CHAPTER 3 *December 1992 Workshop*

Chapter 3, Section A: List of Participants

Workshop on Digital Libraries
Xerox Palo Alto Research Center
December 9-10, 1992

Stanley Besen, Charles River Associates
Christine Borgman, UCLA
Esther Dyson, EDventure Holdings Inc.
Monica Ertel, Apple Library of Tomorrow
Georgia Finnigan, The Information Store, Inc.
Linda Gass, Adobe Systems Inc.
Tony Hall, Dialog Information Services
Pam Jajko, Syntex USA Corporate Library
Michael Lesk, Bellcore
David Levy, Xerox Palo Alto Research Center

Mark Liberman, University of Pennsylvania
Richard Lucier, University of California at San Francisco
Jean Mayhew, United Technologies Research Genter
Peter Menell, Stanford School of Law (1992-93)
James Michalko, The Research Libraries Group, Inc.
Eugenie Prime, HP Labs Research Library
Frank Urbanowski, MIT Press
Gio Wiederhold, DARPA/SISTO
Terry Winograd, Interval Research (1992-93), Stanford University

Xerox/PARC Participants

William Crocca
Margaret Graham
Giuliana Lavendel
Larry Masinter
Geoffrey Nunberg

NSF Participants

Su-shing Chen
Jane Caviness
Y.T. Chien
Larry Rosenberg
Maria Zemankova

Chapter 3, Section B: Agenda

Workshop on Digital Libraries — Xerox Palo Alto Research Center

Wednesday, December 9, 1992

8:30 Arrival (Main reception area-upstairs)

9:00 - 11:00 Introductory Remarks:

Mark Weiser (PARC welcome and overview)
YT Chien & David Levy (Background, Goals, Agenda)
Mike Lesk (The competitiveness issue)
Jane Caviness (NREN & Gigabit Testbeds)
Gio Wiederhold (A perspective from DARPA)

11:00 - 12:30 Around the Table - All participants (Vision statements)
12:30 - 2:00 Lunch
2:00 - 4:30 Group Breakout Session 1
4:30 - 5:30 Report Back from groups & General Discussion

Thursday, December 10, 1992

8:30 - 11:00 Group Breakout session 2
11:00 - 12:30 Report back & general discussion
12:30 - 3:00 Lunch and Summary
3:00 - 5:00 Demos (selected PARC projects)

Chapter 3, Section C: Lesk Summary

The Digital Library: What is it? Why should it be here?

Michael Lesk

The second in a series of NSF workshops on a digital science library took place in Palo Alto, Dec. 9 and 10, 1992. As time goes on, this project is becoming ever more vague; not only did we not agree during the workshop on what should be in the library, but many attendees did not even think this was important. We also reached no resolution on the problems of economic compensation for publishers and authors (not to mention intellectual property lawyers), but we did notice that there is enough non-protected material available to possibly start building a useful collection without addressing that issue. Nor did we settle why such a library would benefit the U. S. more than other countries if it were built here. However, we did see considerable agreement that a digital library would be a Good Thing.

Mark Weiser welcomed us to Xerox PARC with a description of Xerox work in information storage and access, social aspects of electronic information handling, and related areas such as linguistics, ubiquitous computing, and scanning/printing hardware. Xerox has, for example, done a project at Cornell University in which a thousand 19th century books were scanned and reprinted.

Y. T. Chien then explained the structure of his part of NSF and asked for input from the community about the importance of digital libraries and their relation to other technologies. Other government speakers were Jane Caviness who talked about the NREN, and Gio Wiederhold who described some of the information-related work in DARPA.

I described the risk of foreign countries, such as the French using Minitel and their proposals for the Bibliotheque Francais, dominating the next generation of information systems. I also described the original proposal ($50M over five years, half to convert a basic undergraduate science library of perhaps 100,000 books, and half to fund exploitation of this resource). Finally in the introductory speakers, Stan Besen of Charles River explained the increased risks that digital technology poses to intellectual property rights by facilitating piracy, private copying, and the creation of derivative works.

As each person gave their personal statement, we heard a sequence of people emphasizing the importance of social concerns, of bringing information to people rather than just piling it up, and of intellectual property issues. What was noteable was the lack of any fear that technology was unable to do the job; our questions are entirely to deal with context and politics. An interesting question raised was that this library, as envisaged, was in some ways an alternative to bookstores as well as to libraries, but we really did not follow up this point.

We then divided into several groups for the next day. These groups were intended to address more specific issues, with the thought of sketching out a possible report urging that the United States build a digital library, to see if it could be made convincing and if there was any agreement on its content.

One group was supposed to define the library. It came back with a statement that a digital library is a "distributed technology environment which dramatically reduces barriers to the creation, dissemination, manipulation, storage, integration and reuse of information by individuals and groups." They suggested the national initiative should contain some specific testbed projects, but gave no guidance on what these should be. In other words, they not only failed to define the collection, they didn't really even describe the system that would hold it.

The next group was asked what technologies were needed to build such a library: it rounded up the usual suspects, recognizing the particular importance among them of the NREN, OCR and conversion tools, software for filtering, retrieving, summarizing and so on, and user interface research. Security/privacy concerns were also considered important, as were some standardization issues.

The last group on the first day pointed to the conflict between intellectual property rights and free access, as well as the ‘‘political economy’’ problems (e.g., university tenure procedures today are intertwined with publication issues).

On the second day, we then re-formed into a new set of groups to deal with the most important question: why should the U. S. have a digital library? We hoped to sort the answers into three categories: research applications, business applications, and educational applications. I was disappointed by many of the answers, which suffered from circularity (e.g., a well educated work force is good for business), from a confusion of means and ends (e.g., a digital library provides access to multiple formats of information), and from a lack of quantification. Among the possible useful arguments that came out of this process:

Research:

1. One-third of a researcher’s time is spent looking for information, and searching is more efficient in electronic systems;

2. Electronic libraries would let researchers get at existing data reserves which they have trouble finding now, and speed up their work;

3. Rapid identification of related work would reduce duplication;

4. The entry barriers to starting new research projects would be lowered.

Education:

5. An electronic library would help self-education and support distance learning.

6. The United States is falling behind in science education and must do something (although the countries ahead of us don’t have digital libraries either).

7. Electronics can help libraries deal with increasing volumes of information.

Business:

8. Readily available information levels the playing field as between small and large business, and business in small and large towns.

9. Superconductivity and xerography were both found by random browsing, and we need more of that.

10. Better access to regulatory requirements, product literature, standards and the like would help business.

Social:

11. Rapid access to commercial resources in time of national emergency. For example, during the Gulf War it was necessary to identify quickly the manufacturers of the transistors that went into IFF devices.

The most convincing arguments, to my mind, were:

a) Biotechnology is a very productive area and makes enormous use of on-line information – if we could duplicate that productivity in other areas we would be way ahead.

b) Allegedly, studies in Blacksburg, Va. and in Pennsylvania show a positive correlation of information availability and economic growth.

c) Don Swanson (Univ. of Chicago) has demonstrated the ability of purely information investigation for finding medically promising ideas.

But all these benefits were a bit fuzzy, and we have few quantitative measurements as backup for these ideas.

The workshop then closed with final statements by Y. T. Chien and David Levy, and some follow-up discussion.

There was no really good answer to the question: why should the U. S. do this? Why will it benefit our country specifically? We all agree it would be best if such libraries were built internationally, but we also think that the NREN gives the U.S. an advantage in distributing this information within our country, and that we would not be better off letting other countries build such a system. If so, we will find ourselves paying for access to information, much of which was generated within the U.S.

What should be done at the next workshop? Important topics for the discussion at that workshop, I think, should include;

a) Specific consideration of possible compensation schemes for publishers: whether the concurrent-use site licenses now common in the software business, the unlimited use licenses favored by librarians, an adaptation or an equivalent of the Copyright Clearance Center, greater use of page charges, or something completely different (e.g., pledge breaks on the Internet).

b) Specific consideration of the goals list, and a resulting suggestion for content for at least one or two specific projects, including both form and subject matter. We need to know what kinds of material are to be in the library, both in terms of form and content.

c) Specific discussion of why this should be a United States funded activity; would the benefits of having such a library accrue primarily to us if we built it? Our advantages include the NREN and the creation of much of the research, plus our wide number of small businesses.

d) Coordination with other groups having similar discussions. For example, on the same days as our meeting there was another one in Washington with similar goals, also run by NSF, but that one focussed on the social sciences. As part of this, we should continue our efforts to include the publisher and academic library communities in our discussions.

e) And, on the side, perhaps NSF should require that each citation in a grant proposal should include the country of publication and the price of the publication to libraries, and reviewers should be instructed to ignore all of the proposers' publications which either appeared in foreign journals or in journals which charge libraries more than $1000/yr.

On balance, to get a report that would convince Congress to fund something,

a) we need to say what it is,

b) we need to explain why it is good for the United States,

c) we need to show that its economic benefits will exceed its costs, and

d) we need to do all this in simple words, with illustrative graphics, convincing testimonials, and respectable evidence as backup.

In particular, anyone reading this who can provide citations to studies showing that research, education, or business benefit from more information and from more electronic information in particular, is asked to communicate with the author.

Chapter 3, Section D: Statements of Participants

Questions and Assignments for Workshop Participants
by Y.T. Chien

During the workshop, we plan to look at digital libraries from three perspectives, from the perspectives of the users, the technology, and the industry. Below we've compiled a list of questions that can be asked from each of these perspectives (as well as some questions that don't naturally fall into any of the categories).

To get you thinking about the issues, and to help us organize the workshop's breakout sessions, we'd like to ask you to choose one of these topics (or more, if you wish) and prepare answers to some or all the questions associated with it. We're only looking for a one-page statement, but you should feel free to write more.

Questions:

1. The users view

a. In a few words or sentences, what is your vision of a digital library, say 10 to 15 years from now?

b. Given that vision, what are the grand challenge applications of a digital library?

(A Grand Challenge application can be roughly defined, in this context, as an application of digital libraries, which when fully implemented may have a significant impact on the scientific or economic well-being of our society.)

c. Who might use electronic libraries? What information and services do they need? What kind of support, training, and pricing policy is needed to foster more effective use?

d. What is the use of data from foreign sources? How is such use related to international competitiveness?

2. The technology view

a. What research is needed to realize a large-scale digital library? For example, research in computer science, linguistics, sociology/anthropology (work practice studies), business/economics, philosophy/history, etc.

b. Beyond high speed networks, what other infrastructure elements are needed to provide the foundation for digital libraries?

c. What component technologies are essential to the development of a digital library?

d. How do we characterize the functional capabilities of a digital library environment?

3. The Business View

a. What products or services might industry provide for use with or in digital libraries? What types of products or services might be important for your company to consider developing in the future?

b. What services provided by digital libraries would be particularly useful to industry? How would they enhance competitiveness, time-to-market, etc.?

c. What are the non-technical barriers to the realization of digital libraries in the foreseeable future, e.g., copyrights, standards, and other social factors?

d. What kinds of partnerships (e.g., research, training, experimentation) must be developed between industry, academia, and government to realize digital libraries on a national scale?

4. Miscellaneous

a. How is a national digital libraries effort related to other major initiatives, e.g., high performance computing and communications (HPCC), manufacturing, biotechnology, etc.?

b. What should NSF do with respect to development of a technical and social infrastructure linking researchers with each other and with consumers of research?

Christine Borgman, UCLA

NSF WORKSHOP ON DIGITAL LIBRARIES
Xerox PARC, 8-9 December, 1992

Response to “Questions and Assignments for Workshop Participants”

Of these 4 categories (Users, technology, business views, miscellaneous), my interests and contributions are primarily in the Users’s View category. The discussion document I drafted in response to the July 1992 NSF Workshop on a National Electronic Library Initiative is relevant to this perspective, and thus is attached below. The paper also comments on other areas, particularly 2b (technology infrastructure elements), 2d (functional capabilities), 3b (business products or services), 4b (misc - technical and social infrastructure). Also, in relation to 4b (relation between the digital libraries effort and others such as the HPCC), Appendix A.2 (Relation to HPCC) of the July 18, 1992 Workshop report (referenced below) draws specific links among the HPCC and digital libraries efforts and the Grand Challenges. This also is a useful discussion document for the December meeting.

The Information Infrastructure Technology Act (also known as Gore II; distributed to December workshop participants) provides a useful definition and scope statement for “digital libraries” in Section 7, including 7 points for research and development. The role of NSF and other agencies is presented explicitly in that document. The scope appears to include all the major areas listed in my document and others previously generated in this workshop series. The Gore II document seems a valuable starting point for setting a specific funding and development strategy.

While this meeting should not be bound by these prior discussion documents, they are thoughtful pieces and we might make efficient use of our limited time by using them as a point of departure.

William T. Crocca, Xerox

1a. The digital library will be 1/3 equal partners along with publishers and retailers in the distribution and management of intellectual assets. Its role will be navigation, archiving and republishing as well as the traditional roles of access and storage. Distinctions will be made between content and delivery media, removing most artifacts from conservation labs and leaving only those artifacts which are intrinsically precious. Library economics will change. Since major institution holdings have economic value, they will have opportunity to create substantial revenue streams. Also, the shelving costs could come down, depending on collection management strategies.

1b. An important, near term "grand challenge" is indexing and navigation. These are unsolved problems which, when solved, should enable a dramatic opening up of the information base intrinsic to library collections. (I was originally going to use the phrase "knowledge base", but information has to be found, read and internalized before it becomes knowledge . . .)

1c. Potentially, anyone with either a TV set or a telephone. Furthermore, one would expect simple, low-cost printing systems (FAX, photo-CD, etc.) to be ubiquitous, thus relocating hardcopy to a different point in the creation - distribution - consumption chain.

