

much coverage of the literature on intellectual property and copyright issues; an excellent source for these is [U.S. Congress Office of Technology Assessment, 1986]

The opinions expressed here are mine and do not necessarily reflect the views of any other organization or individual. A number of products and projects are mentioned here as examples; these mentions should not be interpreted as endorsements of products. Finally, a large number of trademarks and service marks are mentioned in the text. In particular: UNIX is a registered trademark of Bell Labs; Macintosh is registered to Apple Computers, as is Quicktime. MELVYL is registered to the Regents of the University of California. PostScript, SuperATM and Acrobat are trademarks of Adobe. I apologize in advance for those that I have failed to mention here.

## 2. **The Networked Information Context**

As use of the Internet becomes more widespread, it is becoming clear that access to information resources through the network will become one of the central applications of the network. The term "access," here and *throughout this paper*, is *used in the broadest sense: not simply electronic connectivity to information resource providers through the network, but the ability for people to successfully locate, retrieve and use the information contained within various computer systems*. For a large segment of the Internet user community, such access to information and network-based communications tools such as electronic mail will increasingly be the applications which initially attract users to the Internet and subsequently are most heavily used. Further, we are already seeing both the development of a series of information discovery and navigation tools aimed at the end user and the evolution of a set of services that combine the access and communications capabilities offered by the network into new hybrid information services that are more interactive and immediate than traditional print or broadcast media. Libraries, publishers, and government organizations that, create, collect and provide access to information, are all striving to define and understand their possible new roles in this developing networked information environment.

Discussions about the development of the Internet and the evolution of the National Research and Education Network (NREN) program are increasingly taking a broader view that emphasizes not only the continued expansion and upgrading of the technical communications infrastructure but also the need for policies and planning to foster the development of networked information resources and the tools and educational programs needed to allow network users to create, navigate, and utilize such resources. Another component of this shift of focus is the recognition of the need to transform existing printed and electronic information resources into network-accessible forms. As we look beyond the NREN program, which is targeted to serve the research, education, library and government communities (broadly defined) towards the discussions about a full National Information Infrastructure (NII) it is clear that electronic information content on future networks will play an increasingly dominant and driving role in network evolution.

This shifting focus is evident, for example, in a wide range of bills before the 1993 Congress, including S4, The National Competitiveness Act of 1993 (Hollings), S626, The Electronic Library Act of 1993 (Kerrey), HR1757, The High Performance

Computing and High Speed Networking Applications Act of 1993 (Boucher) and HR1328, The Government Printing Office Electronic Information Access Enhancement Act of 1993,<sup>1</sup> to name some of the most major. Various federal agency based programs (for example, at NASA, the National Agriculture Library, the Library of Congress, and the National Library of Medicine) are also underway to foster the availability of networked information resources.<sup>2</sup> The recent revision of the Office of Management and Budget Circular A-130 and its associated guidance to government organizations also speaks to the need to make information available in electronic form. Paralleling these activities at the federal level is a growing interest on the part of some state governments in the potential of the network to provide improved access to state, regional and local information (for example, in Texas [Stout, 1992]. Colorado, Utah,<sup>3</sup> California and North Carolina). State libraries are using the Internet as a focus for statewide multitype library planning in several states, including Colorado [Mitchell & Saunders, 1991] and Maryland.<sup>5</sup>

Broader based initiatives that span the government, research and education, and commercial sectors recognize networked access to information resources as a key element. For example, the National Science Foundation (NSF) sponsored Synthesis Coalition, which is focused on improving engineering education at all levels, includes a component called NEEDS (the National Engineering Education Delivery System) which focuses specifically on the creation of networked information resources in support of elements of the Synthesis program [Saylor, 1992]. The Council on Library Resources is examining how to improve access to engineering information [Council on Library Resources, 1990; Council on Library Resources, 1992]; here, again, network-based access to information is viewed as playing a key role. Major scientific programs in areas ranging from Molecular Biology to Global Climate Change all view the development and operation of networked databases as essential program components [Olson, 1993; Stonebraker, 1992]. The higher education and research library communities have created the Coalition for Networked Information (CNI), a joint project of CAUSE, EDUCOM and the Association of Research Libraries (ARL) to specifically address

