

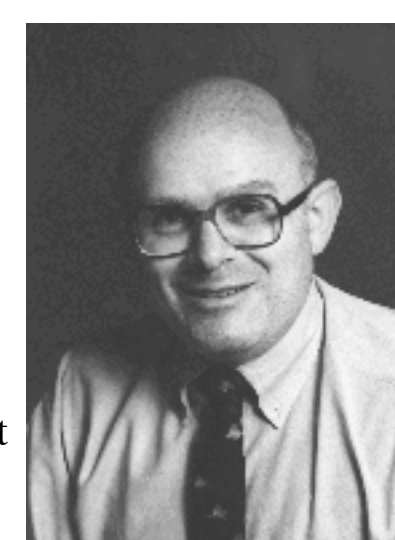
Coalition for Networked Information Spring 1998 Task Force Meeting Report

Introduction

The **Coalition for Networked Information's Spring Task Force Meeting** was held in Crystal City, Virginia on April 14-15, 1998. A wide variety of topics related to networked information were covered in plenary talks and project briefing sessions.

Quantities of Information

The opening plenary session featured a presentation by **Michael Lesk**, on leave from Bellcore, and currently serving as Division Director of Information and Intelligent Systems at the National Science Foundation, where he oversees the Digital Libraries 2 initiative. Lesk has written extensively about information and digital libraries, and recently produced a valuable book, *Practical Digital Libraries*. CNI's Executive Director Clifford Lynch stated in his introduction that Lesk is known for the provocative things that he says, and for his insightful and pithy observations that open up areas of speculation. Lesk's theme was the amount, worth, and usability of the information available in the world.



The management of the sheer quantity of information available in digital form is a daunting task. Lesk gave examples and presented charts to demonstrate how much information of various types there is in the world. His examples ranged from collections of photographs by professionals and amateurs to digitized phone conversations to satellite data. He noted that economists talk about supply and demand and its relationship to value and then asked whether we will have so much information, including freely available information, that it will become worthless.

A related issue is how much information will actually be seen or used by humans. Lesk cited a study by a psychologist who says there are about 200 megabytes in human memory and that people can take in 1 byte/second of information. An average American spends 300 hours/year with some kind of media. In the world of the future, there will be more information than human memory can cope with, so some information will never be looked at by anyone. This is true already of some large datasets, such as those collected by NASA. Lesk concluded, "in a short time there will be so much information that only a small fraction of it will be seen by a human being."

Lesk then moved to his next issue, the value of information, asking if there's so much information, is it worth anything? He noted that since libraries don't charge, no one knows what they are worth. Universities spend an average of 3% of their budget on libraries but do they know if it is well spent? Lesk believes that having a lot of information is, in fact, valuable. The information sector of our economy is growing, and in more and more industries, a larger portion of costs will be information-related instead of materials and/or labor-related. Government investments in information-related research have spawned many successful commercial systems and ventures. Lesk also described some of the economic dilemmas publishers face over whether to publish in print, electronically, or in both formats, and what various charging schemes will do to their overall revenue. Most studies have found that users are very eager to have electronic versions of documents but are generally reluctant to completely give up print.

Another issue of importance in digital information is its persistence. **Brewster Kahle** found that the average web page now lasts only 74 days (up from 48 days). Lesk noted that we know the specifics of the first phone call ever made and the first telegram, but we cannot identify the first e-mail message; no one considered archiving or preserving it when it was sent. He asked, "What do we do about these things?"

Lesk had some specific suggestions for the university community. He encouraged universities to develop high prestige sections of their websites, stimulated by cash payments to authors, that would become desirable venues for scholarly publications. Universities could control the economic system of scholarly publication and the preservation of that information. He also stated that while in a few years it will be possible to record literally all information, we need to decide what we can do with that information; therefore, we need more research on information-seeking and information use.

Lesk concluded with his observations on what these trends in information mean to society and to information professionals. He stated:

- Summarizing is the key problem for us to work on. We need to be able to take quantities of information and abstract the useful parts in all formats - text, audio, video. There has been relatively little research on this but people are starting to attend to it.
- Librarians will be worth more, and libraries may be worth less. People good at managing memory might be the ones who matter.
- Attention is the scarce resource, not information. Organizing information and helping people find their way through it is a "good thing."
- Information professionals will become more valuable as people increasingly rely on an information specialist to help them deal with the quantity of information out there. The focus of information specialists must be on helping people.