1d. This is a large and complicated question even ignoring the competitiveness issue. However, my initial reaction is that the "competitiveness" issue can easily become divisive and should be proscribed from the libraries' role. Let information ownership and national practices deal with this one. The substantive issue seems to be around language and cultural context. This will effect navigation, understanding and presentation.

2a. Navigation and indexing, linguistics and translation, and indeed everything you mention. The question is what of these is special to the library? I'm not sure any of it is -- just the way it all comes together.

2b. High density, high reliability, VERY long term storage. This implies hardware, media and format standards. It also implies internal and mutual consistency across collections w/r/t format and indexing standards. We also need some sort of copyright/royalty/ownership collective in order to allow information to be marketed and used with at least as much facility as the technology provides. Finally, we need cross use of the various electronic common carriers (telephone, TV, etc.) to enable the kind of connectivity this vision will require.

2c. Compression/decompression chips at HD (High Definition) video rates with high resolution and LONG scan lines for large documents. Cheap, high quality display devices (both transient, e.g., CRT and hardcopy e.g., paper). A cultural mandate for real information disclosure, else arcane security systems will become necessary.

2d. . . . not sure I really understand this question

3a. Certainly the hardware, the software and the common carrier services. In addition, one can easily imagine other services such as production capturing, alternative delivery media manufacturing and other production operations a library might not want to manage directly -- especially those, like capture, which are one-time transformations.

3b. Document delivery and research. Identification of others with similar interests (this could also prove to be a liability). Need to brainstorm this one.

3c. The relationships between publishers, sellers and users. Revised skill sets and the necessary storage. New kinds of user terminal capability that is cheap (probably this will happen naturally), TV/cable/phone/etc. channels appropriately organized. The methods of use: the European approach to use it and then pay (in a standard way) for your use is more tractable than ours for the kind of future I am envisioning.

3d. I'd say similar in principle to the one between Xerox and Cornell.

4a. It offers a low-cost, high-performance information distribution infrastructure which is now only partially provided by specialty publishers.

4b. The government information resources could be organized into the first national electronic library. As such, it could experiment with the legal issues which are bound to arise with this kind of effort. The international questions could be explored by partnering with the Library of Congress efforts to provide information to former iron-curtain country government officials. This would be an interesting testing ground for the technologies and a strong signal to our private and quasi-public institutions of expectations for their future participation.

Esther Dyson, EDventure Holdings Inc.

A few answers, and more questions

What is the digital library? That term smacks of "filmed play," "horseless carriage," and the like. The digital library will be less like a library than we think, and more like itself. It won't have a physical location, nor will it really be a logical place. It will sit behind or around everything (electronic) we do. You will be able to use it while you're writing a letter, composing a presentation or analyzing a sudden sales drop in your Hungarian subsidiary; a query tool/knowledge navigator/agent savant will be just one more item on your menu (pick-an-icon or voice input) of choices.

In fact, we will perceive the digital library as an access method to information distributed all around the world. For content, it will contain elements of today's libraries -- books, printed periodicals and reference materials -- as well as videos and other more exotic media. But it will also include realtime access to designated experts (for a fee, at designated times, etc.) and other information sources such as government and private statistical and text bases (with appropriate privacy and property protections).

The main challenge of creating the electronic library is not to build it per se, but to ensure that people make their information (1) digital and (2) reachable. In the business community, that will require an overlay of allocation, accounting and payment mechanisms which should be mostly unobtrusive (although you may well want people to know how much things cost so that they don't spend recklessly).

The most interesting challenge will be to (build technology to) automatically cross-reference all this information, "understand" it well enough to find relevant items and avoid redundancy, and provide useful interfaces to make the content more intelligible. Some examples include hypertext maps such as gIBIS-CM/1, "reading" tools such as those forthcoming from (yes) Oracle, and overall layout and presentation tools.

Users will filter for each other, and good filterers will probably become professionals (like editors or active newsgroup managers today). Who will set the agenda? Today, editors/publishers and advertisers do; in the future, the reader will have more power. Should we leave filtering to imperfect software tools (as opposed to imperfect people)? If information is available all the time everywhere, what's news? What happened yesterday, last hour, last minute? Anything you didn't know yet?

All this of course assumes the requisite infrastructure:

- high-speed bandwidth,
- funding (preferably from competing private sources working hard to make their modules of the library better in a never-ending cycle of improvements),
- excellent software tools,
- ubiquitous information-delivery devices (PCs, personal digital assistants, video screens, information kiosks, etc.),
- appropriate financial underpinnings to manage use and compensation of owners/producers of information,
- appropriate legal and technical safeguards to guarantee authenticity and integrity of information and to protect privacy where appropriate. (Do you want people to know what you read or wanted to know? What are the rules governing anonymous comments? Responsibility for libel or incorrect information?)

Linda Gass, Adobe Systems, Inc.

1. The User's view:

a. In a few words or sentences, what is your vision of a digital library, say 10-15 years from now?

The digital library in the year 2005 will be available directly to end users from their personal computer (or whatever the device may be called) and will come with its own "digital librarian" - an intelligent agent that can assist you in finding whatever information you are trying to track down. You will be able to do unlimited searching - not only by author, subject, title but searches will include the content of the work. Searches will be intelligently filtered, and content will be visually rich and multi-media based with hypermedia links. Users will be able to get multiple views of the information: graphs, photos, etc., in addition to textual information. Users will also be able to access information from a variety of sources through a unified user interface and the details of the infrastructure will be hidden from the user.

b. Given that vision, what are the grand challenge applications of a digital library?

Scientific: A digital library will make it easier for scientists to work outside of academia and pursue hypotheses that are not in vogue and thus not funded or approved for research. This could lead to important medical and scientific breakthroughs.

Economic: A digital library will facilitate the work of the telecommuter, allowing people to work at what they please and live where they choose. This can help alleviate regional economic recessions. The library will also encourage small business startups as people discover that they can easily acquire the data they need to project the success of their business ideas. It will provide larger businesses with better competitive information, leading to less duplication and more innovation.

Social and Political: Impact beyond the scientific and economic should also be considered. For instance, access to digital information through the Freedom of Information Act can help insure that our environments are safe and our political choices are informed. The library can be an important safeguard of democracy.

c. Who might use electronic libraries? What information and services do they need? What kind of support, training, and pricing policy is needed to foster more effective use? Any current library user would use an electronic library: schools, universities, businesses. The key is to draw in a larger user base by making the library more accessible and more appealing than our existing libraries. The information and services needed are the digital librarian mentioned above, excellent searching capabilities, good response time. Support and training will present interesting problems as we will have to change the way people work. Even though computers are widespread in our work environment, they are still not widely used by most businesses in America today. Support will consist of some automated support tools and then there will also be an expert librarian who can assist users in using the system to its full effectiveness. Pricing policy is a tricky issue. Today's public libraries are free and we will exclude a significant portion of our population by charging for access to a digital library. Perhaps some sort of sliding scale is appropriate where users pay per view. The price per view can vary depending upon the amount of information, how current it is, etc.

d. What is the use of data from foreign sources? How is such use related to international competitiveness? This set of questions is actually rhetorical: We ARE in a global marketplace and access to foreign information can only enhance international competitiveness. A key issue is that of translation; much of the foreign information is published in the native language of the country and is of less utility to the average American user.

2. The technology view

a. What research is needed to realize a large-scale digital library? For example, research in computer science, linguistics, engineering sociology/anthropology (work practice studies), business/economics, philosophy/history, etc. The topic of effectively searching information needs research. This spans linguistics and computer science. The end result of a search could be overwhelming for a user unless intelligent relevance ranking techniques are used. Timeliness of search and retrieval on massive amounts of information is also important and we have a great deal to learn about how one can search multi-media (voice, images and video). Another area of research involves the sociological point of view: changing people's work habits. We have had 500 years of people using books and paper and now we are asking them to work in a new way: using electronic information; good user interface design will be critical in encouraging people to change the way they work. Another interesting area of research involves what happens to serendipity. If searching becomes too effective, one loses the accidental discoveries that are made by browsing. How can we preserve this feature? Moving towards electronic information also changes the business/economic structure for information. We are accustomed to paying for "widgets" - books, newspapers, etc. How will our economic structure evolve to account for electronic information? How will our legal system (copyright, piracy, privacy, etc.) evolve to handle the new issues that come with this electronic representation?

b. Beyond high speed networks, what other infrastructure elements are needed to provide the foundation for digital libraries?

- Standards for data interchange
- Visual fidelity of information
- Mass storage
- Security and accounting

c. What component technologies are essential to the development of a digital library?

- Natural language processing

- Filtering
- Better mass storage than CD-ROM

d. How do we characterize the functional capabilities of a digital library environment?
(Answered in question a. under the User's view)

3. The Business View

a. What products or services might industry provide for use with or in digital libraries? What types of products or services might be important for your company to consider developing in the future? The industry at large could provide things such as personalized newspapers, intelligent agents for processing information. Other potential products/services might be "seals of approval" by experts so that people could have some way to assess the quality and authenticity of information. Industry can provide much of the library information (the major publishers, etc.) in electronic form. Industry needs to provide tools to convert information to electronic form. The kinds of products Adobe considers important to develop are technologies that allow people to communicate electronically, especially as it relates to documents. We are working on a truly portable document format and platform independent viewing technology known as the Adobe Acrobat family of products. We are also working on technology that allows people to transform their existing base of documents, either in PostScript or paper form, into the portable document format (Acrobat Distiller and OCR technology).

b. What services provided by digital libraries would be particularly useful to industry? How would they enhance competitiveness, time-to-market, etc.? Timeliness of information and the ability to effectively search information such as new research findings, trends analysis, and general economic information will enhance competitiveness by providing industry with the tools they need to understand the market and the economic climate better. I cannot comment how an electronic library could benefit the general business as I'm only familiar with a software business but issues we need to be concerned with when creating a product span technological development, legal, marketing and sales activities. Being able to search through research papers on algorithm development may save development time. Being able to search patent and trademark databases can save legal time. Marketing and sales could be assisted by being able to search for reports/articles on customers who may have bought similar products.

c. What are the non-technical barriers to the realization of digital libraries in the foreseeable future, e.g., copyrights, standards, and other social factors? Yes, copyrights, standards, social factors such as changing work habits. Also cost.

What kinds of partnerships (e.g., research, training, experimentation) must be developed between industry, academia and government to realize digital libraries on a national scale? Anyone have any good ideas here? The best I could come up with is:

Although the Internet was sponsored by the DoD and developed by academia, I view it as one of the great successes in research. Can we use this model to put together a digital library? Can we prototype something and get it installed at various test sites to turn it into what it really needs to become?

4. Miscellaneous

a. How is a national digital libraries effort related to other major initiatives, e.g., high performance computing and communications (HPCC), manufacturing, biotechnology, etc?

Described above in section on Grand Challenge Applications.

b. What should NSF do wrt development of a technical and social infrastructure linking researchers with each other and with consumers of research?

Do it like the Internet. Start building something and try it out. One could start small by picking one area of research and implement the vision that we come up with. Depending upon the nature of the research, it may be interesting to make it available at the high school level to help put some energy back into our disintegrating educational system.

Vicky Reich, Stanford University

The ideal digital library will incorporate all the positive aspects of print-based libraries and will overcome all the barriers of the print form. Specifically:

Print libraries make it easy to:

1. find some information with little or no instruction
2. self-instruct through signs, brochures, etc.
3. locate new information serendipitously
4. browse
5. get help from a person
6. tailor the environment to specific communities and needs, so that different libraries provide different levels of information and access to different clienteles - but each different type of library has overlapping organizational constructs so the user has some familiarity as s/he moves from library to library.
7. provide a space for social interaction (play, learning, research)
8. offer inexpensive or free access at point of use for most resources
9. retain the principal and practice of fair use
10. provide organization and access to resources
11. be a cultural, scientific and technical public archive

Digital libraries will offer these advantages over print libraries:

1. materials are always available (not lost, stolen, defaced)
2. easy access to all formats, overcomes elitism of multimedia and includes all new media
3. use of resources is not physically limited
4. materials do not deteriorate with time or use/ replacement formatting is cheap, easy and keeps pace with hardware developments
5. materials are hyperlinked, and what is now “background” information (i.e., supporting research data, computer programs) is brought to the fore
6. “cataloging” and “indexing” provide rich and dynamic access - the problems of authority control (insuring that all the works of the same author are retrieved with one search command) have been overcome.

Jean Mayhew, United Technologies Research Center

DIGITAL LIBRARY

1. Users View

The digital library will be a concept from the viewpoint of the consumer or user to describe various forms of electronic information access. The electronic reader tucked in the pocket or briefcase, the self-contained unit in the home or office, and the system accessable halfway around the world will all fall into this concept.

Challenges exist in the evolution of the information culture and how it is regulated. The development of the culture will depend not just on the introduction of technology but also on a much deeper understanding of how the consumer uses information. Current laws and regulations do not reflect where the consumer needs to be in the use of information and are becoming the pacing item in the acceptance of technology into this culture.

Digital libraries will be used by the general populace, often without even realizing it. The newspaper delivered to the home will be an offshoot of the library. The scope of the information available will be limited only by the development of the culture. Services will be valued which sift through the digital library and quickly match user needs.

Training will be non-existent for products with the broadest distribution. Increasingly training will be built into the more sophisticated technologies.

Pricing will be dependent on the perceived value at the level of access. Retrieval of simple directory information will result in minimal cost. Simultaneous machine translation of foreign language documents may be relatively expensive.

Data from foreign sources is critical for entities competing in a global economy. Unless national security is an issue, multinational corporations require freedom to share information across borders.

