Library of Congress Digital Preservation Newsletter

Memento:
Time Travel for the Web

What if you could activate a time machine right in your web browser and explore content from a date in the past? Herbert Van de Sompel of the Los Alamos National Laboratory and Michael Nelson from Old Dominion University discussed this when they visited the Library of Congress on November 16, 2009, to talk about their newest project, Memento.

Herbert Van de Sompel (l) of the Los Alamos National Laboratory and Michael Nelson (r) from Old Dominion University.

Memento refers to an archival record of a resource as well as the technological framework that supports the ability to discover and browse older versions of web resources. Nelson said that "Memento will allow the holdings of content management systems and various archives to be integrated, creating a 'long tail' for archives."

The Memento framework leverages existing capabilities and protocols of the web and applies them in a novel way to add a time-based dimension to searching and browsing. The result is a system in which archived resources can seamlessly be reached via the universal resource identifier of any current resource.

This is not possible today because current and older web content are nonintegrated. There are some archived resources available, but they are limited in scope and accessibility.

"In Memento you always go back to the original resource, but you end up at an archive," said Van de Sompel. He noted that it is cumbersome to navigate the past even when the archival resources are available.

The Memento framework works to address this situation in two ways. Memento researchers have first proposed an additional Hypertext Transfer Protocol header, X-Accept-Datetime, that allows web clients to negotiate with a web resource by specifying the date and time of the archived version they would like to retrieve for the resource.

Following on from this, the researchers realized that a multi-archive view was necessary to ensure that the user would be able to receive the archived version closest in time to the expressed date preference. To that end, they've proposed an application programming interface (API) that will create gateways to available archived versions for any specific original resource across a range of archives.

Read more about the Memento project at digital-preservation.gov. •

Library of Congress Featured in Voice of America Video

A recent video from the Voice of America draws attention to the expanding role of digital materials at the Library of Congress.

"VOA was interested in hearing about how digital technologies were changing the traditional role of the Library," said Jane Mandelbaum, a Library of Congress information technology specialist who is featured in the video. "They wanted to examine how the Library was evolving from the world's largest library in terms of traditional paper formats to becoming the host of the world's largest digital collection."


Screenshot from the VOA video. Jane Mandelbaum of the Library.

The video describes the Library as now hosting "an ocean of digital information" that encompasses more than 50 million individual files. "A terabyte is about 1600 CDs, or about 330 hours of TV, or about 2000 books," says Mandelbaum, describing the size of the Library's digital collections. "And we have about 200 of those terabytes on our website."

Read more about the video at digitalpreservation.gov. •

Digital Preservation Pioneer: Martin Halbert
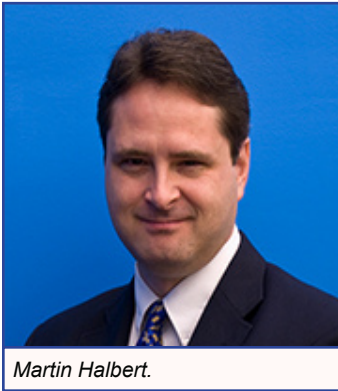
Martin Halbert.

Martin Halbert, dean of libraries at the University of North Texas, has first-hand knowledge that institutional collaboration is essential for digital preservation. He has collaborated on several successful and influential projects throughout his career, including the Transatlantic Slave Trade Database, Southcomb and the MetaArchive cooperative.

Halbert has always moved freely between the sciences and humanities; his education is in philosophy, chemical engineering and librarianship. In the 1980s, he interned at the IBM library in Austin and used their advanced computer system for mediated searches for researchers worldwide. That experience influenced Halbert's ideas about the potential for computer systems and networked information in public libraries.

An avid science-fiction fan, in 1992 Halbert wrote "Knowbot Explorations in Similarity Space" about automated search tools in the future (2010) that gather information from networked databases and graphically display the results as holograms. Though there may not be as much eye candy in today's technology as Halbert imagined in 1992, many of the functions he envisioned are now common, including easy access to web-based databases and query results displayed in graphics.

Halbert feels that his most influential project to date is the MetaArchive, which has become a model for institutions working jointly in digital preservation. "We have been running an effective, low-cost, high-impact distributed digital preservation network since 2004," said Halbert.

Read more about Halbert's interests and career at digitalpreservation.gov. •

Upcoming Event

- Federal Web 2.0 Webinar: The Library of Congress and Digital Technology: Jan. 27, 2010. Registration(free) is required.

NDIIPP Director Participates in Fall 2009 Events

Martha Anderson, director of program management for the National Digital Information Infrastructure and Preservation Program, was among the invited participants at two Fall 2009 events.

In October, Anderson presented at the Stewardship of Digital Assets Workshop, held at the Library of Congress. The workshop featured lectures, case studies and interactive sessions that highlighted the long-term needs of the digital assets and how to plan for their preservation. Anderson's talk focused on digital preservation in the federal sector.

In November, Anderson was among the invited participants to the Digital Tools for Information and Democracy Conference, sponsored by the Jefferson Institute in Charlottesville, VA. The conference allowed leaders from digital stewardship, education and journalism to share current experiences and to discuss synergies among the three disciplines. Anderson talked about how the NDIIPP network shares information across organizations and communities.•

Monticello at Charlottesville, VA. Credit: The Jefferson Institute

Announcements

- The Office of Science and Technology Policy in the Executive Office of the President and the White House Open Government Initiative launched the Public Access Policy Forum in late 2009 to explore future Federal government policy regarding public access to Federally Funded Research. Information on the activity and public comments are available on the OSTP blog or the Open Government blog.

- Fran Berman Awarded for Leadership. Berman received joint inaugural ACM and IEEE award.