Deployment of a Globally Distributed Digital Library. Journal of the American Society for Information Science. 2000; 51(3): 273-280. ISSN: 1097-4571 (Online). Available WWW: http://www3.interscience.wiley.com/cgi-bin/fulltext/69501282/FTSTART

DAY, MICHAEL. 1999. Metadata for Digital Preservation: an Update. Ariadne. 1999 December; (22). ISSN: 1361-3200. Available WWW: www.ariadne.ac.uk/issue22/metadata/

DUSHAY, NAOMI; PAYETTE, SANDRA. 2000. Flexible and Extensible Digital Object and Repository Architecture (FEDORA). Available WWW: http://www.cs.cornell.edu/cdlrg/FEDORA.html.

EDMUNDS, JEFF; BRISSON, ROGER. 2000. Cataloging in CORC: A Work in Progress. Pre-Print article to be published in the Journal of Internet Cataloging. Available from author on WWW: http://www.personal.psu.edu/faculty/r/o/rob1/publications/publications.html

FOX, EDWARD A. 2000. Digital Libraries and the Open Archives Initiative. Powerpoint slides presented at: Louisiana State University; 2000 June 30; Baton Rouge, LA. Available WWW: http://www.ndltd.org/talks.

FOX, EDWARD; MARCHIONINI, GARY. 1998. Toward a Worldwide Digital Library. Guest Editors' Introduction to Special Section on Digital Libraries:

Global Scope, Unlimited Access. Communications of the ACM. 1998 April; 41(4): 28-32. Available WWW: http://purl.lib.vt.edu/dlib/pubs/CACM199804

FRENCH, JAMES C.; VILES, CHARLES L. 1999. Personalized Information Environments: An Architecture for Customizable Access to Distributed Digital Libraries. D-Lib Magazine. 1999 June; 5(6). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/june99/french/06french.html

GOOGLE. Web site: http://www.google.com/.

GORMAN, PAUL; DELCAMBRE, LOIS; MAIER, DAVID; REHFUSS, STEVE; DENG, LONGXING. 1998. Tracking Footprints through an Information Space: Leveraging the Document Selections of Expert Problem Solvers. 1998; Portland, OR. (Project Overview). Available WWW: http://www.cse.ogi.edu/dot/research/footprints/overview.html

GRIFFITHS, JOSÉ-MARIE. 1998. The New Information Professional. Bulletin of the American Society for Information Science. 1998 February/March; 24(3). ISSN: 0095-4403. Available WWW: http://www.asis.org/Bulletin/Feb98/griffiths.html.

HAEKLI, ESKO; HAKALA, JUHA. 1998. URN Implementation in National Libraries: Working Group Report from CENL (Conference of European National Librarians). 1998 October 1; Prague, Czech Republic. Available WWW: http://linnea.helsinki.fi/urn/urnimp.html

HARNAD, STEVAN. 1999. Free at Last: The Future of Peer-Reviewed Journals. D-Lib Magazine. 1999 December; 5(12). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/december99/12harnad.html

HARTER, STEPHEN P. 1997. Scholarly Communication and the Digital Library: Problems and Issues. Journal of Digital Information. 1997 April; 1(1). ISSN: 1368-7506. Available WWW: http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter

HAYNES, DAVID; STREATFIELD, DAVID. 1998. A National Co-ordinating Body for Digital Archiving?. Ariadne. 1998 May; (15). ISSN: 1361-3200. Available WWW: www.ariadne.ac.uk/issue15/digital/

HEDMAN, ANDERS. 1999. Creating Digital Libraries Together—Collaboration, Multimodality, and Plurality. In: Manaris, Bill, Ed. Proceedings of the 4th annual Conference on Innovation and Technology in Computer Science Education(ITiCSE); 1999 June 27-30; Krakow, Poland. ACM SIGCSE/SIGCUE, 1999. 147-150. ISBN: 1-58113-087-2. Available WWW: http://www.acm.org/pubs/contents/proceedings/cse/305786/

