Library of Congress Digital Preservation Newsletter

Federal Agencies Collaborate on Digitization Guidelines, Pt. II

This is the second of three articles about the Federal Agencies Digitization Guidelines Initiative, this time with a focus on the Still Image Working Group. The next and last article in the series will focus on the Audio Visual Working Group.

STILL IMAGE WORKING GROUP

Creating a digital collection with textual content, maps, photographic prints and negatives is a bit like piecing together a giant puzzle – and the technical requirements for the digital copies are one of the most important pieces. A host of decisions must be made before digitization (or scanning) gets under way. What are the standards for an acceptable digital copy in terms of color, resolution and faithfulness to the original? What metadata elements should be captured? A successful project depends on the right answers to these and many other questions.

The new Federal Agencies Digitization Guidelines Initiative aims to help agencies establish the right requirements for cultural heritage materials. The project has launched a Web site to document and share standards, best practices and other useful information.

Two specialized working groups are behind this effort: one for still images and another for audio-visual materials. Each group will produce guidelines covering the most crucial issues for digitizing historical, cultural and archival materials.

The Still Image Working Group has members from ten government agencies. Agencies currently involved include the U.S. Geological Survey, Library of Congress, National Agricultural Library, National Archives and Records Administration, National Gallery of Art, National Technical Information Service, National Library of Medicine, National Transportation Library, Smithsonian Institution, and the Government Printing Office.

In addition to agency representatives, the group has an advisory board of experts to provide a perspective from outside the government. The current board includes Stephen Abrams (California Digital Library), Rob Buckley (Xerox), and Don Williams (consultant).

The Still Image group will focus on developing guidelines for the activities associated with digital imaging and encoding, as well as handling metadata that will be embedded in digital files. A major goal is to implement standard measurements of image quality as set by project requirements. Another area of attention is the "reformatting" of printed, pictorial items, which will aim to develop categories and objectives to help frame questions about the intended use of the copies.

This is a highly collaborative effort, and input and feedback is actively encouraged from outside organizations as well as the public.

While much work remains, the group provides a number of resources through its Web site, including a large (and growing) glossary of technical terms, many of which are designated as specifically image related. Interested in definitions for such imaging terms as "color filter array," "downsampling," or "resolution"? Check them out in the group's glossary.

Other information available includes current institutional imaging guidelines and a list of formal industry standards. A toolkit of public domain software for image quality metrics and image file compliance will be available at a later date.

Michael Stelmach, manager of Digital Conversion Services at the Library of Congress and a key organizer of the group, is optimistic about its prospects. "So far, we've seen a very positive response to the collaboration," he said. "The issues we are tackling seem to resonate with others, both in the public and private sectors, who are working with digital conversion. Going forward, we hope that our work can serve as a useful reference for cultural heritage institutions." •

Read the first article in this series: http://www.digitalpreservation.gov/library/challenge/DigitizationGuidelinesPart1.html

LIBRARY OF CONGRESS

Library Releases Software Tools

The Library of Congress has re-leased software tools that cultural heritage organizations can use to send and receive digital data. All the tools are open source, which means they can be freely used and modified with minimal conditions.

The tools are available through SourceForge.net, the technology community's hub for open source software distribution and services, under the "Library of Congress Transfer Tools" project.

The project is based on use of the BagIt specification, which is a hierarchical file packaging format for the exchange of digital content. The Library's Repository Development Group worked with the California Digital Library to jointly develop the specification.

These are the first software tools the Library has formally released as open source. They support validation and transfer of data that conforms to the BagIt specification.

The Library plans to release additional tools as part of a suite of solutions and software development resources as they are completed over time.

Three tools are available now. Bag Validator is a Python script that validates a Bag, checking for missing files, extra files, and duplicate files. Parallel Retriever implements a simple Python-based wrapper around wget and rsync to optimize the transfer of content between locations through parallelization. It supports rsync, HTTP, and FTP transfers. VerifyIt is a shell script that verifies file checksums within a Bag manifest using parallel processes.•

Announcement

The Interim report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access is now available: http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf. It identifys systemic challenges to sustainable digital preservation based on current practices and funding structures. The final report, due later in 2009, will propose economic models and community scenarios to support long-term digital preservation and access. •

Minn. Initiative Brings Partners Together

Partners in the Model Technological and Social Architecture for the Preservation of State Government Digital Information project met in December to get updates on the latest project work, and to plan activities for the next phase of the project.

The project is lead by the Minnesota Historical Society, and is focused on preserving and making available state legislative information. The overall goal of the project is to create a framework that will be flexible enough for other states to use as a template to create their own methods of capturing and preserving legislative records. The meeting was the first opportunity to bring all the participating partners together to discuss early results.

The first round of site visits conducted throughout 2008 with the state partners, who include California, Illinois, Kansas, Tennessee and Vermont helped the project identify the main points of interest for each state. Common goals include concerns about the authenticity and accessibility of records, preservation, and digitization.

These goals will guide development of a system utilizing an XML wrapper and a core set of metadata elements to tag legislative data for legislative information exchange, Read more at: http://www.digitalpreservation.gov/news/2009/20090112news_article_mn_meeting.html•

Recent Meetings

• Heads of the initial eight NDIIPP partnerships met with Library staff to discuss ideas for continuing collaboration after 2009. http://www.digitalpreservation.gov/news/2008/20081230news_article_NDIIP-PPI_meeting.html

• Library of Congress staff and NDIIPP partners participated in a panel focusing on content transfer at the Digital Library Federation 2008 Fall Forum. http://www.digitalpreservation.gov/news/2008/20081230news_article_DLF_meeting.html