



Library of Congress Digital Preservation Newsletter

Pioneer in Open Sustainable Formats Makes Her Mark



Caroline Arms will retire from the Library of Congress in June.

In 1995, Caroline Arms came to the Library of Congress to work on the [American Memory](#) project, but the looming challenge of preserving digital content eventually caught her attention. Her first article about digital preservation was titled “[Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress.](#)” Caroline went on to make major contributions to key initiatives supporting open digital formats for preservation.

“In the physical world,” Caroline says, “Preservation was something that came years later. The original publisher was long out of the picture by the time you got to think of preserv[ing] the content or the artifact of that book. You have to take steps early in the life-cycle to up the probability that the content can be preserved and reused. And these [steps] include thinking hard about the formats.”

Caroline and her colleague Carl Fleischhauer have done that hard thinking about the formats, and their exhaustive research led to the [Sustainability of Digital Formats Web site](#), the definitive inventory of information about current and emerging digital formats. This Web site has become an essential resource among the international digital preservation community.

Caroline was also active in the development of the [Open Archives Initiative Protocol for Metadata Harvesting](#) and recently represented the Library of Con-

To subscribe to this newsletter, go to http://service.govdelivery.com/service/subscribe.html?code=USLOC_13, type in your e-mail address, scroll down and click on “Digital Preservation.”

gress to prepare Office Open XML for submission as an ISO/IEC standard. She’s also worked as a program officer for several NDIIPP projects, including the Data-PASS and MetaArchive partnerships mentioned in this newsletter.

Caroline will retire from the Library of Congress next month. Her work has benefited cultural institutions nationally and internationally.

Read more about Caroline’s digital preservation work at: www.digitalpreservation.gov.

Meeting the Challenge: Specifications for Digital Formats

The Library of Congress has been actively engaged in creating and supporting the development of the following key open standards for digital content:

Office Open XML

This new standard ([ISO/IEC 29500](#)) is based on the formats used by the latest version of Microsoft Office and supports all features in the various versions of Microsoft Office since 1997. It is estimated that Microsoft Office has more than 400 million users generating billions of documents a year. The new XML-based standard will enable implementation by multiple applications on multiple platforms. The specific document types involved are word processing, presentations and spreadsheets.

As an additional benefit to the digital preservation community, Microsoft has released the specifications of its earlier binary formats and asked the Library of Congress to hold copies, now available on the [Microsoft Office Binary \(doc, xls, ppt\) File Formats page](#).

PDF/A

This standard grew out of an increasing need for the widely used PDF document format to be readily accessible and consistent in appearance over time. PDF/A is a subset of the PDF format suitable for the long-term preservation and archiving of page-oriented text documents. The Library also continues to participate in the working group developing the standard (ISO 19005-1). For more specific information on PDF/A, visit the Library’s [Formats Sustainability page](#).



New Partner Web Site

The Data Preservation Alliance for the Social Sciences (Data-PASS), one of the first projects funded by NDIIPP, recently released a new version of its Web site, www.icpsr.umich.edu/DATAPASS.



The new Data-PASS Web site offers a shared, searchable catalog of social science research.

Data-PASS describes itself as “a broad-based partnership devoted to identifying, acquiring and preserving data at risk of being lost to the social science research community. Examples of at-risk data include opinion polls, voting records, large-scale surveys on family growth and income, and many other social science studies.”

New to this release is a direct link to the partnership shared catalog, which enables anyone to search and browse the entire holdings of most Data-PASS partners. Other significant content includes project guidelines, such as selection criteria and data security standards, created and used by the partnership, as well as publications and presentations by Data-PASS partners. New content will continue to be added to the Web site as Data-PASS continues to identify, acquire and preserve data used for social science research.

Data-PASS is led by the Inter-university Consortium for Political and Social Research at the University of Michigan; the Roper Center for Public Opinion Research at the University of Connecticut; the Howard W. Odum Institute at the University of North Carolina-Chapel Hill; the Henry A. Murray Research Archive, a member of the Institute for Quantitative Social Science at Harvard University; the National Archives and Records Administration; and the Harvard-MIT Data Center, also a member of the Institute for Quantitative Social Science at Harvard University.

MetaArchive Cooperative Adds International Partner

Partners in the NDIIPP digital preservation network continue to attract new participants as organizations large and small devote resources to the preservation of digital content. In April, the San Diego Supercomputer Center reported that its Chronopolis datagrid initiative added new organizations to this demonstration project.

This month, the MetaArchive Cooperative, led by Emory University, has made significant progress in the past few months in both its organizational and technical work. The project has just welcomed its first international sustaining member, Hull University in England. “More than five dozen institutions have now contacted us about membership, and we plan to welcome additional new members this spring,” said Program Manager Katherine Skinner. “These archives will join the established, ongoing Southern Digital Culture archive that we have run since 2004.”

The MetaArchive Cooperative’s new Web site contains documentation on the MetaArchive’s network, processes, business plan and organizational framework; soon, the site will offer a tools, scripts and documentation registry for private LOCKSS networks and a downloadable version of the “Guide to Distributed Digital Preservation.”

In June 2008 the MetaArchive Cooperative and the Networked Digital Library of Electronic Theses and Dissertations will jointly host a workshop in conjunction with the electronic theses and dissertations conference in Aberdeen, Scotland.

The cooperative received notification this quarter that it has been awarded a \$300,330 grant from the National Historical Publications and Records Commission to extend its digital preservation services to libraries, archives, government agencies, historical societies and other cultural heritage repositories throughout the United States. The project will transform the MetaArchive Cooperative, a community-based distributed digital preservation network, from a regional project to a sustainable digital preservation business with a national impact. As part of the project, the cooperative will build a bridge between the LOCKSS software that provides the MetaArchive network’s technical framework with the Storage Resource Broker framework being designed by the San Diego Supercomputer Center, and will make that bridge available to other projects via an open-source license.