Digital Preservation

The scheduled closing plenary speaker, **Janet Murray**, MIT professor and author of *Hamlet on the Holodeck* was unable to make her presentation due to illness. We hope she will be able to join us at an upcoming CNI meeting. A panel on the preservation of digital materials was offered in place of Murray's talk. Clifford Lynch opened the panel by commenting that digital preservation is "a huge problem for scholarship and for those concerned with access to digital materials." It is an area that has been hard to make progress on, and it represents a high stakes arena for the CNI constituency. It is clear when we fail in digital preservation and never know we succeed - we know only that we have not yet failed. Lynch continued by commenting that the Commission on Preservation and Access (CPA) and RLG Task Force report of 1995 did an excellent job of laying out problems on a national and international scale, and addressed the legal problems that hinder this work. However, it is not yet clear what a useful research agenda is. The panelists then updated the attendees on a number of initiatives and laid out key issues.



Howard Besser, University of California, Berkeley, opened the panel with a presentation on "Planning to Maximize Longevity." He encapsulated two of the main problems of preserving digital information: its short life and the "viewing problem," the fact that digital information requires a whole infrastructure to view it and that infrastructure is changing at an incredibly rapid rate. He then identified a number of specific problems and posed a number of questions:

scrambling problem - we have used some technologies to solve immediate problems, e.g. compression and container architecture to enhance digital commerce, that will make access to that information in the future more complex.

- inter-relation problem - information is increasingly inter-related to other information; how do we make our own information persist when it points to and integrates with information owned by others? What are the boundaries of a set of information?
- custodial problem - how do we decide what to save; who should save it; how should they save it? What methods will be used for later access: emulation, migration, or other solutions? How can the custodial body ensure authenticity?
- digital problem - content translated into new delivery devices changes the meaning of that content, e.g. a photo of a painting is different from the painting itself. If information is produced originally in digital form in one encoded format, will it be the same when it is translated into another format? Besser used an interesting example of an early computer game, authored by Jaron Lanier, that was resurrected by some students, but when the inventor saw the "new" version, he said that the pace and timing of the game were completely different from his original version. It wasn't his game.

Besser then described what he sees as pieces of the solution. He said that we need to insist on clearly readable, standardized ways for digital objects to self-identify their formats. He stated that we should discourage scrambling of digital information. He proposed the development of a better understanding of the principles of inter-relationship of information, and what constitutes boundaries of information objects. Finally, he suggested that we develop guidelines on how to make digital information persist.

Besser also commended a number of individuals and groups who are working on these issues, including the CPA Task Force, Peter Graham of Rutgers' study group, those involved in a conference, "Time and Bits: Managing Digital Continuity," convened by the Getty Information Institute, Brewster Kahle's Internet Archive project, which saves snapshots of much of the net; and Stuart Brand's Long Now Foundation.

The second speaker, **Don Waters**, of the Digital Library Federation (DLF), Council on Library and Information Resources (CLIR), described the mission of archiving initiatives as the need to "preserve integrity and ensure persistence." He noted the increasing visibility of this set of issues in the mainstream press partially due to the impact of the CLIR's "Into the Future" video and recent coverage in the *New York Times*.

Waters touched on several initiatives in which progress is being made. In June, 1998, a workshop on digital archive directions will be hosted by the National Archives. Jeff Rothenberg of Rand Corporation is working with CLIR on a type of emulation in which technologists develop ways of encapsulating objects and software and have annotation mechanisms about those encapsulated objects and software that will make them usable on future generations of hardware and software. He also commented on the interplay between the Making of America project and digital preservation issues. The first two phases of that project produced materials, some of which are already becoming inaccessible due to format or system changes. He hopes to pursue digital preservation issues in the context of the Making of America project.



William Arms, Corporation for National Research Initiatives, presented his taxonomy of technical issues for digital archiving. He parsed the research issues into five levels and stated the defining problem(s) of each level. He indicated that each of the problems needed to be addressed not just as a technical research topic, but within an organizational, legal, and social context.

At level one, the bottom level, the issues revolve around the physical media used to store bits. Can the computer industry invent media that will store data reliably over long periods of time? At the next level, problems relate to refreshing bits - copying from one media to another when one media wears out. Arms noted that while copying is not particularly expensive, the refreshing has to be done continually. He asked what organization can be trusted to do this through periods of turmoil? The first two levels do not address the problem that computer environments change, so that although the bits have been archived, they can no longer be interpreted. Hence, the third level is preservation of content by migration from one generation of computer system to the next, transforming formats, protocols, data structures, etc. in the process. Again, a committed organization is needed to do this systematically. The fourth level aims to achieve preservation of content by emulation of the computing environment. Arms questioned how practical emulation is for complete environments, but encouraged its use in narrowly defined areas.

