

Panel on Digital Preservation

Joyce Ray (Moderator)
Institute of Museum and Library
Services
Washington, DC
jray@imls.gov

Robin Dale
Research Libraries Group
Mountain View, CA
Robin.Dale@notes.rlg.org

Reagan Moore
San Diego Supercomputer Center
La Jolla, CA
moore@sdsc.edu

Vicky Reich
HighWire Press
Stanford University Libraries and
Academic Resources
Stanford, CA
vreich@stanford.edu

William Underwood
Georgia Tech Research Institute
Atlanta, GA
william.underwood@gtri.gatech.edu

Alexa T. McCray
(Commentator)
National Library of Medicine
Bethesda, MD
mccray@nlm.nih.gov

ABSTRACT

Digital information in any form is at risk. Software and hardware become obsolete, and versions and file formats change, making data inaccessible. Data stored in even the simplest form are in danger due to computer media degradation and obsolescence. Online information such as e-journals and databases are susceptible. They may become partially or entirely unreadable, and may not be recoverable by the time the problem is detected. Preservation strategies such as emulation (keeping alive the software and hardware needed to access a digital object), migration (converting the digital object to new versions and formats), and other long-term archival methods have been proposed [1-7]. Models such as the Open Archival Information System (OAIS) provide an architecture for conducting digital preservation research and experimentation [8-10]. The importance of preservation metadata has been recognized by a number of groups and efforts to develop and deploy metadata standards are underway [11-14].

As more and more digital information is created, attention must be paid to what information should be preserved and how it can be preserved most economically and effectively. It is clear that for preservation to be successful, we need to pay attention not only to the format of digital objects, but also to the commitment we make to providing long-term access to the information. Thus, decisions about digital preservation will involve technical issues as well as economic, legal, social, and organizational ones. Is it possible or feasible to preserve all digital data automatically and in a cost effective way? How much functionality can or must be preserved? What type of metadata will be needed to ensure both access and preservation? What metrics do we use to evaluate whether our methods will be successful?

Panelists will make short presentations about work in which they have been involved and which reflect a variety of aspects of digital preservation. Reagan Moore will discuss the levels of abstraction that are needed to create infrastructure independent representations for data, information, and knowledge, and he will discuss a prototype persistent digital archive. The persistent

archive infrastructure has been developed for use by the National Archives and Records Administration and other Federal agencies. William Underwood will report on lessons learned in preserving digital records created on personal computers. The records being examined are the digital records created on personal computers during the administration of President George Bush (1988-1992). Vicky Reich will present work on the LOCKSS (Lots of Copies Keep Stuff Safe) project, which is a permanent web publishing and access system. LOCKSS software allows libraries to retain local collection control of materials delivered through the web while preserving the functionality of the original web based content. Robin Dale will report on activities of the preservation program of the Research Libraries Group (RLG). She will focus on the joint work of RLG and OCLC (Online Computer Library Center) on preservation metadata. Following the presentations by the four panelists, Alexa McCray will provide brief comments and then open the discussion for audience participation.

Categories and Subject Descriptors

H.3.7 Digital Libraries.

General Terms: Standardization, Reliability, Experimentation, Measurement, Design.

Keywords: Digital libraries, digital preservation, metadata, archival systems.

1. PANELISTS

Joyce Ray (Moderator) is Director of the Office of Library Services at the Institute of Museum and Library Services (IMLS). She is an archivist by training and before coming to IMLS in 1997 was with the National Archives and Records Administration (NARA) for 10 years. Ray was Assistant Program Director for Technological Evaluation at the National Historical Publications and Records Commission at NARA, where she also served as Acting Program Director. She was previously Special Assistant to the Archivist of the United States and held a number of positions at the National Archives including Archives Specialist for Policy and Program Analysis. She was Head of Special Collections at the University of Texas Health Science Center at San Antonio Library and Technical Services Librarian at Tusculum College Library, Greeneville, Tennessee. Ray has a Ph.D. in history and a master's

degree in library science, both from the University of Texas at Austin.

Robin Dale has been a Program Officer for Member Initiatives with the Research Libraries Group (RLG) for the past five and a half years. In that position, she leads one of RLG's key initiatives, the Long-term Retention of Digital Research Materials, as well as RLG's PRESERV community, a program which focuses on preserving and improving access to endangered research materials. Prior to joining RLG, Dale was Head of the Preservation Reformatting Department at the University of California, Berkeley.

Reagan Moore is Associate Director for Data Intensive Computing at the San Diego Supercomputer Center and an Adjunct Professor in the UCSD CSE department. Moore coordinates research efforts in development of massive data analysis systems, scientific data publication systems, and persistent archives. An ongoing research interest is support for information based data-intensive computing. Moore is an active participant in NSF workshops on digital libraries and knowledge networks. Recent publications include a chapter on data-intensive computing in the book "The Grid: Blueprint for a New Computing Infrastructure". He has been at SDSC since its inception, initially being responsible for operating system development. Prior to that he worked as a computational plasma physicist at General Atomics. Moore has a Ph.D. in plasma physics from the University of California, San Diego, (1978) and a B.S. in physics from the California Institute of Technology (1967).