Geoffrey Nunberg, Xerox PARC

The Digital Library: A Xerox PARC Perspective

Whatever form the digital library finally takes, its success will be measured by its ability to support current practices, media, and forms, while making provision for the new ones that are likely to emerge. And it must serve current constituencies (libraries, researchers, students) while opening itself to new users and communities, whose needs may not have been addressed by traditional libraries. From this simple principle -- the need to accommodate the old and the new -- some more specific principles follow:

1. First, the design of the digital library should be based on a rich understanding of current practice in specific domains -- libraries, research institutions, schools, hospitals, as the case may be. Some of this understanding will come from questionnaires, interviews, and experimental investigations of information-retrieval tasks. But this phase of the design process should also include detailed ethnographic observation of the production and use of documents. The object of this kind of investigation is not to build specific models of work-flow into the design, but rather to understand how the use of technology is accommodated to the work process. And these studies shouldn't be restricted to the narrow use of the technology to accomplish particular goals, but should also look at how people instruct each other in the use of documents and technology, and at how particular practices -- (e.g., looking up a reference) are embedded in the larger context of work activities (e.g., writing a paper, treating a patient). In the end, that is, we want to know how access to information affects the overall operation of organizations and the productivity of work practice. In this connection, it is also important that any digital library project be integrated with network facilities that allow users to exchange information both about the use of the system and about other sources of relevant information.

2. The design should be sufficiently flexible to accommodate the new practices and new document types that are likely to emerge over the coming years, partly as a result of changes in the technology of information management, and partly as a result of independent changes in the nature of work and communication. In addition, we should recognize that the digital library will reach users and communities who currently are not aggressive users of library services, and whose interests and needs may not be identical to those of current users. For these reasons, it would be a serious error to predicate the design of the project on a single model of the work process or on the practices of current users alone, or to presuppose a limited repertory of document types and genres.

3. The design should support all types of document use and all document media. In particular, long-term experience suggests that paper documents will continue to play an important role in all aspects of communicative practice, and that all components of the system should be designed with a view to permitting users to move back and forth between paper and electronic representations as their interests require. In particular, this means that it is important to plan for the creation and distribution of hardcopy documents. In addition, it is unlikely that a digital library will wholly replace traditional print documents, or that new institutions will wholly replace the role of traditional libraries and scientific publishers. Thus the digital library will be most effective if it is considered as one of a battery of forms and institutions that serve scientific communication, rather than as an omnibus substitute for them.

4. The digital library should support a range of search and browsing procedures. To this end, we require not just greatly improved content-based search tools (e.g., intelligent information access, automatic summarizers, translation tools), but also search tools and interfaces that support browsing of document databases and individual documents in ways that are comparably effective as the informal procedures that users now employ in dealing with paper documents.

Gio Wiederhold, DARPA and ACM Publication Board

I am proposing a rapid turn-around experiment to address one specific aspect of Electronic Publication: Referred, Archival Material of a quality that can be used for objectives as academic promotions, recordation of definite hypotheses proofs, etc.

Much of the pressure to publish and increase the number and size of journals is related to publication pressure. The information needs of the public might be as well served by publication on conference proceedings, etc. and requesting full reports in report form from the author's home institution. If we do not address this aspect of publication pressure we fail even with the best technology.

My specific plan is to gain experience with electronic journals by a rapid prototyping approach using predominantly existing resources, in various phases, each justified by success and analysis of the preceding phase:

The first phase establishes electronic dissemination of referred backlogged papers within ACM:

1. Only include completely referred and accepted, but publication backlogged from TODS, likely TOPLAS, and other journals interested in cooperating.
2. Papers to be removed when published in paper form in journals, to remove any nervousness due to competition with the paper world.
3. Papers so available could be listed in CACM, Member Net, the appropriate journal, and an Internet/Usenet listing.
4. To access a paper a single-use authorization code is to be obtained via some number as 900-ACM-TODS, with the charges collected by the telephone company to become income for ACM.
5. Charges are to be set equal to the copyright fee now published.
6. The single use code enables Internet FTP retrieval of the requested paper from an ACM controlled archival site.

7. Papers so distributed will be headed as ‘Printed for the exclusive use of John Smith’, to somewhat inhibit excessive re-distribution, i.e., electronic xeroxing.
8. Monitoring of usage and feedback preparatory to a possible phase 2.

A second phase would let us print the abstracts in the journal as soon as possible, with a reference to the electronic access mechanism.

A third phase could make lengthy appendices, corollary proofs, tables etc. available only in electronic form, but now persistently available.

A fourth phase could avoid publication of full papers, when the expected audience is limited. This would require cooperation of editors and authors.

In all phases high-quality refereeing must continue.

Terry Winograd, Interval Research and Stanford University

2. The Technology view

The term “digital library” is both suggestive and misleading. Misleading because it evokes the image of the traditional library - a storehouse of objects to be examined or checked out by “readers”. As pointed out in some of the background papers, the digital medium will inevitably do away with this notion, which is grounded in the physical contingencies of print media. I will take the long term view here, without addressing the obvious questions about the intermediate waypoints and transformations that will get us from here to there.

We can look at an overall chain of functions in the traditional system as:

author publisher librarian reader

It is the intent of the author to produce materials that will be of interest to some readers, and of the readers to do with those materials what they will. These endpoints are the fixed part of the chain. The middle involves functions that have historically combined physical, commercial and informational roles:

PUBLISHER

Physical: Produce printed, bound objects

Commercial: Market the objects, advertise, collect money, pay royalties,...

Informational: Decide what should be published and what should be rejected. Provide editorial advice and assistance to authors.

LIBRARIAN

Physical: Maintain a physical collection, lay it out in an organized way, make it available to readers

Commercial: Use resources from some (academic or governmental) entity to purchase items and maintain the environmennt.

Informational: Decide what should be stocked, catalog it, provide organization and individualized information services.

When we move to the digital world, the physical functions become unnecessary. It is trivially easy for works to be stored by authors and provided on demand to readers, with no interventions of the kinds required by print. The commercial functions are still necessary if authors are to be remunerated for their work (a fact that is too easily overlooked if we take the model of current internet usage, where most of the proferred material is public domain or provided gratis). The exact form of this in the digital world has yet to be worked out. What is the equivalent of a library paying for one copy of a book (at an inflated price over the consumer version) and then letting many people read it? A site license? Is software a useful model?

The informational roles are still useful: quality control, editorial assistance and authoritative declaration from a respected institution that a work is "publishable"; and selection, cataloging, display (in a suitable medium) and information services by "librarians." However these roles need not be centralized as they are now, since the economies of scale suitable to print collections don't apply (for example, it would be outrageous in cost to have a separate librarian for each subspecialty area of books in a physical library, but not at all so for someone who performed that service nationwide on the net).

So what is the result? There won't be publishers as "manufacturers" of works, or librarians as maintainers of storehouses. But there will be many people doing the informational work they now do, in a more distributed way. New social/economic structures will evolve to suit this change. In particular, it will only work if there are consensus conceptual models that are used in the interchange among these various parties. This will no longer be enforced by the uniform training and professional unity of the publishing or library professions, but rather will be by emergent standards, in the sense that the computer graphic interface with its mice, menus, windows, scrollbars, etc. is an emergent standard that has provided a design language in which interfaces are built.

A. What research is needed?

In this light, a core piece of the research needed is the development of a consistent and manipulable "virtuality" -- the set of objects, relations, properties and views that constitute the world of "works" (which will integrate audio, video, and other media as well as the images and print that make up traditional libraries and most current on-line collections). Physical paper-based metaphors will be a useful starting point, but no more so than the physical desktop can serve as a basis for the computer "desktop" with elements such as scrollbars and icons, which have no direct analog. As with interfaces, it is unlikely that a committee setting out to define the "right guidelines" at an early stage in the evolution of the field will succeed. But people developing multiple versions and trying them out in the new information spaces will gradually converge on usable (not necessarily optimal) designs. By doing this in a thoughtful way from the outset, we can avoid some of the local optima which are blind alleys.

B. What infrastucture elements are needed?

Here again, I will focus on the interfaces. Not the screen and its resolution and portability, but on the set of manipulable "objects" that make up the world of information that the people in various roles can produce, manage and use. Possibilities range from simple keyboard/monitor combinations to imaginative uses of 3-dimensional images, sound, animations, etc. Also, in order to augment traditional methods of conscious cataloging (categories, call numbers, key words, ...), there will be a variety of derived indexing techniques, using statistical properties, heuristic analysis of document structure, massive fast pattern matching, recording of usage patterns, etc.

C. What component technologies are essential?

- High capacity storage with reliable, invisible management of copies, archiving, etc.

- High bandwidth communications nets

- High resolution visual displays with appropriate portability characteristics and networking

D. How do we characterize the functional capabilities of a digital library environment:

A library is a set of functions, not a set of information objects.
- individual search
- browsing, including news grazing, digests, magazine racks, new releases displays, etc.
- current flow of time-critical materials (e.g., newspapers, periodicals)
- forum, meeting place
- chat, etc.
- bboards
- expertise / skilled assistance, e.g., what a good reference librarian does

A library is not for words alone.
- Not just text (or graphics)- the electronic library will include all kinds of media, including video, databases, software, active servers,...

A library is not a physical place or piece of software.
- It is a collection of interfaces, including rooms, workstations, hand-held devices,...

A library's collection is not bounded by the walls of an organization that establishes it.
- It includes the entire network of resources that is available from within.

A library's users are not bounded by the organization.
- It is a place to put things for distribution as well as for use by your customers. It includes partners, people to whom you want to distribute things, the net community...

Librarians and patrons aren't different
- shared functions
- cooperative - users generate items and indices for each other
- find on-line resources
- catalog shared physical resources
- comments, annotations, recommendations

Chapter 3, Section E: Education Working Group Summary

NSF Digital Library Workshop
XEROX PARC
December 1992

Outline Summary from Working Group on Impact of National Digital Library (NDL) on Education

Dan Atkins, Chair

1. Approach
1.1. Try to flesh out a matrix along the following two dimensions
1.2. Functions, services of a digital library
1.3. Constituency in the education pipeline
1.3.1. pre-K
1.3.2. K-12
1.3.3. undergraduate and 2 year schools
1.3.4. graduate
1.3.5. continuing (life-long learning)
2. But first here is some consensus about the vision of a digital library
2.1. As we are using the term, a digital library represents a new, expanded domain for the library. The digital library supports the creation, editing, review, and use of information by individuals and groups. It encompasses both static (images of pages) and dynamic digital documents (software objects, integrated media, compound documents, etc.)
2.2. The inclusion of dynamic documents offers significant opportunities for education and enhances the marketing of the DL vision/initiative to Congress, et. al.
2.3. But we need to be careful in the phasing of provision of static and dynamic documents.
2.4. Need inclusion of static documents to achieve useful coverage of a subject domain.
2.5. Need dynamic documents to explore new applications and to accommodate

library services on new integrated media (multimedia) objects that are being created by academia and industry.

2.6. Suggest a balanced approach which would include moving some paper based collections to on-line images as well as some dynamic documents into a digital library environment.

2.7. Need an architecture which will accommodate both (static as special case of dynamic).

2.8. The digital library is not a single source. It is the aggregation of multiple, heterogeneous resources. In a virtual way, it creates one library out of many. There was some disagreement about whether or not transition from one library to another in the virtual single library should be seamless. Some felt that the context of being in a different type of library would be useful to the user.

2.9. The DL includes a communication (collaboration) function.

2.10. Warning: Need to be sure that we know what is really needed in education.- Focus on augmenting the teacher, not replacing the teacher.

2.11. There are and will continue to be important low-tech interactive media, e.g., crayons and paper.

2.12. Paradox: as digital libraries propagate and become ubiquitous, they recede out of sight and will become invisible, everywhere and nowhere just as dozens of electric motors are buried in modern automobiles but the automobile is not thought of as an instance of electric motors.

3. Functions and services of digital libraries relevant to education

3.1 New media and services for K-12 teachers and learners

3.1.1. Tailored, customized instructional material available on demand.

3.1.2. Augment the teacher with presentation media. “Teacher is the driver.”

3.1.3. Provide interactive media directly to the students.

3.1.4. Real-time access to experts and to scarce facilities in a controlled learning environment.

3.1.5. Simulation of experiences (“learning by doing”, “project-based learning”, “CAD for kids”). Exploratory learning. Learning through “what if” and making mistakes. What are the limits of simulation? Do we need to smell the chemicals?

3.1.6. Customization to different rates and modes of learning

3.1.7. Note: Perhaps we should just document the extensive work under way in academia and industry on the creation and use of multimedia material and computers in general to enhance learning. We could motivate the National Digital Library as a distribution infrastructure for this intrinsically digital media. The NDL would be a repository for resources that can be pulled out on demand. It would facilitate standards and sharing and could serve a clearinghouse function.

3.1.8. Consider placing the first emphasis on the provision of digital versions of reference service, for example, encyclopedias. Focus on a few great products rather than a larger number of weaker ones. Be aware of the traditions of local control of textbook selection and try to avoid getting into that initially. An NDL will encounter tension between global access and local control.

3.1.9. Perhaps the NDL could make it easier to organize workshops, bring together experts, for example, to create new curriculum. Could also help capture and disseminate the results.

3.1.10. Some people expressed the opinion that on-line access was more justifiable for research than it was for education.

3.1.11. Some see the goal of education to be to produce “expert learners.” Can the NDL help do this through, for example, supporting exploration, tutorials, experience? Law Schools are set up, not to expose students to every case, but rather to teach them how to become expert learners who can find what they need for a specific case. West Law and Mead’s Lexus/Nexus are the digital libraries for law schools and are heavily used. Note that these companies heavily subsidize use of their products in law schools in an effort to create future paying customers.

3.2. New forms of information of contemporary importance to the general public and general literacy and welfare of citizens.

3.2.1. Example: Databases of environmental data which can be visualized and used through geographical information systems (GIS). Frank Caizonetti, Associate Dean for Research in College of Arts & Sciences at West Virginia University mentioned a statewide GIS system now being installed in West Virginia. He described increasing public demand for such services and also turf battles about who would provide them.

3.2.2. Example: Expert systems related to health awareness and self-diagnosis. There is a strong and growing consumer health movement in evidence in public libraries now.