---

<sup>1</sup>This was signed into law by President Clinton on June 8, 1993 as Public Law 103-40

<sup>2</sup>While agencies such as NASA seem to have made a commitment to the **networked** information model of access to agency information, other groups—for example, the **Bureau of the Census—have addressed the distribution of government information through the publication** of CD-ROM databases, leaving it to the user communities (i.e. the universities or the Federal depository libraries) to determine how the information they publish on CD-ROM should be made generally accessible through the network. At present most CD-ROM databases are unsuitable for use in the networked information environment, despite efforts by group such as the Air Transport Association and the Intelligence Community Data Handling Committee to define network interfaces to CD-ROM databases [Bowers & Shapiro, 1992]. In still other cases federal agencies such as the SEC have formed alliances with the private sector (Mead Data Corp., in the case of the SEC) to offer access to federal information through the networks [Love, 1993]. There is a great need for more consistent policies for access to government information through the networks.

<sup>3</sup>Utah makes legislative information available through the Internet.

<sup>4</sup>In California, Assembly Bill 1624, currently under consideration in the state legislature, would make legislative information available through network access.

<sup>5</sup>The Maryland State Librarian is leading a major effort to link libraries throughout the state using the Internet. Similar projects are under discussion in Virginia and other states, [Library of Congress Network Advisory Committee, 1992].

networked information issues [Peters, 1992a], as well as devoting substantial attention to broader networking initiatives within the programs of the parent organizations (for example, EDUCOM'S Networking and Telecommunications Task Force); the CNI initiative reaches out beyond the research and education community to reach providers of networks and information technology, publishers, and even, to a limited extent information users in the private sector.

There is a great deal of talk about "digital libraries", "electronic library collections", "electronic libraries", "electronic journals" and "network-accessible library collections"; such visions have captured the imagination of many scholars, educators, librarians, and policy-makers [Lynch, 1991a], and are increasingly attracting the interests of the commercial sector—particularly publishers, mass media, and other information providers—as potentially lucrative new markets. Indeed, the upcoming transition to electronic information resources is viewed hopefully by some as a means of relieving increasingly intolerable financial pressures on the entire system of scholarly publishing.<sup>6</sup> Yet the definition of a digital library remains controversial. Personally, I prefer to consider the viewpoint that stresses electronic library collections; a library is an organization that acquires, structures, provides access to, and preserves information of all kinds, and within this context network-based electronic information is just another type of material that the library can acquire and manage (albeit one with many unique, intriguing and novel properties). Additionally, the availability of digital content and sophisticated information technology of course permit the library as an organization to offer new organizational and access services, and to move from a primarily "passive" organization to one that actively provides information to its users through interactive services and automated, network-based, content-driven information delivery.

When we speak of digital libraries, however, we invoke not only this concept of electronic library collections but also visions of the integration of networked information resources of all kinds (including, for example, numeric databases that libraries have typically neglected and remote sensing databases that are collected as part of various scientific programs outside of the library context) into new collaborative environments (co-laboratories) [Lederberg & Uncapher, 1989]; some term such collections of databases and related networked information resources to be digital libraries. There is discussion of coupling information technology with a new partnership among researchers, information management specialists and information technology experts

---

<sup>6</sup> The **primary** source of these **pressures** is that libraries can afford **to purchase a smaller and smaller part** of the annual output of scholarly books and journals worldwide. The roots of this crisis are complex and form the subject of extensive debate between librarians, publishers, and academics. Many librarians argue that the prices for these materials are rising far in excess of the rate of inflation, in part due to price gouging by the publishers. The publishers submit that the size of the body of annual research is growing rapidly, and that library funding has not kept up with this rate of growth; they also identify other factors such as international currency fluctuations that have contributed to the problem in recent years. For a survey of some of these issues see [Grycz 1992]. A discussion of these issues is outside the scope of this paper; however, it is important to recognize that conversion of scholarly materials to electronic formats may reduce printing, distribution, handling and storage costs for these materials somewhat, but will generate new costs in retooling editorial and production processes and in investments in information technology and infrastructure. Overall, it is not clear that conversion to electronic formats will substantially reduce costs for scholarly materials, though it will undoubtedly redistribute these costs. Further, if, as some librarians argue, much of the problem is publisher profiteering, a shift to electronic scholarly publications will only alter the economics if it causes substantial changes in the system and the role of publishers—particularly commercial publishers—within that system.