Vicky Reich is Assistant Director, HighWire Press, Stanford University Libraries and Academic Resources. She works to facilitate the industry's transition from print to online models in a variety of ways, including directing the LOCKSS (Lots of Copies Keep Stuff Safe) project. Reich represented the Stanford Libraries on the Executive Committee of the Stanford Digital Library Technologies Project from 1995 through 2000. Stanford received federal funding for the Digital Library Initiative Phase 2. Reich has held public services and technical services positions in both public and private institutions. Her broad experience includes research librarian at the Upjohn Company; Head, Mental Health Research Institute Library, University of Michigan; Planning Librarian, Office of the Librarian of Congress; and Head, Serials and Acquisitions Department, Stanford University Libraries.

William Underwood is a Principal Research Scientist with the Information Technology and Telecommunications Laboratory of the Georgia Tech Research Institute. He is the Principal Investigator for the PERPOS project sponsored by the National Archives and Records Administration. The objective of this project is to aid archivists in Presidential Libraries in gaining intellectual control of digital records created on personal computers. Underwood is also a member of the US InterPARES research project sponsored by the National Historical Publications and Records Commission. The objective of this project is to identify technologies and methodologies for long-term preservation of authentic electronic records. He is a member of the Consultative Committee for Space Data Systems, Panel 2, that is developing standards for archival information interchange. His current research interests are in developing formal, theoretical foundations for records management and archival science, experimental investigations of alternative preservation strategies,

and decision support tools for supporting archival description and FOIA review.

Alexa T. McCray (Commentator) is the Director of the Lister Hill National Center for Biomedical Communications, a division of the National Library of Medicine, National Institutes of Health. The Lister Hill Center conducts research and development for the broad purpose of improving health-care information dissemination and use. McCray conducts research at the intersection of computer and information science and medicine. She directs several digital library projects including *Profiles in Science*, a large-scale digital conversion project, and *ClinicalTrials.gov*, a continuously evolving information resource on clinical trials. Before joining NLM in 1986, McCray was a Research Staff Member at IBM's T.J. Watson Research Center. She received the Ph.D. from Georgetown University in 1981, and for three years was on the faculty there. She conducted pre-doctoral research at the Massachusetts Institute of Technology.

2. REFERENCES

- [1] Rothenberg J. Avoiding technological quicksand: Finding a viable technical foundation for digital preservation. A report to the Council on Library and Information Resources. January 1998. <http://www.clir.org/pubs/reports/rothenberg/contents.html>. Accessed April 19, 2002.
- [2] Rothenberg J. Ensuring the longevity of digital documents. *Scientific American*. 1995; 272(1):24-9.
- [3] Granger S. Emulation as a digital preservation strategy. *D-Lib Magazine*, October 2000. <http://www.dlib.org/dlib/october00/granger/10granger.html>. Accessed April 19, 2002.
- [4] Wheatley P. Migration – a CAMiLEON discussion paper. 2001. <http://www.ariadne.ac.uk/issue29/camileon/>. Accessed April 19, 2002.
- [5] Lorie RA. Long term preservation of digital information. *Joint Conference on Digital Libraries*, 2001; 346-52.
- [6] Lynch C. Canonicalization: A fundamental tool to facilitate preservation and management of digital information. *D-Lib Magazine*, September 1999. <http://www.dlib.org/dlib/september00/lynch/lynch.html>. Accessed April 19, 2002.
- [7] Reich V, Rosenthal DSH. LOCKSS: A permanent web publishing and access system. *D-Lib Magazine*, June 2001. <http://www.dlib.org/dlib/june01/06reich/reich.html>. Accessed April 19, 2002.
- [8] Consultative Committee for Space Data Systems- Reference Model for an Open Archival Information System (OAIS) July 2001. http://ssdo.gsfc.nasa.gov/nost/isoas/ref_model.html. Accessed April 19, 2002.
- [9] Lavoie B. Meeting the challenges of digital preservation: The OAIS reference model. *OCLC Newsletter* January/February 2000; 26-30.
- [10] Attributes of a trusted digital repository: Meeting the needs of research resources. An RLG-OCLC report. Draft for public comment. August 2001.

<http://www.rlg.org/longterm/attributes01.pdf>. Accessed April 19, 2002.

- [11] OCLC/RLG working group on preservation metadata: A recommendation for content information. October 2001.
<http://www.oclc.org/research/pmwg/contentinformation.pdf>. Accessed April 19, 2002.
- [12] Research Library Group (RLG) – RLG REACH element set for shared description of museum objects. 1998.
<http://www.rlg.org/reach.elements.html>. Accessed April 19, 2002.

- [13] National Library of Australia. Preservation metadata for digital collections. 1999.
<http://www.nla.gov.au/preserve/pmeta.html>. Accessed April 19, 2002.
- [14] Networked European Deposit Library (NEDLIB). Metadata for long term preservation. July 2000.
<http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>. Accessed April 19, 2002.