3.2.3. Traditional types of library (public, school, academic/research, special) will be blended. Librarians and patrons at any of these will have access to holdings well beyond the local collection. At United Technologies (per Jean Mayhew) the corporate library collection is becoming more diverse to serve the personal needs of employees, for example, material on health, adoption.

3.2.4. Dramatic expansion of the holdings of public libraries and school media centers. Access to the NDL could make the difference between some and none information access in many rural communities. We did not, however, answer the question of where dollars would be available for on-line access if they were not available for print material.

3.2.5. The public library's role as a place and community center should continue. It however becomes an access point to a vast set of collections other than its local holdings. Also an access point to new services, e.g., a field trip through virtual reality.
3.2.6. Public libraries are an important part of a citizens' right (and obligation) to know in a democracy. Must be careful not to jeopardize this right as more and more information becomes available in digital form. Need to fund reinvention of the public library system for the 21st century just as Andrew Carnegie funded inventing the current system a hundred years ago.
3.2.7. NDL could provide lifetime information access, especially in rural communities.
3.2.8. NDL enables more frequent update of reference material.
3.3. Ideas on educational impact at the University level.
3.3.1. Many of the ideas under K-12 (above) apply here as well.
3.3.2. Humanists need access to special collections which frequently require travel. Could the NDL reduce the need for travel? Could it also promote more collaborative work in the humanities? Collaborative work is not as common as, for example in science, but there is increasing discussion in the humanities community that more should be promoted. This is the theme of a major humanities conference at the University of Michigan in April 93. Note too that Bill Wulf, former NSF/CISE Associate Director, is now involved in a "computing infrastructure for the humanities" initiative at UVA.
3.3.3. Could electronic publishing through a NDL help reduce the print publications back log problem?
3.3.4. Could it help bring intellectual control back within the academy?
3.3.5. In making the case for NDL application to higher education, we should probably distinguish between research and pedagogy. Graduate education will in general benefit from initiatives which support research productivity. The more difficult challenge might be to relate it to undergraduate education.
3.3.6. NDL could provide the library resources for "distance learning." Many distance learning environments now make little or no provisions for use of the host school's library. This is one reason that being a student at a distance may not be the same as being on campus.
3.3.7. NDL and collaboratory capabilities help accelerate linkages between research and education. The Arizona Collaboratory Project (Schatz, et al.) is illustrating this phenomena.
3.4. The NDL might be sold in part to help "people at risk."
3.4.1. But they must have access to the technology for the products of the NDL

to have impact.

3.4.2. NDL could help level the access playing field. It might help leverage local investments through new forms of sharing.

3.5. An NDL might be sold (or not be sold depending on current Washington policy) as creating entrepreneurial business opportunities in educational markets.

3.6. An NDL might be sold in part to help support national competitiveness in math, sci-tech, and team-based creative activities.

3.6.1. But note that foreign countries that are doing better than the U.S. in K-12 math education are not doing it with information technology.

3.6.2. Perhaps with technology, both the number of students who are successful as well as the level of their success could go up, i.e., an enhancement in both depth and breadth.

3.6.3. Can the NDL be positioned as a critical support for education in the new information-centered industrial era?

3.7. Better interfaces to information (applicable to all of the above).

3.7.1. There is some evidence that interfaces to digital information can be built which are more productive for humans in some information seeking/use tasks than the traditional print interface. Some of the work by Mike Lesk, et al.. has shown this. Other examples are the work of Robinson and Card at PARC.

3.7.2. Can electronic publishers come up with better learning interfaces than printed pages?

3.7.3. On the negative side, there will be learning curves for new access technology and the inexperienced users might be less productive with the technology "assist" than without it.

3.7.4. Need to think about the NDL technology as helping not so much to find information, but rather to help better manage and use the fire hose of information that is directed at most of us. Use metaphors such as "harvesting" information that we need or want as individuals and as member of groups/organizations.

4. Possible major themes or threads

4.1. Most of our population has no support for learning in the home. The NDL could be a chance to link the home with the school learning environment.

4.2. Perhaps the NDL could help re-create the value system that we had when the community took responsibility for the education of its youth.

4.3. The NDL could enhance the possibility for the community and family to actively participate in education of the community.

5. Other comments and observations

5.1. There appears to be a growing willingness (necessity based on self-interest)

for industry to make investments in education
5.1.1. Adoption programs
5.2. It might be useful to look at the issue of NDL to support education along the following three dimensions:
5.2.1. New media
5.2.2. New practice
5.2.3. New types of communities
5.3. We need more participation from the electronic publishing community.

Chapter 3, Section F: Borgman Notes

NSF Workshop on Digital Libraries
Xerox Palo Alto Research Center (PARC)
December 9-10, 1992

Notes by Christine Borgman

These notes cover the general sessions, with summaries from small group presentations. They are fairly comprehensive but not complete. All omissions and interpretations are the responsibility of the notetaker. Interpretative comments are marked with initials (CB).

December 9, 1992 Wednesday

9:00 Welcome by ***David Levy***

9:15 ***Mark Weiser***, Welcome to PARC

Brief summary of work at PARC related to digital libraries.

He is head of computer science lab, former professor and entrepreneur.

His interests are in social science support for computer science, i.e., a different kind of research:

reverse technology transfer -- from industry to university deploy and use discoveries

PARC: 20 yrs of pioneering basic research

individual change, innovation continues

Reinforce Xerox commitment to long range research

labs ranging from physics to anthropology, all working together

Digital libraries as a nexus for the future of research

Four dimensions: technical + social + educational + political

technical: database, filtering, distribution

social: collaboration, help, reference, editing, communities, psychology, linguistics
education, teaching, browsing

political: copyrighting, intellectual property, international data flow, competitiveness

domains of research at PARC:
forms
conversion
content

Form domain
Cornell -- CLASS and bookstore
highspeed portable print service
work practice
scanning and printing hardware

Conversion domain
OCR
understanding
image processing
System 33 -- conversion among multiple forms

Content domain
information access
information visualizer
tapestry
liveboard
ubicomp -- ubiquitous computing

information access architecture
text database, various retrieval techniques

ubiquitous computing:
situated science: philosophy, psychology, anthropology
poverty of existing computer systems: heavy, GUI, intrusive, VR, agents, multi-media
technology trends: processing, display

goals of ubicomp:
invisible technology, integration of virtual and physical worlds, through desk, rooms, buildings, life

enhanced ability to act

using digital information should be as pleasant, refreshing, enjoyable as using a community library today.

YT Chien, NSF: Background on National Digital Libraries Initiative

Overview of IRIS supported research

HPCC -- support for educ and technology

Jane Caviness (next speaker) is in charge of NREN implementation for NSF

While digital libraries are part of NREN scope, no specific funds have been allocated for it. We need a next wave initiative to take advantage of the NREN infrastructure to develop digital libraries.

Gore II bill -- addresses specific work to be done. Digital libraries is one of most prominent activities in that bill. (CB note for record -- Gore II not passed; expect it to be resubmitted in some form in the 1992-93 Congress. It will have support in the

White House.)

The July Digital Libraries Workshop was for academic researchers. This one is for wider participation, especially participants from industry. Consider user and business point of view.

Goals of workshop:
Bring together people with expertise to help design blueprint for digital libraries. Get the communities involved. Design programmatic activities that have the endorsement and support of the community.

Digital libraries of the future

This term is used to denote a number of related concepts:

a knowledge network -- initially based on the NREN, which is expected to be in place in mid or late 1990s.

an enabling technology -- for the capture, store, distribution, connection, and access of a variety of information sources, materials, and tools: text, image, sound database, software, and scientific instruments

new information services -- packaged in unconventional ways beyond traditional library media and environments

public, personal, and ubiquitous

Technical issues that arise in developing electronic libraries from Computing the Future NRC report, 7/92, distributed to participants of this meeting.

database
information retrieval
pattern recognition
human computer interaction
data structures
algorithm engineering
data compression
distributed systems
parallel computation

linguistics
reliability
file system and management
networking
storage architectures

digital libraries
users view
technology
business view

Jane Caviness -- NSF

One of the reasons for being here is the need for such activities as the digital libraries initiative. NREN and digital libraries are complementary activities.

Focus of development in her group at NSF:
development of gigabit net with DARPA -- research
infrastructure -- NSFnet as interim NREN

Status report on NREN/NSFnet:

NSFnet -- backbone of 45 Megabits, 45 nodes(?), 6-700 major universities plus 1000 research labs

connections to 50 countries
started at 56 kbits, linking super computers

The network already is far beyond computational use -- need these many other applications. Even the rudimentary use of library catalogs on the net today have brought in multiple disciplines and groups.

155 Mbits next stage, next contract
already having a billion packet day on the net!

Building gigabit network testbeds; 5 testbeds:

AURORA, BLANCA, CASA, NECTAR, VISTANET
3 year research project

co-funding by NSF and DARPA
participants contribute from industry, university and national labs, super computer centers

different focus in each testbed

heavy industrial and academic participation -- she showed useful map of testbeds

showed applications list. Includes multimedia digital library applications

Moving to higher speeds very rapidly.

Knowbot project with CNRI is part of this effort.
We are trying to encourage library participation, but we are not funded to support it in our initiative.

We would like to see the network live up to its potential.

Gio Wiederhold -- DARPA perspective
I3 -- intelligent integration of information : an architecture for using the global information network

Objective of DARPA for information networks:

get useful information to
a user
in the appropriate form
amount and level of detail
at the right time

We want to exploit the many data and computing sources available and emerging.

data, data everywhere, and nothing helps you think. The ancient planner.

databases
experience and knowledge bases
simulations

They are working heavily in simulation technology.

Banking is still running at 55 baud!! on large pipelines, for reliability. Established standard around the world. (CB: Is this an example of setting standards too soon?)

Emerging technology advances needed: showed detailed list.

List includes
information retrieval,
description,
standards,
knowledge sharing standards,
data abstractions,
formal models.

They don't want to build one right system today that will be a great weight later.

Example of elaborate military inventory control system. Integrated system for 150 databases and applications around the world -- cannot change any part, because cannot know the effect on the rest.

Client server model: want intelligence in local system, not mainframe core -- too hard to adapt.

standards and scaling.
huge range of standards needed at different levels. data, telecomm, etc.

DARPA is not trying to set standards and improve them; rather to support testing and validation.

ex. KGML -- knowledge query and manipulation language

superset of SQL -- highly structured. (CB: based on example given, this interface is not for novice users).

new section documentation
syntax in BNF
semantics
semantic model
message ontology

Applications that should change:

electronic libraries with journal articles which permit replication of experiments. Where wet experiments are needed, interface to the experiments' equipment must be provided (doing this in genome data)

a design and utilization record for devices which can be effectively summarized.

an automated individual medical record, machine processable, relevant history of patient complaint, at an appropriate level for the care provider (physician or nurse)

materials database which can be plugged into design environment
current resources sparse
text which permits simulation of all examples
ditto for assignments given to students

Michael Lesk -- Bellcore

We have a shared vision -- people can sit at screens, call things up for use.

The vision works -- people can do it. CORE experiments at Cornell -- students with electronic access do much better than on equivalent paper tools.

People really do use the library resources on the net.
computer manuals are on-line or CDROM -- rarely get on paper now.
get software via ftp sources.

Right now the US dominates in these areas (computing and information services delivery). We have huge economies of scale -- we have control so far.
People in Europe go to US to get information and software. We'd like to keep it that way.

We don't dominate information services such as Minitel -- France has taken the lead.
France is close to ISDN, UK putting it in next year. Low penetration here in US.

Image areas -- Europe is taking off very quickly. Scanning books, European consortium for medical literature.
These need to be large scale efforts to happen.

Videotext -- failed in UK as customer supported technology.

UUCP -- Lesk is one of the inventors. This service/network has really taken off.

Who are the largest publishers in the world? -- 2 Japanese printing companies -- $6 billion companies. Printing everything from books to circuit boards and wallpaper. Europeans next tier of very large publishers.

Who cares if it's done in the US or not? WE do -- important strategic business. Good growth here, but much bigger in Europe and Japan

It wouldn't do any good to put US data on-line because others could read it too. Response of one reviewer of Lesks proposal (CB: This is the isolationist side of the competitiveness argument; discussed again on Dec 10)

We are still largest producer of scientific information.

The real issue is scientific education -- if were going to continue to be the best educators, we need to maintain control of scientific information.

Others will package our data and sell it back to us -- this is happening already.

what matters is:
putting the infrastructure in place
focusing on specific applications
economics and politics -- publishing and tenure.

summarized earlier proposal to Wash DC Digital Libraries workshop (July 92)
Get basic undergraduate science information on-line first -- get enough for everybody to use.
suggested $50 million over 5 years
1/2 for infrastructure, 1/2 for related research

get 5 or 10 areas, cluster of publisher, user, technology community

coordinating committees to keep things on track.

Proposal was presented to NSF advisory board in Dec 91 by ***Michael McGill***, where it was received favorably. This is why were doing more workshops.

The information business will have a few really big players. We really need to develop those businesses before we lose it to the Europeans.

So far we have the basic infrastructure, networks, information in place. We need ways to get the data available for lots of people to use. This is strong motivation for building a national digital/electronic library.

Stan Besen - Charles River Assoc.
Intellectual property issues.

Focus of talk: how does moving from print to electronic affect the infrastructure of the publishing industry?

Where would parts of the present industry be redistributed?

He is not focusing on the law of intellectual property. Rather when one goes to electronic form, can rights be appropriated? Can some or all of the rights be taken across the boundaries?

Scenarios:
Large scale interlibrary loan -- make copies available to everyone else. Only one copy of a journal would be sold.

Pirated copies of tapes, videos, are sold overseas.

Right now you find the guy and sue him -- threat is dis-incentive to violate intellectual property laws. Detection is hard to avoid.

In electronic environment it is harder to detect who is doing copying and distribution; easy to borrow copies of tapes, software, and hard to detect or trace.