HEDSTROM, MARGARET. 1999. Digital preservation: A Time Bomb for Digital Libraries. Paper presented at: NSF Workshop on Data Archival and Information Preservation; 1999 March 26-27; Washington, DC. Available WWW: http://www.uky.edu/~kiernan/DL/hedstrom.html

HEDSTROM, MARGARET; WALLACE, DAVID. 1999. And the Last Shall be First: Recordkeeping Policies and the NII. Journal of the American Society for Information Science. 1999; 50(4): 331-339. ISSN: 0002-8231. Available WWW: http://www3.interscience.wiley.com/

HEERY, RACHEL. 1998. What is… RDF? Ariadne. 1998 March;(14). ISSN: 1361-3200. Available WWW: http://www.ariadne.ac.uk/issue14/what-is/

HENRY, LINDA J. 1998. Schellenberg in Cyberspace. The American Archivist. 1998 Fall; 61: 309-327. ISSN: 0360-9081.

HICKEY, THOMAS B. 1998. CORC—Cooperative Online Resource Catalog. Annual Review of OCLC Research. 1998. ISSN: 0894-198X. Available WWW: http://www.oclc.org/oclc/research/publications/review98/hickey/corc.htm

HODGE, GAIL M. 2000. Best Practices for Digital Archiving. D-Lib Magazine. 2000 January; 6(1). ISSN: 1082-9873. Available WWW: www.dlib.org/dlib/january00/01hodge.html

KAHN, ROBERT; WILENSKY, ROBERT. 1995. A Framework for Distributed Digital Object Services. 1995 May. Available WWW: http://www.cnri.reston.va.us/home/cstr/arch/k-w.html

KANTOR, P. B.; BOROS, E.; MELAMED, B.; MEJIKOV, VLADIMIR. 1999. The Information Quest: A Dynamic Model of User's Information Needs. Available WWW: http://aplab.rutgers.edu/ant/

KANTOR, PAUL. 1999. How the AntWorld Works. 1999 September. Available WWW: http://aplab.rutgers.edu/ant/.

LAGOZE, CARL. 2000. The Cornell Digital Library Research Group: Architectures and Policies for Distributed Research Libraries. Invited Paper for Digital Libraries Workshop 17, Tsukuba, Japan. 2000 February. Available WWW: http://www2.cs.cornell.edu/lagoze/papers/DLW17/cdlrg.htm

LAGOZE, CARL; FIELDING, DAVID. 1998. Defining Collections in Distributed Digital Libraries. D-Lib Magazine. 1998 November; 4(11). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/november98/lagoze/11lagoze.html

LAGOZE, CARL; FIELDING, DAVID; PAYETTE, SANDRA. 1998. Making Global Digital Libraries Work: Collection Services, Connectivity Regions, and Collection Views. In: Witten, Ian; Akscyn, Rob; Shipman, Frank M. III, Ed. Proceedings of the Third ACM Conference on Digital Libraries; 1998 June 23-26; Pittsburgh, PA. New York, NY: The Association for Computing Machinery; 1998. 134-143. ISBN: 0-89791-965-3. Available WWW: http://www2.cs.cornell.edu/lagoze/papers/lagoze.pdf

LAGOZE, CARL; FUHR, NORBERT. 1998. Resource Discovery in a Globally-Distributed Digital Library Working Group Report: Digital Library Collaborative Working Groups. 1998 February 26-27; Washington, D.C. Available WWW: http://www.iei.pi.cnr.it/DELOS/NSF/resourcediscovery.htm

LAGOZE, CARL; LYNCH, CLIFFORD; DANIEL, RON JR. 1996. The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata. 1996 June. (Cornell University Computer Science Technical Report TR96-1593). Available WWW: http://cs-tr.cs.cornell.edu:80/Dienst/UI/2.0/Describe/ncstrl.cornell/TR96-1593

LASSILA, ORA. 1997. Introduction to RDF Metadata. 1997 November 13. (W3C NOTE 1997-11-13). Available WWW: http://www.w3.org/TR/NOTE-rdf-simple-intro

LEINER, BARRY M. 1998. The NCSTRL Approach to Open Architecture for the Confederated Digital Library. D-Lib Magazine. 1998 December; 4(12). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/december98/leiner/12leiner.html

LESK, MICHAEL. 1997. Practical Digital Libraries: Books, Bytes, and Bucks. San Francisco, CA: Morgan Kaufmann Publishers, Inc.; 1997. 297. ISBN: 1-55860-459-6.