The fifth level is digital archaeology - the process of regenerating digital content when it has not been systematically archived. He suggested that we need to scatter some Rosetta stones around the network to enable decoding of digital objects in the future.

Arms suggested that authors should create digital documents with archiving in mind. He had two specific suggestions for improving the current situation: pay particular attention to the metadata associated with the document, and assign copyright to libraries for the purpose of long-term archiving.

Melissa Levine, Media Advisor for the National Digital Library in an Library of Congress, has been engaged in a five-year project to put on the web many historical materials for free public access. The collection includes multi-media materials and items that exist in both in analog collection as well as materials that exist only in digital form. She described their experience with materials from Coolidge, which include some clippings, some photos, a mix of materials out of copyright protection and some that are under current copyright restrictions.

Levine stated that the issues that she looks at deal with copyright, privacy rights, indecency and obscenity. She stated that the issues can be dizzying in their range and complexity. She described some challenges in the context of current legislation, including the proposal to increase the term of copyright protection, which will have an impact on when materials enter the public domain. Part of the complexity of copyright as seen through the perspective of digital preservation is that the legal status of works can change over time, due to changes in the copyright law and longevity, for example. She advised putting in the metadata associated with a digital object some information about its provenance. She cautioned the audience that it is very important now and will be in the future to be able to determine who has what rights for what version of a document or software.

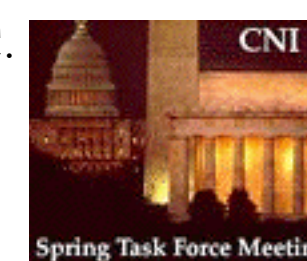


Levine stated in closing that the jurisdictional and global issues posed by the Internet are very interesting. A bill before Congress would make it illegal to circumvent any type of protection of a digital object. She would have liked to have seen more careful exceptions either for libraries or for certain preservation activities; right now, those exceptions are not in the language of the legislation.

Clifford Lynch wrapped up the session by stating that we're still stuck in the mind-set of producing a physical object and preserving an artifact. We have situations in the digital environment where we want to preserve the evolution of digital objects, not just freeze an artifact. He asked whether we understand the requirements of such preservation? Lynch asked the audience to help CNI identify what we can do to make tangible progress in the area of digital preservation.

Project Briefings

Project briefings were held on a wide array of topics, from Internet 2 and NLI's Instructional Management System (IMS) to information literacy to ARL's scholarly publishing initiative, SPARC. CNI's website has information on many of the project briefings and links to some related sites. Information is located at: <http://www.cni.org/tfms/1998a.spring/>.



CNI Projects

A number of project briefing sessions were devoted to updates of CNI projects. In a session on the CNI Program on Authentication, Authorization and Access Management, Clifford Lynch invited comment on CNI's recently issued second draft of a white paper on the topic. (The current version of the paper is available on CNI's website.) Two sessions highlighted a total of four projects affiliated with CNI's Institution-Wide Information Strategies (IWIS) project, led by Gerry Bernbom of Indiana University. Susan Perry of Mt. Holyoke College, Philip Tompkins of Indiana University - Purdue University of Indianapolis, and Joan Lippincott of CNI gave an overview of the successful CNI project New Learning Communities, and described the conferences, workbook, video, and other products and outcomes of that project. They also detailed what they had learned from the participating institutions concerning success factors and roadblocks to developing new learning communities. In a session on the CNI project Assessing the Academic Networked Environment, attendees heard from project leaders Chris Peebles of Indiana University and Charles McClure of Syracuse University and leaders of two of the participating projects.

Paul Evan Peters Award and Scholarship Fund

Robert Heterick, President of Educom, reported that nominations for the **Paul Evan Peters Award** are due by May 1, 1998. The award is sponsored by ARL, Educom, CAUSE, Microsoft, and Xerox. Additional information is available on Educom's website at <http://www.educom.edu/>.



Duane Webster, Executive Director of ARL, encouraged additional contributions to the **Paul Evan Peters Scholarship Fund**. A committee headed by **Charles Henry** of Rice University has been soliciting gifts and will make the first scholarship award later this spring. Information on the scholarship fund is available on CNI's website.

Fall Task Force Meeting

The **Fall 1998 Task Force Meeting** will be held at the Sheraton Seattle Hotel and Towers in Seattle, Washington on December 7 and 8, immediately preceding **CAUSE '98**.



What is CNI?	Projects	Meetings	Conferences
What's New?	Net Services	Search	Archives

CNI

21 Dupont Circle Suite #800
Washington, DC 20036-1109
202.296.5098
<http://www.cni.org/>