It is hard to determine economic impact on the rights holder.
Part of the issue is how good is the copy as a replication -- it may be quite good, in the electronic case; i.e., a good substitute for the original. This is the problem for rights holders. What is the original in the electronic world?

The price of the original has to go down to the price of the copy, or it will be difficult to sell the product; would only get one sale. Result is a very messy economic model. The private copying problem is getting worse.

Derivative works -- e.g., novel and movie, foreign language translations; character becomes doll, toy. Rights holders make money from their work in a variety of ways. All revenue streams belong to originators in present scheme.

Court ruling: Someone purchased a book, sliced it up, and framed pictures to resell. The publisher sued over control of image and won.

When someone takes and manipulates another person's work electronically it will be much harder to detect the original source. Thus it is much harder to plan the revenue stream of derivative works.

How are firms going to compete? It is not clear. How to control intellectual property? They will try to do vertical integration and not put out lots of raw materials that might be used for other purposes.

We should have less restrictive rights, which lower the cost of this industry, and make more information available to people. Could have effect on the supply of works. Set of adaptations need cannot yet be anticipated.

Designation of breakout groups: ***Levy***

Day 1 Groups:

1. What is a digital library? Leader: ***Borgman***
Determine functions, dimensions. What are we trying to do?

2. What technology requirements must be met for a digital library? Leader: ***Lesk***
Topics: infrastructure, research

3. What are the economic, social, and political issues in the digital library? Leader: ***Besen***

Day 2 Groups:

Grand challenge applications from the point of view of

1. business
2. education
3. research

Charge to the groups:

1. Select one person to represent the group. Come back with a 15 minute presentation. Address these questions:

2. How is the digital library going to happen? What are the infrastructure issues? How are we going to build it, what has to be in it?

3. How much? Intellectual property questions. Industry structure. Political economy questions.

Group reports:

Group 1: What is a digital library? Leader: ***Borgman***

[Editor's note: see charts at end of this section.]

Summary from flip charts:

For the national initiative to develop digital libraries, the definition of a digital library is a service that enables users to access the information they need from any place at any time

an environment for creation, dissemination, and use of information

an archive for preservation over time -- stable versions of documents

that integrates data in multiple formats

that will select, acquire, manage, and organize information

a set of user friendly tools to locate, retrieve, and utilization the information resources

a distributed technology environment that dramatically reduces barriers...

Question from ***Michael Lesk***: what's in, what's out? (Definition is too general to be usefull)

CB response: This is intentionally a very inclusive definition. It would start with science and technology and move outward, but does not have a specific content as a stated scope. The breadth is deliberate -- YT Chien of NSF wants to be able to sell it to other federal agencies in multiple disciplines (medicine, engineering, etc.) and get their support and contribution to development. It is a service, not a place or a product.

Group 2: What technology requirements must be met for a digital library? Leader: ***Lesk***

[Editor's note: see charts at end of this section.]

Issues: collection/distribution/reception

collection: OCR/scanning. conversion. processable representations. develop collection. shared reference resources.

distribution: NREN, ubiquitous, distributed retrieval, natural language processing, tools to support indexing, summarization, classification; filtering; multimedia

computational requirements
security, privacy, authentication, protection, accounting
standards with extensibility

Technology for organization of database not included here; assumed it will be included in group 1 scope.

Group 3. What are the economic, social, and political issues in the digital library? Leader: ***Besen***

[Editor's note: see charts in the next session.]

This group issued no shoulds

Forecasting future of digital libraries is difficult.

The diversity of publications exacerbates problem. More research is needed.

Appropriability vs. free access: new mechanisms for appropriability in a digital environment

What will be the significance of gatekeeper function? We need to reward gatekeepers (CB: Besen is referring to publishers as gatekeepers, which is a much different definition than in scholarly communication research.)

How will the gatekeeper function change in a digital environment?

We expect changes in the number, identity, roles of participants in publishing.

Some interaction between intellectual property law and industry structure is expected.

What will be the political economy of introducing digital libraries?

How are digital publications different from other types of intellectual property that also have appropriability problems?

What is the role of government?
research and development
support for infrastructure
support for producers
support for users

Home delivery of information -- how can it piggyback on extant infrastructure?

December 10, 1992 Thursday

Day 2. Breakout into 3 groups; meet 2 hours and report back.

Reports of 3 groups:

What are the benefits of a national digital library to:

1. Research community (leader: ***Christine Borgman***)
2. Education community (leader: ***Daniel Atkins***)

3. Business community (leader: ***Michael Lesk***)

Reports:

Group 1. Research community. (***Borgman***)

[Editor's note: see charts in the next session.]

Group 2. Education community. (***Atkins***)

[Editor's note: see charts in the next session.]

Very detailed presentation with many flip charts; this is incomplete.

3 dimensions:
functions of digital libraries
constituencies K-12, college, research, continual life learning
subject domain

An infrastructure that supported new forms of dynamic media might be strongest case argument.

Communication processes are a very important part of the National Digital Library.

We must be careful not to suggest that high technology will supplant low technology -- still need paper and crayons.

Experiential learning still important -- chemistry labs and smelling -- olfactory senses.

There is the potential for customization to different rates and modes of learning.

Consider value of pulling things out that can be shared -- clearinghouse dimensions.

School libraries have minimal resources, as do public libraries -- a digital library could lead to expansion of virtual holdings.

The traditional role of a public library; the right to know, right to acquire information is a fundamental social right (in the US, anyway). We are reinventing the public library for the next century.

There is a growing demand for environmental, geographic information systems.

In the humanities, people spend much time to travel to resources. Some of this travel could be overcome.

Potential for personalization, customization of information resources.

Issue of selling our intellectual property offshore and buying it back -- can we recover some of this? (CB - discussion ensued later, part of isolationist issues.)

Graduate education and research are tightly coupled.

People at risk in not having adequate information access, particularly in K-12 (K-9?) level.

There is a tension between local control (textbooks, etc.) and remote access to materials.

Bootstrapping effect. Existence of digital library may make it easier to create materials for classroom use.

If we want to produce expert learners, we need to expose them at an early age to the resources that will be available to them later in life.

Consider a law school model of creating expert learners. Lexis and Westlaw can be seen as digital libraries. They provide the systems to law schools -- could we do the same for schools?

Teleconferencing technology is popular for education, but one of the major problems is that remote students do not have access to library materials; this might help compensate.

We need new types of interfaces -- better, more efficient access to resources.

Also consider health care as a national issue, what is relation to digital libraries?

Library useful for team based activities, competitiveness.

Students in low resource countries such as China are ahead of our students in science, which may make it difficult to make the argument that better information resources will improve our education system.

Could digital library increase the linkages between home and school?

Group 3. Business community. (***Lesk***)

[Editor's note: see charts in the next session.]

Three main issues for business community:

1. user impacts
2. producer impacts
3. social impacts

Assumption: the national digital library is good for business if it leads to a well educated workforce and higher productivity.

User benefits

Saving redundant work; processable and refinable information can reduce overhead.

Equal information access contributes to a level playing field
small/large/global business
businesses in small towns/large cities
isolated technology areas

Ways to improve productivity
superconductivity, xerography were discovered by random browsing
faster design making, shorter product cycles

industrial equivalent --as National Library of Medicine is to health

Example of identifying suppliers when technology needed quickly for Desert Storm. Who made IFF transistors? Identified a source, placed a contract in 2 days and got friend or foe detectors in 10 days.

Standardization, interagency agreements -- if we could get information out easily and quickly, it would increase use and dissemination (technology transfer).

We need rapid access to commercial resources in national emergency -- hurricane, need water purification information, etc.

We need access to grey literature -- government documents, etc. need better access and coverage.

a couple of months in the lab will save you a couple of hours in the library.

Other issues for business community:

advertizing, alerting, problem identification

technology transfer, both to and from universities and research

enable libraries to cope with greater volumes of materials

patent information

Some scenarios:

Blacksburg/Penn studies -- sharing correlation of information availability and economic growth

Swanson studies on medical searching

Productivity of biotechnology and their use of information -- new field, everything on-line.

software vendors -- for agents programs which display, search, filtering, abstracting

communications: services, telecommunications, etc.
equipment -- modems, portable devices

information vendors for other information sharing platforms
federal regulations,
materials information
product information
zoning building codes
new or substitute data

hardware: printers, displays
internal information system development

publishers alternative outlet for publication.

General discussion (after group reports)

social benefits of the national digital library:

small business opportunities as intermediaries, vendors, etc.
enable work at home
media changes for disabled
support democracy -- transfer Ben Franklin's argument for the post office to the NREN today
promote joint ventures -- anything that makes it easier for people to find each other will encourage joint ventures
business will locate where information is
easy to get
easy to distribute

Stated as nasty question: why won't benefits just accrue to foreign countries? (isolationist argument). Audience response:

1. Most of our information will be in English,which means more benefits to us.

2. Others already building national digital libraries and we need to trade with them.

3. If we don't trade, they are taking out information and selling it back to us.

4. Our economic survival is based on infrastructure and workforce -- we need it first to have the competitive advantage.

5. Same as general issues of international trade -- we need to have something to trade!

Other discussion point -- encouraging our staff to be multidisciplinary -- need multi disciplinary access

Closing session

YT Chien, NSF. Summary of what we've accomplished

Thanks to everyone, especially ***David Levy*** for making workshop happen.

Thanks also to Larry Masinter, Guiliana Lavendel, Geoff Nunberg.

NSF now has enough information that staff can assemble.

Our goal is to invent the future and what better place to do it than here at PARC, where Alan Kay started?!

What is the future of the national digital library?

Are we in the information business or the library business?

What economic model do we employ?
Free to the end user?
Bookstore?
Information kiosk?

How do you draw lines between who has free access and who does not?

One of the great future issues will be the models of who pays and how much and in what levels, etc.

Internal contradictions
free public and academic libraries
corporate libraries are willing to pay for Dialog, Lexis, etc.

Information delivery services are getting popular among academics.

People will pay for freelance services that are not called libraries.

We must not fail to resolve conflicts among these competing information sources.

NSF will distribute a resource book to all participants of the workshop.

December 9, 1992 - What is a Digital Library?

NSF workshop on National Digital Libraries
Breakout session on What is a national digital library?

Leader: Christine Borgman
Notetaker (these notes) Terry Winograd
Overheads for presentation are a subset of these notes

[Editor's Note: Formatting information has been partially lost for this breakout session report.]

1 Official notes on "breakout" session
2 What is a doable image of an electronic library that can be sold to Congress?
3 Gore bill says:
3.1 development of advanced storage
3.2 thousands of users
3.3 instantaneous
3.4 conversion of print to electronic
3.5 data base software for filtering and summarizing
3.6 standards for electronic data
3.7 training of users and librarians
3.8 simplify utilization
3.9 visualization technology for large volumes
4 Prototypes
4.1 digital libraries of scientific data
5 To be seductive to Congress as something worth doing it has to be a radical departure from what we have now
6 Break boundaries of traditional view of library
6.1 information is continuous
6.2 everything from classroom to homework to scholarship
7 Vision statement should be broader than the experiment
7.1 Talk about what it could be and have an experiment that fits in with today's structures and technologies
7.2 vision could be threatening to existing parties (e.g., publishers)
7.3 don't couch the vision as an attempt to put them out of business
8 A distributed technology environment which dramatically reduces barriers to the creation, dissemination, manipulation, storage, integration and reuse of information by individuals and groups
9 Any time, from any place, having access to the world's knowledge

9.1 reduce constraints of time and distance
9.2 all media
10 across communities
10.1 potential to bring connections from researchers to teachers to schools...
10.2 community of scholars can communicate their works and enter into dis course about it
10.3 collaborative group use is a key advantage of digital technology
10.4 Must be desirable because of what it does
10.5 leverage through digital replicability
11 A national initiative to develop digital libraries
12 Would include one or more projects, one or more libraries
12.1 includes research and specific experiments
12.2 continuum between different styles of information
12.3 bridging between roles of consumption and production
13 Testbed on a specific topic, e.g.,
13.1 Horizontal: Undergraduate science education (earlier proposal)
14 Vertical: e.g., chemistry from kindergarten to Nobel Laureate
15 allows synergy between different levels / teaching and research
15.1 as in the best institutions
15.2 Information for decision makers (e.g., Congress)
16 multiple levels of capacity
16.1 something completely portable
16.2 higher capacity available with some local concentration (e.g., large view, high resolution)
16.3 servers providing high-cost capacities
17 Capacity for archiving / preserving over time
17.1 long term preservation over technology and economics change
18 repository for guaranteed referentiality
18.1 stable versions of documents
18.2 Generally, widely and easily usable set of tools and capabilities to locate, retrieve, and utilize the information resources available.
19 Electronic vs. paper: electronic is better when materials are
19.1 highly volatile, e.g., air schedules
19.2 manipulation desired
19.3 scanning for names or words or phrases in lengthy texts
19.4 light use of remote materials
19.5 rapid communication in dispersed group
20 References
20.1 Redesigning Libary Services: A Manifesto, Michael Buckland
20.2 Rogers and Hurt, in Chronicle of Higher Ed on moving libraries to

community of scholars
21 Silicon Dreams, Bob Lucky
22 Digital information is fundamentally different
22.1 copiable, processable, rapid transmission,....
23 Different kinds of finding activities
24 Navigation in the “info space”
24.1 defuzzing and finding
24.2 negotiating the query
25 Access
25.1 How do I get it - obtaining access to the material, fact, or object
26 Serendipity
27 Browsing
27.1 encountering display
27.2 table of contents, distribution lists
27.3 general collections of abstracts
27.4 environmental scanning
27.5 network surfing
27.6 within structure of a conscious selection process
27.7 possibility of positive surprise
27.8 current awareness browsing
28 Reuse
28.1 Archival access: getting it over again
29 Putting it into something else
29.1 idea vs. expression
30 Participants
30.1 Chris Borgman
30.2 Jean Mayhew
30.3 Eugenie Prime
30.4 Vicky Reich
30.5 Tony Hall
30.6 Terry Winograd
30.7 Daniel Atkins
30.8 Maria Zemankova

