

"Digital Reality II: Preserving Our Electronic Heritage"

by Kaylyn Hipps, Assistant Editor of ARL

On 5 June 2000, I attended--along with 350 others--the "Digital Reality II: Preserving Our Electronic Heritage" conference cosponsored by the NELINET Preservation Advisory Committee, John F. Kennedy Library, and Northeast Document Conservation Center (NEDCC) and hosted by the Kennedy Library in Boston. The conference brought together five experts in digital technology and preservation to present their views of the future of the World Wide Web and libraries and archives, and how we might best preserve the growing amounts of digital information being produced.

After opening remarks by Megan Desnoyers of the Kennedy Library, Ann Russell of NEDCC, and Arnold Hirshon of NELINET, the morning was dedicated to speculation about the future of the Web and its impact on libraries and archives. Tim Berners-Lee, of the Massachusetts Institute of Technology and World Wide Web Consortium (W3) and inventor of the Web, noted that the Web has made more apparent to more people the necessity--and difficulty--of preserving our digital information. When asked what the greatest problem with the Web is, many people cite the frequency with which they encounter "Error 404: URL not found." Berners-Lee observed that companies do not change their phone numbers or product I.D. numbers without building in a referral to the new number, but they often change URLs and fail to provide referrals because websites are difficult to manage. He would like to see permanent domain names and policies enacted to encourage persistent URLs; he believes that librarians should lead the charge for persistent URLs because that will require the establishment of social standards related to information--librarians' territory.

Walt Crawford of the Research Libraries Group presented three different scenarios for the future of the Web: "the Web as life," "the Web as CB radio," and "the Web in complexity." "The Web as life" is the scenario where everything is digital, where people obtain all they need (products, information, entertainment) via the Web and live in isolation from one another. Crawford sees this as highly improbable. He believes that digital resources (and experiences) will complement, not replace, analog ones. The vision of "the Web as CB radio" sees use of the Web moving from general to specific. Crawford believes this future is as unlikely as "the Web as life" although he does believe that the Web will lose some of its general appeal as the hype surrounding it dies down. The most probable future, in Crawford's opinion, is that the Web will become more complex as web-based tools for creating services and publications evolve. The Web will not replace buildings but will be used to develop new resources. Libraries will continue to encompass resources, services, and place.

In the afternoon, the conference sharpened its focus on the issue of digital preservation. Paul Conway of Yale University Library set the stage with a description of the context and issues involved in discussing digital preservation. He asked, "What is the place of preservation theory and practice in a world dominated by information created, shared, sold, and used in digital form?" In the past 10 years, the library community's perspective on digital preservation has been transformed. Conway observed that best practices for preserving digital text and grayscale images are emerging and information architectures have been implemented. He sees digital preservation as "the creation of digital products worth maintaining over time," and emphasized that use must drive preservation choices and technology. This philosophy challenges the foundation of the research library collection built in anticipation of future use, but Conway believes use validates the creation of these products.

Fynnette Eaton, of the Smithsonian Institution Archives and formerly of the National Archives and Records Administration (NARA), presented her experience with the preservation strategy of migration--the transfer of digital data from one technological configuration to another to ensure readability without substantive loss of content, structure, or context. First, Eaton described the 30-year evolution of the

NARA procedures for creating preservation copies of electronic records, culminating with the modifications to their Archival Preservation System (APS) required by the task of creating preservation copies of the email backup tapes from the Reagan and Bush White Houses. Eaton also outlined three current research projects focusing on migration of electronic records:

- the [InterPARES \(International Research on Permanent Authentic Records\)](#) project, whose goal is to develop the essential knowledge for preserving electronic records and formulate model policies, strategies, and standards that will ensure preservation;
- the [Collection-Based Persistent Digital Archives](#) project at the San Diego Supercomputer Center, which is testing the feasibility of preserving various types of digital information over time by integrating archival storage technology from the supercomputing community, data-grid technology from the computer science community, information models from the digital library community, and preservation models from the archival community; and
- the [Open Archival Information System \(OAIS\)](#) conceptual model for archival systems dedicated to preserving and maintaining access to digital information over time, which is a draft NISO standard initiated by NASA's Consultative Committee for Space Data Systems and developed in open forums.

Eaton concluded that there is no single answer to the problem of preserving digital information but we are beginning to ask the right questions.

Jeff Rothenberg of the Rand Corporation argued the case for preserving digital documents via emulation. Digital documents depend on software (and, therefore, also on an operating system and hardware) for interpretation. Rothenberg believes it is not sufficient to save the bits to preserve a digital document; we must save the interpreter as well. (He nicely illustrated this point with, among other examples, text from *Winnie the Pooh* whose format conveys meaning that is lost when the text is stripped of formatting.) Rothenberg argued that standards can keep documents readable and migration can be used when standards become obsolete, but migration is expensive and labor-intensive. Saving the original software is a straightforward way to maintain access to documents, but running obsolete software requires obsolete hardware or emulation of that hardware. Rothenberg argues that emulating hardware is cost effective because, once a platform has been emulated, that emulator can be used to run any software written for that platform. He admits there are problems with emulation: future users will need to know how to run obsolete software, so we may need use-copies for non-scholarly access to documents; we may have to emulate more than hardware; and emulation requires an emulator specification and environment per platform, which is an ambitious goal. Rothenberg believes emulation's strengths outweigh its weaknesses; he also believes emulation is a superior strategy to migration. Migration requires repeated conversion as technology changes, which requires understanding of individual documents, while emulation can be performed and paid for once and then it is accomplished for all documents created on that platform.

The day closed with Jan Merrill-Oldham of Harvard University Library moderating a lively panel discussion involving Crawford, Conway, Eaton, and Rothenberg as well as heavy audience participation. [More information](#) about the conference, including speakers' biographies and presentations, is available online.

[Table of Contents for Issue 211](#) | [Other Preservation of Research Resources Articles](#)
[Other Networked Information Articles](#) | [Other Collection Management Articles](#)



ARL policy is to grant blanket permission to reprint any article in the Bimonthly Report for educational use as long as full attribution is made. Exceptions to this policy may be noted for certain articles. This is in addition to the rights provided under sections 107 and 108 of the Copyright Act. For commercial use, a reprint request should be sent to ARL Director of Information Services, [Julia Blixrud](#).

© The Association of Research Libraries

Maintained by: [ARL Web Administrator](#) Site Design Consultant: [Chris Webster](#) **Last Modified:** 7 August 2000