LEVY, DAVID M.; MARSHALL, CATHERINE C. 1994. Washington's White Horse? A Look at Assumptions Underlying Digital Libraries. In: Schnase, John L.; Leggett, John J.; Furuta, Richard K.; Metcalfe, Ted, eds. Proceedings of Digital Libraries '94: The First Annual Conference on the Theory and Practice of Digital Libraries; 1994 June 19-21; College Station, TX. Available WWW: www.csdl.tamu.edu/DL94/paper/levy.html

LEVY, DAVID M.; MARSHALL, CATHERINE C. 1995. Going Digital: A Look at Assumptions Underlying Digital Libraries. Communications of the ACM. 1995 April; 38(4): 77-84.

LEVY, DAVID M. 2000. Digital Libraries and the Problem of Purpose. D-Lib Magazine. 2000 January; 6(1). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/january00/01levy.html

LUKESH, SUSAN. 1999. E-mail and Potential loss to future archives. First Monday. 1999 September; 4(9). Available WWW: www.firstmonday.dk/issues/issue4_9/lukesh/index.html

LYMAN, PETER; KAHLE, BREWSTER. 1998. Archiving Digital Cultural Artifacts: Organizing an Agenda for Action. D-Lib Magazine. 1998 July/August. ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/july98/07lyman.html

LYNCH, CLIFFORD. 1997. Identifiers and their Role in Networked Information Applications. A.R.L.: A Bimonthly Newsletter of Research Library Issues and Actions. 1997 October; (194). ISSN: 0066-9652. Available WWW: http://www.arl.org/newsltr/194/identifier.html

LYNCH, CLIFFORD. 1999. Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information. D-Lib Magazine. 1999

September; 5(9). ISSN: 1082-9873. Available WWW:
www.dlib.org/dlib/september99/09lynch.html

MALY, KURT; NELSON, MICHAEL L.; ZUBAIR, MOHAMMAD. 1999. Smart Objects, Dumb Archives: A User-Centric, Layered Digital Library Framework. D-Lib Magazine. 1999 March; 5(3). ISSN: 1082-9873. http://www.dlib.org/dlib/march99/maly/03maly.html

MARCUM, DEANNA B. 1997. A Moral and Legal Obligation: Preservation in the Digital Age. International Information and Library Review. 1997; 29(3-4): 357-65. ISSN: 1057-2317.

MARSHALL, CATHERINE C. 1997. Annotation: From Paper Books to the Digital Library. In: Proceedings of the 2nd ACM International Conference on Digital Libraries; 1997 July 23-26; Philadelphia, PA. New York, NY: The Association for Computing Machinery; 1997. 131-140. ISBN: 0-89791-868-1. Available WWW:
http://www.acm.org/pubs/articles/proceedings/dl/263690/p131-marshall/p131-marshall.pdf

MCNAB, RODGER J.; SMITH, LLOYD A.; BAINBRIDGE, DAVID; WITTEN, IAN H. 1997. The New Zealand Digital Library MELody inDEX. D-Lib Magazine. 1997 May. ISSN: 1082-9873. Available WWW:
http://www.dlib.org/dlib/may97/meldex/05witten.html

MIKSA, FRANCIS L.; DOTY, PHILIP. 1994. Intellectual Realities and the Digital Library. In: Schnase, John L.; Leggett, John J.; Furuta, Richard K.; Metcalfe, Ted, eds. Proceedings of Digital Libraries '94: The First Annual Conference on the Theory and Practice of Digital Libraries; 1994 June 19-21; College Station, TX. Available WWW: http://www.csdl.tamu.edu/DL94/paper/miksa.html

MILLER, ERIC. 1998. An Introduction to the Resource Description Framework. D-Lib Magazine. 1998 May. ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/may98/miller/05miller.html

MOURA, ANA MARIA DE CARVALHO; CAMPOS, MARIA LUIZA MACHADO; BARRETO, CASSIA MARIA. 1999. A Metadata Architecture to Represent Electronic Documents on the Web. In: Proceedings of the IEEE Meta-Data Conference; 1999 April 6-7; Bethesda, MD. Available WWW: http://www.computer.org/proceedings/meta/1999/papers/15/amoura.html

MULTIVALENT DOCUMENT HOME PAGE. Web site: http://http.cs.berkeley.edu/~wilensky/MVD.html.