31 What is it?
S = structure
B = behavior / function
31.1 S audience - education
31.2 S materials - science
31.3 S platform - NREN
32 S/B interface

32.1 agents
33 B economic model
33.1 B fee for service vs. tax/subsidy
33.2 S/B where is the librarian?
33.3 B structure for authoring / information generation
33.4 B scholarly communication process
34 B anticipate user needs / active vs. passive
34.1 B "don't tell me what I want to know, tell me what I should know"
34.2 S functions

Copies of Flip Charts

LEADER: CHRISTINE BORGMAN

BENEFITS #1: RESEARCH

1. LIBERMAN
2. BESEN
3. ZEMANKOVA
4. WINOGARD
5. MASINTER
6. LUCIER

LEADER: MICHAEL LESK

BENEFITS #3: BUSINESS

1. JAJKO
2. DYSON
3. ERTEL
4. PRIME
5. MAYHEW
6. GASS
7. WIEDERHOLD
8. LAVENDEL

LEADER: DAN ATKINS

BENEFITS #3: EDUCATION

1. CHEN
2. CAVINESS
3. ROSENBERG
4. MENELL
5. NUMBERG
6. REICH
7. CROCCA
8. URBANOWSKI

COLLECTION - OCR/SCANNING - CONVERSION
DISTRIBUTION - NREN - NII (NATIONAL INFO. INFRASTRUCTURE)
RECEPTION - USER INTERFACE (GRAPHIC)
UBIQUITOUS REPRESENTATION
ACCESS TO TRAINING
PROCESSABLE REPRESENTATION
PRINTING DISTRIBUTED RETRIEVAL
INTEGRATION DEVELOP COLLECTIONS
NL CONFERENCE
SHARED REFERENCE RESOURCES
ENTERTAINMENT
TOOLS TO SUPPORT ACCESSIBILITY
INDEXING
SUMMARIZATION DOMAINS
USER CLASSIFICATION, BACKGROUND
FILTERING
MULTIMEDIA RETRIEVAL
COMPUTATIONAL REQUIREMENTS
SECURITY/PRIVACY/AUTHENTICATION/PROTECTION/ACCOUNTING
STANDARDS WITH EXTENSIBILITY

LEADER: STAN BESEN
GROUP #3 ECONOMIC, SOCIAL, POLITICAL ISSUES

1. DYSON
2. LUCIER
3. MENELL
4. URBANOWSKI
5. LIBERMAN
6. GASS
7. ERTEL
8. JAJKO
9. ROSENBURG
10. NUNBERG

LEADER: MICHAEL LESK

GROUP #2 TECHNICAL REQUIREMENTS FOR A DIGITAL LIBRARY?

1. WIEDERHOLD
2. CAVINESS
3. CHEN
4. FINNIGAN
5. MASINTER
6. GIULIANA

LEADER: CHRISTINE BORGMAN

BREAK OUT GROUP #1 WHAT IS A DIGITAL LIBRARY?

1. BORGMAN
2. MAYHEW
3. PRIME
4. ZEMANKOVA
5. WINOGRAD
6. ATKINS
7. REICH

WHY

- LEVERAGE TECHNOLOGY TO IMPROVE RESEARCH & EDUCATION

- MULTIDISCIPLINARY SYNERGY/COORDINATION
(THAT WON'T HAPPEN OTHERWISE WITHIN NSF DISCIPLINES)

- CREATE/ENABLE NEW INDUSTRIES

- DRIVER APPLICATION FOR NREN

- WIDER ACCESS/AVAILABILITY

- MORE COST EFFECTIVE

WHAT DOES THE INITIATIVE HAVE TO CONTAIN FOR YOU TO BUY INTO IT

I'M ALREADY IN!

RESEARCH INITIATIVES IN OPEN, DISTRIBUTED, EXTENSIBLE HYPER-BASE
SYSTEMS CLEARLY ADVANTAGEOUS FOR THE NATION'S EDUCATION (K-12 THROUGH POST DOC)

OPEN MINIMUM BARRIERS TO ENTRY

WHAT IS NEEDED TO ACCOMPLISH THIS

CRITICAL MASS OF RELEVANT INFORMATION
- ADD VALUE BY GOING BEYOND ELECTRONIC FORM OF EXISTING ITEMS
- CREATE NEW INTERACTIVE INFORMATION BASE

INFORMATION PEOPLE WANT AND NEED TO DO EXISTING WORD.

- $50m
- COOPERATION OF DIVERSE RESEARCH GROUPS
- COOPERATION OF CONTENT PROVIDERS
- INSTITUTIONAL BUY-IN (UNIVERSITIES)

- NEED TO DEFINE ROLE FOR TRADTIONAL LIBRARIES & PUBLISHERS
- ACCOMODATE LEGAL (COPYRIGHT) + BILLING CONCERNS
- CONSORTIUM BUILDING
- EVALUATION
- CONFERENCES / WORKSHOPS

WHY WILL IT BE GOOD FOR THE U.S.

PROVIDE FOR CROSS-FERTILIZATION OF IDEAS
- - APPLE PIE: EDUCATION, RESEARCH,...
- - SO WE WON'T BE A 3RD WORLD COUNTRY BY 2001
- - BECAUSE U.S. EDUCATION COULD STAND A LIFT!

MOON SHOT
MAINTAIN LEAD IN SCIENCE/TECHNOLOGY/ENGINEERING PUBLISHING

WHAT SHOULD BE IN THE NAT'L INITIATIVE

5 PARTS

(1) DISTRIBUTED/EXTENSIBLE/HETEROGENOUS/DATA + METADATA
DEFINE AN OPEN TECHNICAL/ARCHITECTURAL/STANDARDS FRAMEWORK TO CONTAIN THE MODULAR COMPONENTS OF AN ELECTRONIC LIBRARY
INFO SERVERS, REFINERS/BROKERS & CLIENTS - THAT FORMS AN OPEN PLAYING FIELD.

(2) FUND DEVELOPMENT OF SPECIFIC INSTANCES OF EITHER RESEARCH OF PRODUCTION LEVEL INSTANCES OF SERVERS, CLIENTS, ETC. THAT FILL SPECIFIC NEEDS OR INVESTIGATE/
LEVERAGE RESEARCH ACTIVITIES - BOTH NOW AND LONG TERM

(3) FUND LONGER - TERM RESEARCH ON SCALE UP, NAVIGATION, USER STUDIES,...
(4) DEVELOP PEER REVIEW MECHANISMS FOR VARYING COMPONENTS
- EVALUATION

(5) PROVIDE BACKBONE NETWORK INFRASTRUCURE THAT CAN SCALE TO MILLIONS OF USERS AND TERABYTES OF DATA IN PRODUCTION QUALITY, SECURE ENVIRONMENT.

HOW SHOULD IT BE COORDINATED NOW

TECHNICAL, ADMINISTRATIVE

USE E-LIB AS COORDINATION TOOL

PANEL AT NSF FOR NEW INITIATIVE IN IRIS

(1) MUST TAKE A VERY MULTIDISCIPLINARY APPROACH, AVOID PAROCHIAL INTERESTS OF NARROW RESEARCH COMMUNITIES

(2) SPONSOR CONFERENCES

(3) HAVE REPS OF: FEDERAL AGENCIES, ASSOCIATIONS

INVOLVE END USERS PROFESSIONAL SOCIETIES

NIST

RESEARCHERS

NEL Workshop Report

Christine Borgman
September 4, 1992

Response to workshop outcomes, as presented by Edward Fox on 21 July 1992.

Background:

A workshop was held in Chicago, 10-11 October, 1991, on Future Directions in Text Analysis, Retrieval and Understanding, in which I participated. Part II of the Final Draft Report (18 July 1992) of that workshop is a proposal for “A National

Electronic Science, Engineering, and Technology Library" (Lesk, Fox, & McGill). The proposal formed the basis of discussion for the 20-21 July 1992 workshop at NSF on developing a multi-agency initiative for a "National Electronic Library." Participants in the latter workshop were asked to respond to a number of issues raised in this report and in two days of discussion. This short paper is my response, organized by the suggested categories.

The above-mentioned reports do not contain an explicit, succinct definition of the term "National Electronic Library." For the purposes of discussion here, I am using the following definition, adapted from the workshop summary discussion:

The proposed National Electronic Library is

A service

An architecture that includes
—network requirements and expectations
—formal and de facto standards where they exist
—requirements where standards do not exist
—open framework for system component development

A set of information resources, databases of text, numbers, graphics, sound, video, etc.

A set of tools and capabilities to locate, retrieve, and utilize the information resources available

Users of the National Electronic Library include:
—students
—teachers/professors
—researchers/scholars
—librarians
—authors
—publishers
—information providers
—practitioners

Contributors of information resources to the National Electronic Library include:

—publishers
—universities
—professional societies
—libraries
—authors
—editors
—compilers

Responses to specific workshop questions

What does the National Electronic Library look like to different categories of users?

The National Electronic Library (NEL) initiative as outlined in the report of 7/92 would begin by building a set of information resources aimed at undergraduate level science, engineering, and technology, and would support related research in information technology, standards, and information behavior. The NEL would exist on the NREN. Research and development would continue to expand, refine, and support the services and system. It is assumed that most users of the NEL also would be users of other information network services such as those provided on the Internet: electronic mail, file transfer, and remote login. The following sketches are the images the network might offer to each of these groups after the first 5 years of development.

Undergraduates
Undergraduate students in science, engineering, and technology at U.S. colleges and universities would access databases of bibliographic, text, numerical, and graphic material supporting their coursework via their campus networks. (The NEL should be available to those outside the U.S. as well, but we discuss domestic users here as our primary interest for the NEL initiative.)

Specific implementations would vary depending on resources of the institution and individual student from dialup access in the academic library to campus workstations scattered through classrooms, laboratories, libraries, offices, and dormitories, connecting resources on the campus backbone and the NREN. Students with personal computers would be able to dialup or otherwise connect into the network from home or away.

At the low end, undergraduate applications of the NEL would replicate current library use; e.g., searching indexing and abstracting databases to obtain information for a term paper, then seeking the full publication in print form from the library shelf or interlibrary loan. For all but the most resource-rich libraries, NEL access would be a considerable service expansion by providing access to databases not held locally, or providing new databases created for the NEL. The broad access to a common network would equalize information resources between richer and poorer institutions and could greatly increase access for the part-time or commuting student, by providing easy access from home or non-university workplace. As the educational marketplace shifts to the older, part-time student, distributed access to information becomes increasingly important to the educational process.

Local computing resources will grow along with the addition of NEL databases, allowing the NEL to be used for higher end applications, such as manipulating numerical and graphical databases of scientific information and for conducting empirical analyses of data collected by others, such as geophysical data, genome data, or historical trends in social or political data.

While the NEL, as described in the proposal document, would be of greatest initial value to undergraduate students majoring in science, engineering, and technology, it would also be of immediate value to other undergraduates in supporting their basic liberal arts courses in these areas. Conversely, science cannot be taught with scientific data alone. Science must be applied to application domains, such as the environment, which require information drawn from the range of human knowledge. As the scope of databases expand, so will the scope of undergraduate students served, directly and indirectly.

K-12 Students

An NEL initially targeted at undergraduates will have immediate value for the high school student. Advanced high school students begin studying undergraduate material, sometimes placing out of lower division courses.

High schools have far less computing infrastructure in place to take advantage of an NEL than do colleges and universities, and thus adoption will be slower. As the databases expand beyond primary science materials and interfaces are improved, usage will expand downward into lower grades. Early uses of the NEL for K-12 are more likely to be teacher-driven, using the resources for classroom demonstrations and in-class explorations, than student driven, as students may have insufficient access to computers.

In the long term, the NEL offers great potential for enhancing education in the full K-12 range. Our studies of the Science Library Catalog (discussed further below) show that elementary school children (aged 8-12) are able to search for science information in a browsing-oriented system that is tailored to their level of cognitive development, reading level, and interests. They like to use computers and they are interested in looking for information. Wide adoption at the elementary school will require more computing resources in the schools, NEL databases designed for this age group, and rich but simple browsing mechanisms for general science and other information.

Graduate students

Graduate students may make the greatest use of the NEL, even if the system is aimed initially at undergraduates. Graduate work is much more oriented toward individual research and information seeking than is undergraduate work. Graduate students are more likely to have their own personal computers and more apt to invest their own resources in information seeking than are undergraduates. They are older, often more affluent than undergraduates, and have the incentive to pursue individual research so they can complete their degrees in a timely manner. Undergraduates are more oriented toward pursuing specific assignments using specifically-assigned materials in a fixed curriculum of courses.

Graduate students are rewarded for seeking and identifying new information and new sources not only in support of their own coursework and independent research, but in support of their professors' research. Graduate research assistants often have the responsibility of literature searching and locating information resources (textual or otherwise) in support of research projects. Thus they are even more likely than scholars, in many cases, to take advantage of the exploratory capabilities of the NEL and NREN.

Researchers

Researchers can be expected to use the bibliographic and textual databases selected for undergraduates for their own research, to the extent that they represent the basic sources of their discipline. Non-textual databases of numerical and graphic data may be of considerable value to researchers, both for reference purposes (e.g., engineering handbook data) and for empirical research (e.g., geophysical data). Given that research and teaching are two sides of a coin, university researchers will be seeking databases that are valuable for their own research and that can be used to teach undergraduates and graduate students. Better yet, if the information resources support projects that can be assigned to students that will advance the researchers' own projects, they will be even more readily adopted. As the NEL grows to include more primary scientific data, it will be increasingly attractive to this group.

