

Cooperative Approaches to Digital Preservation: Panel

Martin Halbert
University Library
Emory University
Atlanta, GA 30322 USA
martin.halbert@emory.edu

Richard Pearce-Moses
Arizona State Library, Archives and
Public Records
Phoenix, AZ 85007 USA
rpm@lib.az.us

Tyler Walters
Georgia Institute of Technology
Atlanta, GA 30332 USA
tyler@gatech.edu

Aaron Trehub
RBD Library
Auburn University
Auburn University, AL 36849 USA
trehuaj@auburn.edu

Jonathan Crabtree
Odum Institute for Research in Social Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599 USA
jonathan_crabtree@unc.edu

ABSTRACT

In this panel, representatives from four archives – the MetaArchive Cooperative, Alabama Digital Preservation Network, Data-PASS, and the Persistent Digital Archives and Library System – discuss the versatility, low cost, and compelling benefits of using cooperative distributed digital preservation networks to safeguard categories of digital content that define our culture, identity, and history and that might otherwise be lost as a result of natural disaster, human error, or neglect.

Categories and Subject Descriptors

H.3.4 [Systems and Software]: Distributed Systems, Information Networks.

K.6.4 [System Management]: *Centralization/Decentralization*

General Terms

Management, Design, Economics, Reliability, Legal Aspects, Verification.

Keywords

Distributed Digital Preservation

1. INTRODUCTION

Participants in this panel will describe the forces that drove them to preserve their institutions' unique collections using a distributed digital preservation approach called PLNs (Private LOCKSS Networks), and the ways they were able to adapt their networks to meet their institutions' specific needs. The discussion will address the versatility, low cost, and benefits of using a distributed digital preservation network to safeguard categories of digital content that define our culture, identity, and history and

This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license. You are free to share this work (copy, distribute, and transmit) under the following conditions: attribution, non-commercial, and no derivative works. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>.

DigCCurr2009, April 1-3, 2009, Chapel Hill, NC, USA



that might otherwise be lost as a result of natural disaster, human error, or neglect.

Each panelist represents an institution deeply committed to safeguarding our cultural heritage for future generations. Researchers at each institution discovered the PLN to be a flexible, effective tool with which to accomplish their goals. Additionally, they discovered the inherent value of building networks with other like-minded institutions. Each panelist was able to encourage other libraries and archives to collaborate in building an infrastructure that would safely and reliably protect assets and records of national significance.

The panelists will describe how their networks differ from the others. Examples of these differences include:

- single-state vs. multi-state model
- different business strategies and pricing models
- technology and work flow integration decisions (including hardware needs and network management)
- types of materials being preserved (range from digital collections about the American South to electronic public records)

Distributed digital preservation networks built on LOCKSS' open-source technology have been growing in popularity among research and government institutions that wish to collect and preserve materials of high importance to specific communities.

2. TECHNICAL ISSUES

While the technologies behind PLNs are affordable and standard, much work has taken place to date to optimize these networks' performance. In some cases, new technologies have been created—as in the case of the MetaArchive Cooperative and its *conspicuum* database, which is now available as an open source module to be used with the LOCKSS software. Some of these networks are pushing technological limits in regards to the size and number of collections (or “archival units”, in the LOCKSS vernacular) that are being harvested. Many other issues will be discussed by the panel, including issues relating to indexing network holdings, ingest challenges, explorations in format migration and emulation, networking challenges such as open port

management, overseeing and ensuring the successful multiple node harvest of collections, and the overall aligning of technologies such as hardware, operating systems, and middleware to ensure that nodes are successful in their interactions and communications.

3. SUSTAINABILITY ISSUES

The cornerstone reason for founding the organizations behind PLNs is to sustain the preservation network and its capacity to maintain and manage its content over long periods of time. The experiences of the PLNs on this panel will be discussed with a view to how they have approached the need to sustain their organizations to date. Approaches range from making use of existing state consortia (ADPNet), to grafting distributed digital preservation management onto multi-institutional consortia (Data-PASS), to creating new multi-institutional multi-state consortia (PeDALS), to creating a brand new cooperative with its own independent management entity (MetaArchive Cooperative). Financial issues are key to sustainability, and each cooperative will expand upon its approach to garnering adequate finances to sustain preservation activity, including working to maintain extremely low costs to participation, managing the number of members, and devising an annual dues structure to support and sustain cooperative business.

4. ORGANIZATIONAL ISSUES

Each cooperative has faced specific challenges in initiating, growing, and maintaining its multi-institutional organization. Foundational issues have been addressed, such as establishing governance and related documents, defining rights and responsibilities of the participating institutions, financial arrangements, technology specifications, and processes for adding new as well as withdrawing members and collections. Panelists will share their cooperatives' experiences and the best practices they have derived from their organization-building activities. The emerging theme from this work is the development of formalizing trust relationships in new digital preservation federations like those being represented in this panel session.

5. REPOSITORY ISSUES

PLNs are beginning to be explored with regard to preserving the content of popular, open source, institutional repository systems such as DSpace and EPrints. Consortia in the U.S., Italy, the U.K., and South Africa have expressed interest in or are currently exploring the establishment of PLNs to harvest and preserve their repository content. One of the largest DSpace repositories in the world is Georgia Tech's SMARTech repository, which is being harvested as part of the MetaArchive Cooperative's network. The MetaArchive – DSpace experience will be examined for early lessons learned. Panel members will discuss how the process is being refined, anticipate improvements that can be made with new protocols becoming available, and share best practices currently being documented for LOCKSS harvesting of repository systems.

6. ACKNOWLEDGMENTS

We would like to acknowledge the support of the Library of Congress National Digital Information Infrastructure and Preservation Program (NDIIPP), Institute of Museum and Library Services (IMLS), and the National Archives and Records Administration National Historical Publications and Records Commission (NHPRC).

7. REFERENCES

- [1] MetaArchive Cooperative website.
<http://www.MetaArchive.org>
- [2] Data-PASS website.
<http://www.icpsr.umich.edu/DATAPASS/>
- [3] Alabama Digital Preservation Network [ADPNet] website.
<http://www.adpn.org/>
- [4] Persistent Digital Archives and Library System (PeDALS) Project website. <http://rpm.lib.az.us/pedals/>
- [5] U.S. National Digital Information Infrastructure and Preservation Program [NDIIPP] website.
<http://www.digitalpreservation.gov>