MURAMATSU, BRANDON; AGOGINO, ALICE M. 1999. The National Engineering Education Delivery System: A Digital Library for Engineering Education. D-Lib Magazine. 1999 April; 5(4). ISSN 1082-9873. Available WWW: http://www.dlib.org/dlib/april99/muramatsu/04muramatsu.html

MURAMATSU, BRANDON. 2000. A Digital Learning Space for Science, Mathematics, Engineering and Technology Education. Paper presented at: Mathematics/Science Education and Technology 2000 (M/SET); 2000 February 5; San Diego, CA. Available WWW: http://www.smete.org/smete/info/papers/MSET-0200/muramatsu-mset2000.pdf

NATIONAL ENGINEERING EDUCATION DELIVERY SYSTEM. Web site: http://www.needs.org.

NATIONAL LIBRARY OF CANADA. ELECTRONIC COLLECTIONS COORDINATING GROUP. 1999. Networked Electronic Publications Policy and Guidelines. 1999 April. Available WWW: http://www.nlc-bnc.ca/pubs/irm/eneppg.htm

NELSON, MICHAEL L.; MALY, KURT; CROOM, DELWIN R. JR.; ROBBINS, STEVEN W. 1999. Metadata and Buckets in the Smart Object Dumb Archive (SODA) Model. In: Proceedings of the IEEE Meta-Data Conference; 1999 April 6-7; Bethesda, MD. Available WWW: http://www.computer.org/proceedings/meta/1999/papers/53/mnelson.html

NELSON, MICHAEL L.; MALY, KURT; SHEN, STEWART N.T.; ZUBAIR, MOHAMMAD. 1998. NCSTRL+: Adding Multi-Discipline and Multi-Genre Support to the Dienst Protocol Using Clusters and Buckets. In: Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries (IEEE ADL '98); 1998 April 22-24; Santa Barbara, CA. 128-136. Available

WWW: http://techreports.larc.nasa.gov/ltrs/PDF/1998/mtg/NASA-98-ieeedl-mln.pdf

NEW ZEALAND DIGITAL LIBRARY. 2000. Meldex demonstration home page. Available WWW: http://www.cs.waikato.ac.nz/~nzdl/meldex/demo.html.

ODLYZKO, ANDREW. 1997. Silicon Dreams and Silicon Bricks: The Continuing Evolution of Libraries. Library Trends. 1997 Summer; 46(1): 152-167. Available WWW: http://www.research.att.com/~amo/doc/complete.html

ONLINE COMPUTER LIBRARY CENTER. Web site: http://www.oclc.org/oclc/corc/index.htm.

OPEN ARCHIVES INITIATIVE. 1999. Proceedings of the First Meeting of the Open Archives Initiative. 1999 October 21-22; Santa Fe, NM. Available WWW: www.openarchives.org/ups1-press.htm

PASKIN, NORMAN. 1999. DOI: Current Status and Outlook. D-Lib Magazine. 1999 May; 5(5). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/may99/05paskin.html

PASKIN, NORMAN. 1999. The Digital Object Identifier. Summary of seminar given as part of the Online Computer Library Center's Distinguished Seminar Series; 1999 February 23. Available WWW: http://www.oclc.org/oclc/research/publications/review99/paskin/doi.htm

PAYETTE, SANDRA; BLANCHI, CHRISTOPHE; LAGOZE, CARL; OVERLY, EDWARD A. 1999. Interoperability for Digital Objects and Repositories: The Cornell/CNRI Experiments. D-Lib Magazine. 1999 May ; 5(5). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/may99/payette/05payette.html