University researchers' access to the NEL will be dependent upon university resources. Researchers are more likely than students to have ready access, through computers in their offices, laboratories, and homes. Non-university-affiliated researchers (government laboratories, corporate researchers, consultants, etc.) will have needs similar to those of university researchers, though without the complementary support of teaching.

An important aspect of the current Internet for researchers that was not addressed in the workshop report is the ability to mount locally-produced databases for access by others. We can mount files of data that we want to share with others, including research data, preprint files, backfiles of electronic newsletters, bibliographies, and various other material. Mounting our own databases allows easy dissemination of research materials, with minimal overhead of copying and mailing papers and disks. Provision of research data should increase the dissemination of research results and thus technology transfer as well, spreading resources of NSF and other funding agencies even further.

The workshop report did address the need for intelligent agents to locate network resources, such as the Archie and gopher utilities already extant. As we continue to mount large, formally-produced databases along with the informal locally-produced databases, researchers and other users will find the line between material formats increasingly fuzzy. The mixing of resources also will put pressure on standards initiatives and enforcement on the network, with extensive need for open standards and searching mechanisms that can handle a range of formats.

Practitioners

Practitioners is taken to mean people who do not fall into the other categories. Depending on what they are practicing, their needs may be comparable to students or researchers. The differences may be in infrastructure. Students, faculty, and non-university researchers typically have access to an infrastructure that assists in training, documentation, selection of hardware and software, alerts to new resources, etc. Independent practitioners may rely more heavily on sources internal to the NEL and NREN for such information. A new market may develop for providers of information about the NEL for the non-affiliated practitioner.

Publishers

Publishers are a large group with unclear boundaries. Traditional publishers of trade materials and scholarly books and journals will see the NEL as a large library to whom they may sell their wares. They may use it as practitioners as well. Attention to standards will be an increasingly important responsibility of publishers, who will be supplying computer-readable data directly. Even with the extensive use of computer-readable text in publishing now, some publishers are using the computer-readable data only as an interim step to typesetting, with disks later destroyed. While this approach means printing quickly and attractively, it complicates publishing the text electronically later and avoids some of the need to adhere to standards.

Currently, some publishers are contracting with large universities to mount their databases on local systems. These systems may be accessible over the networks, but with access restrictions to certified users on a database-by-database basis. For example, one can sign onto MELVYL anonymously, without fee, to search the bibliographic database of University of California holdings. To switch to a MELVYL database supplied by a commercial publisher, such as the MEDLINE or Inspec files, one is prompted for a password that is available only to UC affiliates. This barrier limits the access to these databases to members of institutions that have paid directly for this resource, while access to other resources is paid indirectly, in terms of each institution offering up some of its own files and expecting others to do the same. The solutions to more open access are not clear; policy studies are needed to address the access and revenue tradeoffs such that information is affordable while providing sufficient revenue to keep publishers in the marketplace.

The definition of "publishing" is stretched further when researchers begin mounting their own databases, essentially publishing them outside of the usual formal mechanisms. When information bases are in the public domain, it appears to be more cost-effective to mount them and let others search them than to deal with requests individually. I have no data to support this claim, except for the observation that the number of mounted files appears to be increasing exponentially, and those mounting them must consider it cost effective or would not otherwise do so.

Information Providers

Information Providers appears to be a general term to include publishers, who make available the work of authors, editors, and compilers; and the authors, editors, and compilers who make direct contributions to the NEL. All will view the NEL as a place to contribute their work for others to read and use. The degree to which they view the NEL as a competing entity to other formats or as a new opportunity will depend on their current distribution and payment mechanisms, if any. The NEL will pose the greatest challenge to those with established distribution mechanisms, who must determine how this new channel complements prior channels.

At the same time, the low overhead of contributing information in machine-readable form compared to traditional publishing should create a whole new class of information providers who see this as a grand opportunity to be heard. As multimedia applications increase and the capabilities of the NEL increase, so will the number and types of information providers. The usual problems of copyright, payment, and access tradeoffs apply.

Authors (and editors and compilers)
Authors are assumed to be academic authors for the most part, as the initial NEL will be supplying information aimed at the undergraduate. University faculty have competing interests in electronic publishing between being authors and readers. As authors, we want control over our writing and payment for its use. As readers, we want ready access to the writings of others at minimal or no cost! Academic authors routinely assign copyright to publishers in return for getting their work published, and get royalties for books, though rarely for journal materials, complicating the roles further.

Development of the NEL will contribute to the dismantling of the formal scholarly communication infrastructure of authors, publishers, and libraries. Various schemes have been proposed to bypass publishers and make materials directly available over the information networks, such that authors claim royalties directly. Similarly, publishers have various schemes to bypass libraries and gain more control over dissemination. Technical and policy issues abound, which are discussed further in Borgman (1992).

Current libraries
Libraries already have expanded their services well beyond their walls, having used commercial information retrieval systems since their development in the early 70s (and the batch systems prior to that time), and more recently using the existing Internet services. Academic libraries would be the most directly and immediately involved with the NEL, given the emphasis on undergraduate education. They would use the NEL to expand their own services, providing access to the databases on the NREN by training their user community, providing hardware and software access, and incorporating references to NEL resources in their own catalogs. They would also consider NEL resources in their own collection development, shifting duplicate resources to other materials and services, extending already strapped budgets.

Corporate and special libraries would be immediate users of the NEL, just as they are heavy users of other electronic sources. Special libraries often are staffed by one full or part-time professional serving dozens of practitioners, whether corporate researchers, managers, or attorneys. They have minimal print collections and may do the majority of their information searching by accessing retrieval systems. When they need the full text of materials for which only bibliographic data is available on-line, they get it through reprint services and by sending messengers to area university and public libraries. Large special libraries, such as those of international firms, may be well-staffed and have extensive local collections, but they also rely heavily on information retrieval services, company-developed databases, and electronic networks of their own corporate libraries. They will see the NEL as an extension of other on-line retrieval services and will be more than willing to pay reasonable fees for usage.

Public and school libraries currently make the least use of automated retrieval services. They may not yet have automated their own internal operations and lack the computing resources to make external networks available to their patrons. The exception are the large central and regional public library collections, which provide information services that individual branches cannot supply. Many of these services are fee-based; businesses without their own libraries often rely on these services or those of university libraries. They also may hire information brokers to find information in support of larger projects.

Public and school libraries have the potential to be the critical point of access to the NEL, as the databases expand and the service broadens to more general access. Public libraries can provide the training and access infrastructure for the general public. School libraries can provide the training and access infrastructure for the schools, especially as school libraries and classrooms become networked. Given the current funding level for public education, such networking probably is much more than 5 years off, however.

Educators

Educators are by no means last in consideration, but are an umbrella group, as are information providers. These include K-12 teachers, college and university faculty, and practitioner-teachers, training in environments other than full-term courses. Teaching applications run the full gamut of users of the NEL, including searching for information for one's own edification, designing curricula, demonstrating scientific topics by manipulating scientific data on the NEL in the classroom in real time, or making teaching assignments.

Early science database applications in education include not only resources intended for student use, but resources for teacher use (Science CD-ROM Conference, 1992). These databases include lesson plans for a variety of science topics and laboratory experiments that generate inventory lists of instructional materials.

Learning science requires not only the subject content of science, but how to find out more information about the topic. The NEL can be extremely useful in this respect, allowing students to search for new information beyond what exists in their classroom and texts, and allowing the teacher to locate new information in response to student questions. The current educational structure, with its minimal library support and lack of teacher training in information seeking, does not support well the teacher in assisting students in continuing their explorations. The NEL could fill a critical gap here.

Educator uses will vary depending on resources available and on technical sophistication. With one workstation and a display unit, a teacher can search for information for class preparation and can conduct classroom demonstrations. With multiple workstations, students can search on their own. With multiple networked stations, students can conduct searches and projects jointly with other students in the school, the district, and elsewhere, and librarians and teachers can work closely in supporting classroom instruction.

Problems with the National Electronic Library:

The National Electronic Library is a powerful concept and one whose time has come. The potential payoff in terms of shared resources, expanded library and educational resources for the individual institution, and technology transfer is vast. The problems will lay in coordination, implementation, and policy.

One problem, though a relatively minor one in my mind, is the inappropriate use of NEL resources. Students, faculty, and other users of the NEL will use the databases available on the NEL, often adapting their questions to the information rather than vice versa. When we mounted the New York Times Information Bank at the University of Pittsburgh in 1973 for free access, all undergraduate term papers seemed to become NYT topics. When Medline was mounted on MELVYL for open access in the UC system, the university community made massive use of the database, far beyond the needs of the health sciences alone. This follows from the principle of least effort -- people use the information available, whether or not it is the “right” or most appropriate information. NEL services will compete with traditional library services, particularly on a self-service basis. The alternative to using the wrong information often is getting no information at all. If people are searching and are finding useful material, they will continue to search when it is appropriate, and will eventually become more sophisticated users.

We must provide the service, police it in a cost-effective manner, and accept some level of misuse. We cannot fail to provide an information service due to misuse, anymore than we can fail to provide city streets because they will be used by prostitutes and drunk drivers. Rules of good management apply here as elsewhere.

Effect of delay on the National Electronic Library:

The time for the NEL is NOW! We have the necessary building blocks in place for the NEL: the outline of the NREN, with the Internet and the NSF backbone; large numbers of valuable databases that are underutilized due to lack of access; basic information technology standards for telecommunications interconnection and file sharing; a rapidly expanding base of personal computers in the hands of institutions and users; basic understanding of computing networks by the tens of millions of electronic mail users. The user community on the Internet is expanding exponentially and they are becoming increasingly sophisticated in the use of electronic mail, file transfer, and remote login services. Hardware costs are plummeting, which will continue to expand the base of workstations with access to the information networks.

If we do not take advantage of the infrastructure we have built, others will. We are faced with increasing competition in the information technology and information services marketplace by other countries, as outlined in the workshop report. We can maintain our (shrinking) competitive lead by developing the NEL, which in turn will set the information technology standards in communications, information retrieval, and interface design, that others will follow. We can mount the databases and design a reasonable cost and access structure before others control so much of the information that we are priced out of the market. Or we can wait, let others set the standards and control the information, so we are priced out of the marketplace, and lose control over the technology so we buy it from others rather than they from us.

The sooner we implement the NEL, the sooner we can take advantage of shared information resources and supply needed information to our classrooms and researchers. The longer we wait, the more we will lose in both quality of education and international competitiveness.

Approaches to the National Electronic Library:

The approach to building the NEL outlined in the workshop report begins by constructing a basic core of databases aimed at the undergraduate, while funding supporting research on the NEL infrastructure. This seems a reasonable approach, as the size of the undergraduate population is known and manageable, and is in great need of expanded access to science, engineering, and technology information. Databases aimed at the undergraduate population should be of immediate value to the advanced high school population as well as to the graduate student and research population, thus providing a fairly broad base of support.

The proposed approach also recognizes the need for research in a variety of areas, from specific standards to interface design. It needs also to recognize the need for policy research to address the complex issues of the shifting scholarly communication infrastructure, including copyright and ownership issues.

Beginning with a few databases and then expanding is useful, though reliance on a few large projects carries risks of large failure and the added costs of coordination. An approach favored by at least one workshop discussion group favored more, smaller projects in database building and implementation, in concert with the other research efforts.

The NEL initiative should include a prioritized funding plan for expanding the services and resources beyond purely science, engineering, and technology to a broader base of knowledge that supports the nation's educational and research and development infrastructure, from K-12 through graduate work.

Projects related to the National Electronic Library; Success stories, results from evaluations:

1. Science Library Catalog Project, UCLA and CalTech.
Abstract from Borgman, Gallagher, and Walter, 1992:

Computers are being introduced into libraries and schools, often without an adequate understanding of people's abilities to use them effectively. This is especially the case in elementary school libraries, where little is known about how children search for information in either manual or automated environments. Studies of cognitive development suggest that the skills required by standard, keyword command-driven on-line catalogs are beyond the abilities of elementary school children, and that a hierarchical, browsing-oriented hypertext system would be a better match.

We report results of two information-seeking experiments in two elementary schools. The first experiment compared children's search success and time on the Science Library Catalog, a hierarchical browsing system, in two schools with their own library databases. The second experiment compared a later version of the Science Library Catalog to a keyword system. We found that children aged 10 through 12 were able to use both systems and that younger children were more successful searching the browsing system than the keyword system. Younger children aged 9 through 11 were successful on the Science Library Catalog in both experiments at both schools, and when compared to the keyword system, preferred the Science Library Catalog. Children achieved comparable overall success rates and search times between the two systems, but the keyword system was more sensitive to differences in age, question set, and computer experience. We conclude that a hierarchical, browsing-oriented on-line catalog interface is more suited to the skills, abilities, and interests of children in the concrete-operational (age 7-11) stage of development. Children entering the formal-operational stage (age 11-), with adequate computer experience, perform equally well on a keyword system and prefer it, at least for simple search tasks.

2. Conferences related to the National Electronic Library

Creating CD-ROMs for Science Education Conference. April 1-3, 1992, Colorado Springs. Organized by the Center for the Improvement of Physics Instruction, University of Nebraska, with funding from the National Science Foundation, Educational and Human Resources Directorate, Award MDR-9253138, Gerhard Salinger, Program Officer. Proceedings available from University of Nebraska.

This conference brought together about 45 people representing 8 projects developing databases for science education, K-12 and undergraduate, plus representatives of professional organizations and industry.

For further information, contact, Robert G. Fuller, Center for the Improvement of Physics Instruction, University of Nebraska, 1110 Ferguson Hall, Lincoln, NE 68588-0109. Phone 402/472-2790 or 800/332-0265(ask for Toolkit); fax 402/472-6234. For subscription to mail server, contact Norman Chonacky: nchonacky@james.bowdoin.edu

Invitational Symposium on Knowledge Management. October 27-29, 1991, Coolfont Conference Center, Berkeley Springs, WV. Sponsored by the Council on Library Resources, Johns Hopkins University, and the University of California, San

Francisco. Synopsis available from Johns Hopkins.

