**Peer Reviewed**

**Title:**
Distributed Digital Preservation: Technical, Sustainability, and Organizational Developments

**Author:**
Walters, Tyler, Georgia Institute of Technology
Bishoff, Liz, BCR (Bibliographic Research Center)
Gore, Emily B., Clemson University
Jordan, Mark, Simon Fraser University
Wilson, Thomas C, University of Alabama - Tuscaloosa

**Publication Date:**
10-05-2009

**Series:**
iPRES 2009: the Sixth International Conference on Preservation of Digital Objects

**Publication Info:**
iPRES 2009: the Sixth International Conference on Preservation of Digital Objects, California Digital Library, UC Office of the President

**Permalink:**
http://www.escholarship.org/uc/item/38g232wc

**Multimedia URL:**
http://www.cdlib.org/services/uc3/iPres/video.html?file=ipres/GorePanel&title=Panel%20Discussion%3A%20Distributed%20Digital%20Preservation%3A%20Technical%2C%20Sustainability%20and%20Organizational%20Developments

**Abstract:**
Representatives from a variety of distributed digital preservation initiatives will serve as panel members and discuss the technical adaptability, economics, and functionally compelling benefits of using cooperative distributed digital preservation networks to preserve the vast array of at-risk digital content produced by our societies and their institutions.

**Supporting material:**
Presentation

# iPRES 2009

THE SIXTH INTERNATIONAL CONFERENCE ON THE PRESERVATION OF DIGITAL OBJECTS

# Proceedings

October 5-6, 2009
Mission Bay Conference Center
San Francisco, California


California Digital Library

# Distributed Digital Preservation: Technical, Sustainability, and Organizational Developments

**Tyler O. Walters**
Georgia Institute of Technology
704 Cherry Street
Atlanta, GA 30332
Tyler@gatech.edu

**Liz Bishoff**
BCR (Bibliographic Research Center)
14394 E. Evans
Aurora, CO 80014-1408
LBishoff@bcr.org

**Emily B. Gore**
Clemson University
S. Palmetto Blvd.
Campus Box 343001
Clemson, SC 29634-3001
egore@clemson.edu

**Mark Jordan**
Simon Fraser University
Burnaby, British Columbia, V5A 1S6
Canada
mjordan@sfu.ca

**Thomas C. Wilson**
University of Alabama
Box 870266
Tuscaloosa, AL 35487-0266
tcwilson@ua.edu

**Abstract**

Representatives from a variety of distributed digital preservation initiatives will serve as panel members and discuss the technical adaptability, economics, and functionally compelling benefits of using cooperative distributed digital preservation networks to preserve the vast array of at-risk digital content produced by our societies and their institutions.

## Introduction

To date, our community's current digital preservation practices largely consist of geographically and institutionally homogeneous replication of content by one institution. However, this approach leaves content vulnerable to the vagaries of one institution's technical infrastructure and more susceptible to man-made and natural disasters. A network of geographically and institutionally diverse digital repositories adhering to best practices, such as those set forth in RLG's *Trusted Digital Repositories: Attributes and Responsibilities* and the *Reference Model for an Open Archival Information System*, greatly reduces these threats.[1]

Private LOCKSS networks (PLNs) have worked to establish successful strategies for archiving copies of content in secure, geographically distributed locations. Using leading software for distributed digital replication (the LOCKSS system from Stanford University), PLNs have developed a viable means to distribute, replicate, and monitor digital collections. [2] This approach provides the geographic and institutional diversity required to significantly reduce the risk of digital collections loss.

Distributed digital preservation networks built on the LOCKSS open-source technology have been growing steadily in popularity among research and government institutions with a mandate to collect and preserve information of significance to their communities. The PLN approach is recognized as a versatile and low-cost one that can protect from disaster at any one geographic location the critical digital content that documents our culture, public institutions, and private organizations. This paper offers a range of perspectives and experiences in distributed digital preservation from an international PLN-based consortia − the MetaArchive Cooperative; a state-level organization − the Alabama Digital Preservation Network (ADPNet); and a large, regional organization in Canada – the Council of Pacific and Prairie University Libraries PLN (COPPUL PLN). It portrays how their participating institutions embarked on the distributed digital preservation approach and how their work meets the current digital preservation needs of



[1] *Trusted Digital Repositories: Attributes and Responsibilities*, RLG-OCLC, May 2002, Mountain View, CA, (http://www.rlg.org/en/pdfs/repositories.pdf) and *Reference Model for an Open Archival Information System (OAIS)*, Blue Book, Issue 1, January 2002, ISO 14721:2003, (http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf ).