under a model called "knowledge management" that relies heavily on networked information resources to directly support and integrate with the research process itself as well as the diffusion of new knowledge to the research community and the dissemination of research results [Lucier, 1990; Lucier, 1992]. I view these new networked information resources to be something fundamentally new, and different from library collections (though they might, in some cases, be part of a library's collection, or a part of the services offered by a library); certainly they are different from libraries (as organizations), though in some cases libraries may be the organizations that create, manage or fund these new networked information resources. We will need a new terminology and taxonomy for the networked information age. But, in any event, the focus of this paper will be collections of network-accessible information resources and the roles of libraries in maintaining and providing access to them.

There are a vast number of experiments underway at present to use the network to deliver or provide access to bitmapped images of print publications, including document delivery services being offered by CARL, Engineering Index, University Microfilms, Faxon and others, often in complex business partnerships with traditional secondary (i.e. bibliographic) database access providers such as OCLC, RLG, or Dialog. Primary scientific journal publishers such as Elsevier and Springer-Verlag are conducting experiments with various universities under which they are making the contents of some of their journals available electronically to institutions either under site license or pay per view arrangements. In addition, various third-party aggregators and relicenses such as UMI and Information Access Corporation are now licensing full text or page images of journal contents for sizable collections of journals in specific areas directly to institutions, and a number of publishers are making the text of their newspapers, magazines, or other publications available for searching through database access providers such as Dialog or BRS on a transactional basis.

Indigenous network-based electronic journals are now well established, and their number continues to grow (though it is important to recognize that they are still a very minor force, compared to the existing print journal system, in most disciplines). The vast majority of these are made available for free on the Internet, and they include both journals structured in analogy to peer-reviewed print journals, such as *Postmodern Culture* (which Oxford University Press has recently agreed to market in a parallel print version), *Psychology*, *Public Access Computer Systems Review* and many others (see Michael Strangelove's bibliography [Strangelove, 1993] and other regular publications that are similar to edited newsletters (*Newsletter on Setials Pricing Issues*, *TidBits*, etc.). The edited newsletters are part of a continuum that fades off into "moderated discussion lists" implemented through LISTSERVERs or other mail reflectors, which might be viewed as continuously-published electronic newsletters that exploit the electronic medium of the network to avoid the need to gather submissions and commentary together into discrete issues. There are also thousands of unmoderated discussion groups which provide additional forums for discussion and information interchange; these have some elements in common with newsletters or other publications, but are really a new and unique artifact of the networked environment. Recently, a few for-fee journals have begun to publish either solely in electronic form (i.e. the OCLC/AAAS *Current Clinical Trials* experiment, a fully peer-reviewed journal [Keyhani, 1993; Palca, 1991], or the *Computist's Communiqué* by Ken Laws, more of a newsletter), or in parallel print and electronic editions (e.g. John Quarterman's *Matrix*

News). It seems probable that the development of for-fee journals on the network has been inhibited by publisher concerns about the acceptable use policies that govern traffic on much of the Internet; as it appears that the acceptable use policy may well either be abandoned or interpreted liberally enough to comfortably accommodate this type of network based publication and publishers can find other publishers distributing journals in the network environment without problems, the number of for-fee journals will grow rapidly.

