

ARTICLE
EXCERPTED
FROM:

ISQ

INFORMATION STANDARDS QUARTERLY

SPRING 2010 | VOL 22 | ISSUE 2 | ISSN 1041-0031

SPECIAL ISSUE: DIGITAL PRESERVATION

DIGITAL PRESERVATION
METADATA STANDARDS

TRUSTWORTHY
DIGITAL REPOSITORIES

UNIFIED DIGITAL
FORMATS REGISTRY

AUDIO-VISUAL
DIGITIZATION GUIDELINES

DIGITAL PRESERVATION
EDUCATION



← LOCKSS

Of the approaches discussed, the guide particularly recommends the use of the LOCKSS software to create Private LOCKSS Networks (PLNs), a model which the MetaArchive Cooperative has used for over six years for their own shared infrastructure.

Distributed Digital Preservation Guide from the MetaArchive Cooperative

Touted as being the first in a series of preservation publications, *A Guide to Distributed Digital Preservation*, produced by the MetaArchive Cooperative, contains a collection of articles by Cooperative members on the still-emerging field of using digital replication and distribution for preservation.

Of the approaches discussed, the guide particularly recommends the use of the LOCKSS software to create Private LOCKSS Networks (PLNs), a model which the MetaArchive Cooperative has used for over six years for their own shared infrastructure. The PLN model is used as the underlying framework and case example for the Guide's discussions of architecture; technical considerations; organizational considerations; content selection, preparation, and management; content ingest, monitoring, and recovery;

cache and network administration; and copyright practices—although much of the content is considered extensible to other distributed digital preservation solutions as well.

In providing the guide, the authors hope to forestall a trend towards outsourcing of digital collection management. "The central assertion of the MetaArchive Cooperative... is that cultural memory organizations can and should take responsibility for managing their digital collections, and that such institutions can realize many advantages in collaborative long term preservation and access strategies." A distributed strategy with institutional collaboration and investment supported by a robust technical infrastructure is the proposed alternative. ■

📄 Download the guide at: www.metaarchive.org/GDDP

Planets Survey Gauges Organizational Readiness for Digital Preservation

The Preservation and Long-Term Access Through Networked Services (Planets) Project has issued a white paper summarizing a market survey of over two hundred organizations, mainly European archives and libraries, investigating their digital preservation activities and needs.

Findings of the survey included:

- » The volume of digital content that organizations expect to archive will increase 25-fold over the next ten years. By 2019, 70% of survey respondents expect to hold over 100TB of content.
- » Within a decade, over 70% will need to preserve video, audio, databases, websites, and e-mail in addition to the current needs for documents and images.
- » The three most important capabilities of a digital preservation system were:
 - maintaining the authenticity, reliability, and integrity of records; checking that records have not been damaged; and planning the preservation of content to deal with technical obsolescence.
 - » Open-source and proprietary software are used equally by respondents, and often combined in the same solution.
 - » Respondents were much less interested in emulation than in migration as a preservation solution for technology obsolescence.
 - » Only 27% think that they have complete control over the file formats that they will accept and store in their digital archives.
 - » Compliance with metadata standards is regarded as fairly important, but there is less agreement on which standards. Dublin Core was the most popular (used by 51% of respondents), followed by MARC (31%) and ISAD(G) (28%).
- » Organizations are only starting to commit to funding digital preservation; just 47% have allocated a budget to it.
- » There is evidence that digital preservation is emerging as a profession in its own right; where previously the work was carried out by IT and preservation or curation staff, now it is starting to be carried out by specialists.

The white paper concludes with a summary of needed next steps including the importance of having a digital preservation policy and the need for more and better tools to automate the preservation process. ■

📄 The Planets white paper and a separate survey analysis report are available from: www.planets-project.eu/publications/

Report on Sustainable Economics for Long Term Preservation

For its final report, The Blue Ribbon Task Force on Sustainable Digital Preservation and Access has taken an economic perspective on all the resources—human, technical, and financial—that are needed to ensure that digital assets will be available for future use. *Sustainable Economics for a Digital Planet* identifies three requirements:

- 1 articulate a compelling value proposition;
- 2 provide clear incentives to preserve in the public interest;
- 3 define roles and responsibilities among stakeholders to ensure an ongoing and efficient flow of resources to preservation throughout the digital lifecycle.

National and international agencies, funders, and sponsors of data creation, stakeholder organizations, and individuals are all called upon to take particular actions to ensure preservation and access.