PAYETTE, SANDRA; LAGOZE, CARL. 1998. Flexible and Extensible Digital Object and Repository Architecture (FEDORA). . In: Proceedings of the Second European Conference on Research and Advanced Technology for Digital Libraries (ECDL) '98; 1998 September 19-23; Heraklion, Greece. New York, NY: Springer-Verlag, 1998. pp. (Lecture Notes in Computer Science, 1513) ISBN: 3-540-65101-2. Available WWW: http://www2.cs.cornell.edu/payette/papers/ECDL98/FEDORA.html

PERSISTENT UNIFORM RESOURCE LOCATOR TEAM. 2000. Persistent URL home page. Available WWW: http://purl.oclc.org/.

PHELPS, THOMAS A.; WILENSKY, ROBERT. 2000. Robust Hyperlinks Cost Just Five Words Each. 2000 January 10. (UCB Computer Science Technical Report UCB//CSD-00-1091). Available WWW: http://www.cs.berkeley.edu/~wilensky/robust-hyperlinks.html

POCKELEY, SIMON. 1995. Killing the Duck to Keep the Quack: Using the World Wide Web for Digital Preservation. Available WWW: www.cinemedia.net/FOD/FOD0055.html

ROBERTSON, SCOTT; JITAN, SHERIF; REESE, KATHY. 1997. Web-Based Collaborative Library Research. In: Proceedings of the 2nd ACM International Conference on Digital Libraries; 1997 July 23-26; Philadelphia, PA. New York, NY: The Association for Computing Machinery; 1997. 152-160. ISBN: 0-89791-868-1. Available WWW: http://www.acm.org/pubs/citations/proceedings/dl/263690/p152-robertson/

ROTHENBURG, JEFF. 1999. Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Council on Library and Information Resources; 1999 January. 35. ISBN: 1-887334-63-7. Available WWW: http://www.clir.org/pubs/reports/rothenberg/contents.html

SANTA FE CONVENTION FOR THE OPEN ARCHIVES INITIATIVE. 2000. Summary of Results from Meeting of Open Archives initiative; 1999 October 21-22; Santa Fe, NM. Available WWW: http://www.openarchives.org/sfc/sfc_entry.htm

SCHATZ, BRUCE; CHEN, HSINCHUN. 1999. Digital Libraries: Technological Advances and Social Impacts. Computer. 1999 February; 32(2): 45-50. ISSN: 0018-9162.

SCHATZ, BRUCE; MISCHO, WILLIAM; COLE, TIMOTHY; BISHOP, ANN; HARUM, SUSAN; JOHNSON, ERIC; NEUMANN, LAURA; CHEN, HSINCHUN; NG, DORBIN. 1999. Federated Search of Scientific Literature. Computer. 1999 February; 32(2): 51-59. ISSN: 0018-9162.

SCHWARZWALDER, ROBERT. 1999. Librarians as Knowledge Management Agents. Econtent. 1999 August/September; 22(4). ISSN: 0162-4105.

SHAFER, KEITH. 1998. Mantis Project: A Toolkit for Cataloging. In: Annual Review of OCLC Research 1998. ISSN: 0894-198X. Available WWW: http://www.oclc.org/oclc/research/publications/review98/shafer/mantis.htm

SHAFER, KEITH. 1999. ARMs, OCLC Internet Services, and PURLs. In: Annual Review of OCLC Research 1999. ISSN: 0894-198X. Available WWW: http://www.oclc.org/oclc/research/publications/review99/shafer/arms.htm

SHEPARD, THOM. 1999. Notes Toward a Universal Preservation Format. Paper presented at: NSF Workshop on Data Archival and Information Preservation; 1999 March 26-27; Washington, DC. Available WWW: http://www.cecs.missouri.edu/NSFWorkshop/upfpp.html

SHIPMAN, FRANK M.; FURUTA, RICHARD; BRENNER, DONALD; CHUNG, CHUNG-CHI; HSIEH, HAO-WEI. 2000. Guided Paths through Web-Based Collections: Design, Experiences, and Adaptations. Journal of the American Society for Information Science. 2000; 51(3): 260-272. ISSN: 0002-8231.