Paralleling these initiatives in the creation of content, a great deal of effort is being devoted to the development of tools for network navigation and information retrieval. The development of standards for resource description, location, and access in a distributed environment are also a key part of the development of the tools themselves. Major efforts in this area include the Gopher project at the University of Minnesota [Alberti, Anklesaria, Linder, MacCahill, & Torrey, 1992; Wiggins, 1993], the World Wide Web system from CERN [Berners-Lee, Cailliau, Groff, & Pollermann, 1992], the WAIS system that was originally developed as joint project of Thinking Machines, Apple, Dow Jones and KPMG which is now being carried forward by a number of organizations, including the new NSF-funded Clearinghouse for Networked Information Discovery and Retrieval in North Carolina [Kahle, Morris, Davis, & Tiene, 1992a], thearchie system developed at McGill University [Emtage & Deutsch, 1991; Simmonds, 1993], the resource discovery work carried out by Mike Schwartz and his colleagues at the University of Colorado, and others [Schwartz, Emtage, Kahle, & Neuman, 1992; Schwartz, 1989; Schwartz, Hardy, Heinzman, & Hirschowitz, 1991]. Recently, the National Science Foundation awarded a sizable contract to AT&T for the development of directories for the Internet; while this contract is primarily to compile and operate such a directory using existing technologies and standards, and the resource directory being developed does not seem to incorporate any sophisticated classification or retrieval approaches, the AT&T effect is likely to focus and stimulate further effort in this area. More research-oriented work is also underway in developing cataloging and directory tools through the CNI Topnode project,<sup>7</sup> the Department of Education funded project for cataloging Internet resources at OCLC [Dillon, 1993], the work of the Library of Congress on cataloging practices for electronic information resources [Library of Congress, 1991a; Library of Congress, 1991b; Library of Congress, 1993], and the efforts of various working groups within the Internet Engineering Task Force on document location and identification.<sup>8</sup> Other important standards work includes activities such as the Z39.50 Implementor's Group, which is addressing both functional extensions to the Z39.50 computer-to-computer information retrieval protocol, a basic tool for searching information resources on the network, and also attempting to resolve interoperability issues as the Z39.50 protocol moves toward widespread implementation [Lynch, 1991b; Lynch, Hinnebusch, Peters, & McCallum, 1990]. In addition, of course, standards developed within broader communities to describe various types of record and document interchange formats are of critical importance to the development of networked information retrieval tools.

---

<sup>7</sup>Information on the current status of this project can be obtained from the Coalition for Networked Information, or by searching CNI's file server at <ftp.cni.org>.

<sup>8</sup>While not much has been published on this yet, the IETF should be issuing a series of draft RFCs within the next six months. The general approach that is being pursued is described in Section 13 of this paper.

Institutionally based projects at universities such as Carnegie-Mellon (Project Mercury) [Arms, 1992], Cornell (various projects) [Lesk, 1991], the University of California (various projects) [Lynch, 1989; Lynch, 1992]. The University of Southern California Watkins, 1991] have focused on developing systems for the storage and delivery of electronic information resources to institutional user communities, in some cases integrating and building upon tools and standards developed on a national and international level. Some other projects, notably at Yale, Cornell, Barry Shein's Online Book Initiative (hosted at world.std.com) and Michael Hart's Project Gutenberg are working with public domain collections (either out of copyright materials or government material not subject to copyright) as prototypical electronic library collections. In some ways, these out of copyright collections are very attractive test cases as they permit the institution to focus on the technology and end user requirements of the application without becoming enmeshed in economic and legal (intellectual property) concerns.

Not all work on networked information access is rooted in the higher education and library communities or the efforts of commercial firms that primarily serve these communities. For example, a number of communities have deployed versions of the Freenet system developed by Tom Grunder in Cleveland, Ohio [Grunder, 1992; Watkins, 1992]. This is a system which provides the electronic analog of a town, complete with a town hall, libraries, schools, medical facilities, discussion groups, and other areas. While some implementations have been supported in part by libraries and/or universities, Freenets may equally well be sponsored by municipal governments or private citizen groups outside of the higher education and research communities. In addition, commercial services such as CompuServe and America Online are now well established and offer access to sizable collections of information; their primary user communities are outside of the academic or library worlds.