Four domains—scholarly discourse, research data, commercially owned cultural content, and collectively created web content—were analyzed for their sustainability risks and domain-specific recommendations were made. For example, in the domain of scholarly discourse, the Task Force recommended that “publishers reserving the right to preserve should partner with third-party archives or libraries to ensure long-term preservation.” For collectively produced web content, the Task Force suggests that “creators, contributors, and host sites could lower barriers to third-party archiving by using a default license to grant nonexclusive rights for archiving.”

The Blue Ribbon Task Force on Sustainable Digital Preservation and Access was created in late 2007 with funding from the National Science Foundation and The Andrew W. Mellon Foundation, in partnership with the Library of Congress, the Joint Information Systems Committee of the United Kingdom, the Council on Library and Information Resources, and the National Archives and Records Administration. ■

📄 All of the Task Force’s publications, including *Sustainable Economics for a Digital Planet*, are available for free download from: brtf.sdsc.edu/publications.html

Keeping Research Data Safe

Building on the Phase 1 report's cost model for the long-term preservation of research data, the Keeping Research Data Safe (KRDS) project has just published their Phase 2 report, which reports the results of testing and validating that cost model. Survey cost data was received on 13 collections and the cost information from four organizations—Archeology Data Service, National Digital Archive of Datasets, UK Data Archive, and University of Oxford—were analyzed in depth and presented in case studies. A benefits framework was also developed and illustrated with case studies from the National Crystallography Service at Southampton University and the UK Data Archive at the University of Essex.

Among the report's conclusions were:

- » The costs of acquisition/ingest and access are far greater than the costs of the archiving activities. Thus, it is likely that the largest potential cost benefits will come from the development of tools that support the ingest and access activities.
- » Once core fixed costs are in place (largely staff resources), increasing levels of economies of scale can be demonstrated as content is added.
- » The documentation of the dataset can be as beneficial to archive as the data itself. (In one example, the documentation was downloaded 10 times more often than the actual dataset.)
- » The OAIS reference model fits better with preservation services focused on data archives and institutional repositories. It is less ideal for focusing on "near-term preservation and curation work from a researcher perspective."
- » The benefits taxonomy has great potential for further development and implementation. [Ed. Note: While the study focused on datasets, the benefits taxonomy could be easily transferable to other types of preservation collections and activities.]

The KRDS2 study was funded by JISC with support from OCLC Research and the UK Data Archive. ■

ⓧ The full report can be downloaded from the KRDS2 webpage: www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx

JHOVE2 Beta Released

A new beta version of JHOVE2, the open-source Java framework for format-aware characterization of digital objects, has been released to the public. Characterization not only provides information about an object but can also function as a surrogate, which is especially useful in preservation environments. JHOVE2 uses the processes of identification, validation, feature extraction, and assessment to derive the characterization.

JHOVE2 is being designed as a next-generation improvement to the original JHOVE software, based on over four years of extensive use. While the original JHOVE assumed that one object was equivalent to one file and one format, JHOVE2 supports a single object having multiple files and formats. The redesigned architecture also configures the modules so they can be iteratively applied to each object. Among the other improvements are a plug-in interface, de-coupling the identification and validation, performance enhancements, and better error reporting.

This beta release has been provided to give interested users an early look at the new JHOVE2 architecture and APIs. While the processing modules are fully functional, there is limited format support at this time. Additional format modules will be added as they are completed.

The JHOVE2 project is a collaborative undertaking of the California Digital Library, Portico, and Stanford University, with funding from the Library of Congress as part of its National Digital Information Infrastructure Preservation Program. ■

ⓧ For further information and to download the beta release, visit the JHOVE2 website: <https://confluence.ucop.edu/display/JHOVE2Info/>

Codecs Primer for Archives

AudioVisual Preservation Solutions has published *A Primer on Codecs for Moving Image and Sound Archives: 10 Recommendations for Codec Selection and Management* by Chris Lacinak. In addition to providing introductory material on what encoding and compression are and how they work, the paper emphasizes that the choice of these schemes can impact the ability to preserve the digital object.

Since moving images and sound generally require a codec for the decoding process, the selection process is critical, whether dealing with born digital content, reformatting older content, or converting analog materials. Ten recommended approaches are explained: adoption, disclosure, transparency, external dependencies, documentation and metadata, pre-planning, maintenance, obsolescence monitoring, maintenance of the original, and avoidance of unnecessary transcoding or re-encoding.

There is no single "right" codec; each archive needs to make the decision as part of an overall preservation strategy. ■

ⓧ The Codecs Primer can be downloaded from: www.avpreserve.com/wp-content/uploads/2010/04/AVPS_Codec_Primer.pdf