SIMMONS, L. M. 1997. Collaboration Dreams: Guest Editorial. D-Lib Magazine. 1997 March. ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/march97/03editorial.html

SMITH, ABBY. 1999. Preservation in the Digital Age: What is to be Done? American Libraries. 1999 March; 30(3): 36-9. ISSN: 0002-9769.

SOLLINS, K.; MASINTER, L. 1994. Functional Requirements for Uniform Resource Names. 1994 December. (Network Working Group RFC 1737). Available WWW: http://www.w3.org/Addressing/rfc1737.txt

SUN, SAM X.; LANNOM, LARRY. 2000. Handle System Overview. (Network Working Group, INTERNET-DRAFT.) Available WWW: http://www.ietf.org/internet-drafts/draft-sun-handle-system-04.txt

THIBODEAU, KENNETH. 1999. Resolving the Inherent Tensions in Digital Preservation. Paper presented at: NSF Workshop on Data Archival and Information Preservation; 1999 March 26-27; Washington, DC. Available WWW: www.cecs.missouri.edu/NSFWorkshop/thibpp.html

THOMAS, SPENCER W.; ALEXANDER, KEN; GUTHRIE, KEVIN. 1999. Technology Choices for the JSTOR Online Archive. Computer. 1999 February; 32(2): 60-65. ISSN: 0018-9162.

UNIVERSITY OF CALIFORNIA AT BERKELEY DIGITAL LIBRARY PROJECT. Web site: http://elib.cs.berkeley.edu/vision.html.

VAN DE SOMPEL, HERBERT; KRICHEL, THOMAS; NELSON, MICHAEL L.; HOCHSTENBACH, PATRICK; LYAPUNOV, VICTOR M.; MALY, KURT; ZUBAIR, MOHAMMED; KHOLIEF, MOHAMED; LIU, XIAOMING;

O'CONNELL, HEATH. 2000. The UPS Prototype. D-Lib Magazine. 2000 February; 6(2). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html

VAN DE SOMPEL, HERBERT; LAGOZE, CARL. The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine. 2000 February; 6(2). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html

WALLACE, JONATHAN. 1996. The Internet is a Library. Sex, Laws, and Cyberspace Bulletin 1. Available WWW: http://www.spectacle.org/cda/lib.html

WANG, JAMES ZE; LI, JIA; CHAN, DESMOND; WIEDERHOLD, GIO. 1999. Semantics-sensitive Retrieval for Digital Picture Libraries. D-Lib Magazine. 1999 November; 5(11). ISSN: 1082-9873. Available WWW: http://www.dlib.org/dlib/november99/wang/11wang.html

WARD, SANDRA. 1999. Information Professionals for the Next Millennium. Journal of Information Science. 1998; 25(4): 239-247. ISSN: 0165-5515.

YAHOO. Web site: http://www.yahoo.com/.

WEIBEL, STUART; JUL, ERIK; SHAFER, KEITH. 2000. PURLs: Persistent Uniform Resource Locators. Available WWW: http://purl.oclc.org/OCLC/PURL/SUMMARY

WITTEN, IAN H.; MCNAB, RODGER J.; JONES, STEVE; APPERLEY, MARK; BAINBRIDGE, DAVID; CUNNINGHAM, SALLY JO. 1999. Managing Complexity in a Distributed Digital Library. Computer. 1999 February; 32(2): 74-79. ISSN: 0018-9162.

WILENSKY, R.; PHELPS, T. 1998. Multivalent Documents: From Presentation to Collaboration. 1998 January. (Powerpoint presentation). Available WWW: http://www.dli2.nsf.gov/dlione/berkeley/ucb-mvd/sld001.htm

WILENSKY, ROBERT. 2000. Digital Library Resources as a Basis for Collaborative Work. Journal of the American Society for Information Science. 2000; 51(3): 228-245. ISSN: 0002-8231.

YOUNG, PETER R. 1996. Librarianship: A Changing Profession. Daedalus. 1996 Fall; 125(4): 103-125. ISSN: 0011-5266.