Recently, Carl Malamud established a project called Internet Talk Radio which offers a mixture of interviews and live coverage of speeches and other newsworthy events; the content is somewhat similar to that of the C-SPAN cable network, although it includes announcements from commercial underwriters similar to those found on Public Television (not really full scale advertising by sponsors) [Markoff, 1993]. Internet Talk Radio has coverage into the National Press Club in Washington, DC and is scheduled to "broadcast" its first coverage of a congressional hearing later this summer, Internet

---

<sup>9</sup>It should be noted that, while out of copyright material will be a very important resource for libraries that wish to explore the electronic storage and dissemination of material exactly because this material is not subject to copyright constraints, such material is substantially difficult to identify; worse, the identification of such material is growing more complex as the issues are explored in more depth. Consider first simple textual materials. In the US, currently, the period during which a work is subject to copyright is defined by the life of the author plus a certain number of years, rather than the old definition which was just a fixed number of years from the date of copyright. This means that a library that wants to determine whether a work is still under copyright protection must attempt to determine whether the work's author is still living or when he or she died. This is a major research problem. Further, international copyright issues have become extremely complex. For example, Project Gutenberg recently made a copy of Peter Pan available, since the work appeared to be out of copyright in the US, only to subsequently discover that there is apparently a special exemption for this work under UK copyright law that permanently assigns copyright protection to this work and donates the proceeds to a hospital in the UK, and thus the electronic text could not be distributed in the UK legally. How this strange exemption in UK copyright law relates to the Berne Convention and the internationally reciprocal copyright agreements that the US has agreed to honor is a subject for legal scholars that I will not speculate upon here; however, it is a good illustration of the problems of identifying material that is no longer subject to copyright.

Talk Radio captures audio from these events and distributes it over the Internet in real time using multicast technology to interested listeners; in addition, digital files containing the audio for the broadcast events are archived on the network and can be retrieved and played by individuals with appropriate audio boards in their machines at any time on demand. Even for simple audio, these files are quite large and stretch the capabilities of many machines on the net. The Internet Engineering Task Force has been experimenting with digital audio and video distribution of key sessions at its meetings using similar multicasting technology, though this is considerably more taxing for the network due to the data volumes involved. As the network capacity expands and the supporting capture, playback and multicasting technologies become more mature and more widely available it seems likely that this type of audio and video coverage of events of interest, both multicast real time and stored in archives for later demand playback, will become more commonplace.

Yet, despite this rich and wide-reaching series of projects (and what has been described here is only a sampling intended to give the reader a sense of current developments) which we hope will yield knowledge and insight that will inform future efforts,<sup>10</sup> there is little consensus about the future structure of electronic libraries, digital libraries, network-accessible libraries or whatever one wants to call them—or even if these terms refer to the same things. Some people refer to collections of network-accessible files as a digital library; this is common in some parts of the computer science community, for example. Some from the publishing community speak of digital libraries when a perhaps more accurate term might be a digital bookstore or information store. Those viewing the evolution of electronic information resources from the library tradition tend to think of networked information as simply one more component of a traditional library's collection, subject to the basic library functions of selection, organization, provision of access, and preservation, suitably adapted for the unique characteristics of the network environment (for example, you can select a network resource that you provide access to without physically making it a part of a given library's collection—in other words, performing acquisitions without taking physical possession, as distinct to providing some form of subsidized access to a resource that the library continues to regard as “external” and available through mechanisms such as interlibrary loan or short term contract to subsidize access in the networked environment). Indeed, with the network's ability to dissolve geographic boundaries and unify access to autonomous, physically and organizationally distinct resources, fundamental questions are being raised about the nature of these future electronic information collections—for example, might there just be one logical “library” of information for each discipline [Loken, 1990], perhaps managed by a professional society, in the sense that the user would communicate with only a single access-providing organization for the discipline's literature?

---

extent to which current prototypes will in fact help to resolve the open questions is problematic. Many of the prototypes are being rushed into production use, without any systematic evaluation of the human or economic outcomes. Too often there is funding to build prototypes, but no funding to evaluate them rigorously. In some cases the economic viability of projects beyond the prototype state is unclear, and there is a real lack of common economic models that permit comparisons to be drawn between projects. The definitional difficulties surrounding the concept of “digital libraries” are indicative of the severity of this problem.