This conference brought together about 70 people from universities, funding agencies, and related organizations to discuss the changing infrastructure of scholarly communication in the age of electronics. The major discussion groups were Scholarly and Scientific Communication, Intellectual Property, Institutional Infrastructure, Education, and New Technologies for Knowledge Management.

For further information, contact Nina W. Matheson, William H. Welch Medical Library, School of Medicine, 1900 East Monument St, Baltimore, MD 21205. Phone 410/955-3411; fax 410/955-0985.

User Interfaces for On-line Public Access Catalogs. November 1, 1991, Library of Congress, Washington DC. Sponsored by the Library of Congress.

This conference brought together about 40 people from LC, other National Libraries, and researchers in information retrieval and interface design, most of whom were in Washington for the Annual Meeting of the American Society for Information Science. Organizers of the conference include Edward Fox, Ben Shneiderman, Gary Marchionini, Ray Larson, and Dean Wilder. A larger conference is planned for 1993, with invited papers.

For further information contact Ray Larson, Graduate School of Information and Library Studies, University of California, Berkeley, Berkeley, CA 94720. phone 510/524-2275; ray@sherlock.berkeley.edu. A brief synopsis of the 1991 meeting is available from Trudi Bellardo, School of Library and Information Science, Catholic University, Washington, DC 20064. phone 202/319-5085; fax 202/319-5574. bellardo@cua

CIKM-92, First International Conference on Information and Knowledge Management. November 8-11, 1992, Baltimore, MD. Sponsored by ISMM and the University of Maryland Baltimore County in cooperation with AAAI, IEEE, ACM (SIGART and SIGIR), Bellcore, UMBC.

For more information, send mail to cikm@cs.umbc.edu

Testimonials related to the National Electronic Library:

Support for the initiative will be needed from all corners of the information world. This list of organizations is anecdotal, from my personal knowledge of the field. It would be useful for the developers of the initiative to do a thorough search of the directories of professional organizations and foundations to find other groups as well. Addresses and contact names for these organizations and others can be located in standard bibliographic sources. I've added some comments on the organizations where useful.

- American Library Association (50K plus members, active lobbying office in Washington, DC, major player, very important source of support for initiative.)
- Library and Information Technology Association (a subgroup of ALA, large, represents those involved in library automation and networks)
- National Commission on Libraries and Information Services (federally funded, organizes White House Conferences on Libraries and Information Services; held hearings on NREN and libraries in July 1992).
- Computer Professionals for Social Responsibility (interested in workplace issues, responsible use of technology, privacy issues. Testified before NCLIS on NREN issues. Washington office headed by Marc Rotenberg.)
- Association for Computing Machinery (especially special interest group on Information Retrieval)
- Coalition for Networked Information (Washington, DC. Working on standards for information delivery on Internet, cooperative issues. New but very active group.)
- Association for Research Libraries (Washington, DC. Umbrella for all major university and other research libraries in the U.S.)
- American Society for Information Science (Silver Spring, MD. Scholarly and professional society for broad range of theoretical and technical issues in information and delivery. Mix of librarians, information retrieval researchers, government and private industry.)
- Council on Library Resources (Washington, DC. Primary private funding agency for library-related research in the U.S.).
- Information Industry Association (Represents small and large electronic publishers)
- Corporation for National Research Initiatives (Vinton Cerf, Robert Kahn, et al. Active in developing network infrastructure.)
- Society for Scholarly Publishing (Represents university and private scholarly presses)

- Society for the Social Studies of Science (Scholarly society, studies scholarly communication infrastructure)
- Special Libraries Association
- Medical Library Association
- Art Libraries of North America (ARLIS)

(other library associations representing special groups)

- Museum Computing Network (building databases of museum information and networking; often modeled on work of libraries)
- Professional societies of major scientific disciplines who would be served (American Chemical Society, Physics, Biology, Geology, etc.)
- American Educational Research Association (interest in classroom technologies)
- Professional education societies (teachers, university educators, etc.)

Figures, tables related to the National Electronic Library:

I don't have much to contribute here. Will keep my ear to the ground and send snippets as I see them. One useful figure would be the rate of Internet growth. I've seen figures of 15-25% per month of users, though these numbers change quickly.

Bibliography related to the National Electronic Library:

Reference list from Borgman (in press).

Bailey, C.W. (1989). Public access computer systems: The next generation of library automation systems. Information Technology and Libraries, 8(2), 178-185.

Barlow, J.P. (1992). Electronic frontier: The great work. Communications of the ACM, 35(1), 25-28.

Barron, B. (1992). Internet on-line public access catalogs. Internet Society News, 1(1), 24-25.

Bates, M.J. (1992). Scholarly Communication. Manuscript, Graduate School of Library and Information Science, University of California, Los Angeles.

Bates, M.J.; Wilde, D.; & Siegfried, S. (1992). The Getty on-line searching project report no. 1: Categories of information sought by humanities scholars. Manuscript submitted for publication.

Borgman, C.L. (in press). Round in circles: The scholar as author and end-user in the electronic environment. In Hazel Woodward & Stella Pillig, (eds.) The International Serials Industry: Current and Future Issues for Publishers and Scholars, LIbrarians, and Suppliers. Aldershot, Hampsire, Eng: Gower Pub Co. (publication February 1993).

Borgman, C.L. (1990). Editor's introduction. In Borgman, C. L. (ed.), Scholarly Communication and Bibliometrics, Newbury Park, CA: Sage, pp.10-27.

Borgman, C. L. (1989). Bibliometrics and scholarly communication: Editor's introduction. Communication Research, 16(5), 583-599.

Borgman, C. L. (1986a). Why are on-line catalogs hard to use? Lessons learned from information retrieval studies. Journal of the American Society for Information Science, 37(6), 387- 400.

Borgman, C. L. (1986b). The user's mental model of an information retrieval system: An experiment on a prototype on-line catalog. International Journal of Man-Machine Studies, 24(1), 47-64.

Borgman, C.L.; Case, D.O.; Ingebretsen, D. (1985). University Faculty Use of Computerized Databases: An Assessment of Needs and Resources. On-line Review, 9(4), 307-332.

Borgman, C.L., Gallagher, A.L., & Walter, V.A. (1992). The design and evaluation of an information retrieval system based on children's abilities: The Science Library Catalog. Manuscript submitted for publication.

Borgman, C. L., Gallagher, A. L., Walter, V.A., & Rosenberg, J.B.

(1991). The Science Library Catalog Project: Comparison of children's searching behavior in a direct manipulation and a keyword search system. In J-M. Griffiths (ed.), Proceedings of the 54th American Society for Information Science Annual Meeting. October 27-31, Washington, DC. Medford, NJ: Learned Information, pp. 162-169.

Carrigan, D.P. (1991). Publish or perish: The troubled state of scholarly communication. Scholarly Publishing, 22(3), 131-142.

Carrigan, D.P. (1990). The political economy of scholarly communication and the American system of higher education. Journal of Academic Librarianship, 15(6): 332-337.

Case, D.O. (1991). Conceptual organization and retrieval of text by historians: The role of memory and metaphor. Journal of the American Society for Information Science, 42(9): 657-668.

Case, D.O. (1986). Collection and organization of written information by social scientists and humanists: A review and exploratory study. Journal of Information Science, 12, 97-104.

Case, D.O; Borgman, C.L.; & Meadow, C.T. (1986). End-User Information-Seeking in the Energy Field: Implications for End-User Access to DOE RECON Databases. Information Processing & Management, 22(4), 299-308.

Computer-Readable Databases. 8th ed. (1992). K.Y. Marcaddio (ed.), M.E. Williams (founding ed.). Detroit: Gale Research.

Crane, D. (1972). Invisible colleges: Diffusion of knowledge in scientific communities. Chicago: University of Chicago Press.

Crane, D. (1971). Information needs and uses. In C.A. Cuadra and A.W. Luke (eds.), Annual Review of Information Science and Technology, 6. Chicago: Encyclopedia Britannica, pp. 3-39.

Directory of On-line Databases, 13(1). (1992). Detroit: Cuadra Associates, Gale Research.

Directory of Portable Databases, 3(2). (1991). Detroit: Cuadra

Associates, Gale Research.

Duda, K. (1992). Real world Kerberos: Authentication and privacy on a physically insecure network. Intertek, 3(3), 22-25.

Fayen, E.G. (1989). Loading local machine-readable data files: Issues, problems, and answers. Information Technology and Libraries, 8(2), 132-137.

Garvey, W. D.; & Griffith, B. C. (1964). Scientific information exchange in psychology. Science, 146, 1655-1659.

Glicksman, M. (1990). Changing patterns of scholarly communication: Implications for libraries. Library Acquisitions: Practice & Theory, 14, 341-346.

Hayes, R.M. (1992). The promise and the peril of electronic information delivery in research libraries. Speech to OCLC Directors' Conference, 9 March 1992.

Hayes, R.M. (1990). Desktop publishing: Problems of control. Scholarly Publishing, 21(2), 117-123.

Hurych, J. (1986). After Bath: Scientists, social scientists, and humanists in the context of on-line searching. The Journal of Academic Librarianship, 12(3), 158-165.

Kent, S. (1992). Security initiatives in the Internet. Internet Society News, 1(1), 34.

Landweber, L. (1992). International connectivity. Internet Society News, 1(1), 3-5.

Latour, B., & Woolgar, S. (1979). Laboratory Life: The social construction of scientific facts. Beverly Hills, CA: Sage.

Lievrouw, L.A. (1990). Reconciling structure and process in the study of scholarly communication. In Borgman, C. L. (ed.), Scholarly Communication and Bibliometrics, Newbury Park, CA: Sage, pp.59-69.

Line, M.B. (1973). Information needs of the social sciences. INSPEL, International Journal of Special Libraries, 8, 29-39.

Lynch, C.A. (in press). Networked information: A revolution in progress. Proceedings of the Illinois Clinic on Library Applications of Data Processing, 1991. Urbana, IL: University of Illinois.

Lynch, C.A., & Preston, C.M. (1990). Internet access to information resources. In M.E. Williams (ed.), Annual Review of Information Science and Technology, 25. Amsterdam: Elsevier, pp 263-312.

Lyons, P. (1992). Rules of the road: Network law. Internet Society News, 1(1), 41.

McCain, K.W. (1990). Mapping authors in intellectual space: Population genetics in the 1980s. In Borgman, C. L. (ed.), Scholarly Communication and Bibliometrics. Newbury Park, CA: Sage, pp.194-216.

Meadows, A.J. (1974). Communication in Science. London: Butterworth's.

Meadows, A.J. (1989). Communicating Research - Past, Present, and Future. Inaugural lecture, professor of library and information studies, Department of Library and Information Studies, Loughborough University of Technology, Loughborough, Leicestershire, U.K.

Menzel, H. (1967). Planning the consequences of unplanned action in scientific communication. In A. DeReuck and J. Knight, Communication in Science: Documentation and Automation. London: Churchill. pp 57-77.

Merton, R.K. (1957). Priorities in scientific discovery. American Sociological Review, 22(4), 635-659.

Morrison, P. (1979). Since Bath: A review of published information studies in social and behavioral sciences, 1974 through 1978. Behavioral and Social Sciences LIbrarian, 1, 5-22.

Nelson, T.H. (1987). Literary Machines. Edition 87.1. Self published; no place of publication given. Also published by The Distributors, South Bend, IN.

Okerson, A. (1990). Scholarly publishing in the NREN. ARL, 151, 1-4.

Osburn, C.B. (1989). Structuring of the scholarly communication system. College & Research Libraries, 50(3), 277-286.

Paisley, W.J. (1965). The flow of (behavioral) science information: A review of the research literature. Stanford, CA: Institute for Communication Research, Stanford University.

Pierce, S.J. (1990). Disciplinary work and interdisciplinary areas: Sociology and bibliometrics. In Borgman, C. L. (ed.), Scholarly Communication and Bibliometrics. Newbury Park, CA: Sage, pp. 46-58.

Price, D.J.deS. (1965). Networks of scientific papers. Science, 149, 510-515.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? Journal of Documentation, 25, 348-349.

Quarterman, J.S. (1990). The Matrix. Bedford, MA: Digital Press.

Rawlins, G.J.E. (1991). The new publishing: Technology's impact on the publishing industry over the next decade. Technical report. Department of Computer Science, Indiana University, Bloomington, IN 47405. Available via ftp (file transfer protocol) from iuvax.cs.indiana.edu: /usr/ftp/pub/techreports/tr340.ps.z

Rogers, S.J.; & Hurt, C.S. (1989). How scholarly communication should work in the 21st century. Chronicle of Higher Education, 36(7), A56. Reprinted in College & Research Libraries (1990), 51(1), 5-8.

Silverman, R.A. (1989). Desktop publishing: Its impact on the academic community. Scholarly Publishing, 21(1), 57-63.

Skelton, B. (1973). Scientists and social scientists as information users: A comparison of results of science user studies with the investigation into information requirements of the social sciences. Journal of Librarianship, 5, 138-155.

Small, H.G. (1978). Cited documents as concept symbols. Social Studies of Science, 8, 327-340.

Smith, L.C. (1981). Citation analysis. Library Trends, 30, 83-106.

Sproull, L., & Kiesler, S. (1991). Connections: New Ways of Working in the Networked Organization. Cambridge, MA: MIT Press.

Stone, S. (1982). Humanities scholars: Information needs and uses. Journal of Documentation, 38(4), 292-313.

Uncapher, W. (1991). Trouble in cyberspace. The Humanist, 51(5), 5-14.

White, H.D. (1990). Author co-citation analysis: Overview and defense. In Borgman, C. L. (ed.), Scholarly Communication and Bibliometrics. Newbury Park, CA: Sage, pp. 84-106.