[2] See the LOCKSS Program (http://lockss.stanford.edu/).

their institutions. In this paper, we illustrate that new digital preservation federations are resulting from compatible institutions reaching out, identifying each other, and pooling resources to reach common goals. The growth and expansion of existing PLNs can be seen through the BCR-led MetaArchive West project, which involves seven institutions that are new to PLNs. Individual institutions also are joining these efforts and their decisions to join are examined through the example of Clemson University joining the MetaArchive Cooperative. Further, this paper describes and analyzes aspects of the organizational networks represented, such as business plans and strategies, technologies, sustainability issues, and other organizational challenges.

## The MetaArchive Cooperative

The MetaArchive Cooperative was the first Private LOCKSS Network, established in 2004 as one of eight original digital preservation partnerships with the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP). It stemmed from the collaborative initiatives of six university research libraries and archives -- Emory University, Georgia Tech, Virginia Tech, Florida State University, Auburn University, and the University of Louisville. The MetaArchive has promulgated a successful cooperative charter and membership agreement that form the core of the multi-institutional relationships. By 2009, the MetaArchive has grown from the original six to 14 member institutions, adding the following: Boston College, Clemson University, the Folger Shakespeare Library, PUC Rio de Janiero, Penn State, Rice University, the University of Hull (UK), and the University of South Carolina. There are currently more than 40 other institutions who have expressed interested in joining the MetaArchive Cooperative.

Once initiated, the Cooperative's leadership, which forms its steering committee, began articulating a sustainability model for the Cooperative, consisting of components that address sustaining the organizational model, its technology model, its economic model, and finally the collections themselves. The digital sustainability model has guided specific actions taken by the MetaArchive Cooperative to further its goals and operations.[3]

**Organizational Management and Sustainability.** The MetaArchive Cooperative's organizational sustainability strategy has been shaped by its experiences in inter-organizational collaboration and learning. Working through NDIIPP has given the MetaArchive contact with many institutions and organizations working on digital preservation. The National Historical Publications and Records Commission (NHPRC) of the National Archives and Records Administration have also supported the MetaArchive financially, which has facilitated the MetaArchive's transformation from project to business. Of course, NDIIPP and NHPRC have contributed greatly to the MetaArchive's organizational sustainability and have given it support that allows it to plan strategically for the long-term while it builds its partner base today. The mixed revenue model of grant funding, membership fees, and consulting fees also has contributed to the Cooperative's success and stability. The other major point of organizational sustainability is the member institutions themselves. The organizational plans for the MetaArchive are embedded within its charter. The charter establishes the roles, rights, and responsibilities of the Cooperative's partners. There are two main member categories in operation currently. They are:

*Sustaining Members:*

These are the leaders of the MetaArchive Cooperative, its governance, member services, and strategies. Sustaining Members also operate preservation nodes (i.e. storage servers on the MetaArchive network) and they preserve their content on them. These nodes also preserve content coming from all Cooperative members. The Sustaining Members test, develop, and maintain the Cooperative's hardware, software, and network connectivity. They provide a high level of input, which creates, in effect, a distributed staffing model that includes an intellectual property attorney, many subject specialists and curators, and a large number of software engineers and systems administrators. The member institutions' staffs do more than take care of their own node; they work to support the overall network and organization as well.

*Preservation Members:*

These members perform two essential roles: they participate in the MetaArchive networks by operating preservation nodes and they contribute content that is preserved on them. They do not have a decision-making role in the governance and business operations of the Cooperative. They also do not bear additional responsibility for supporting the technical and organizational work of the Cooperative through staff contributions.

The MetaArchive Cooperative Steering Committee is the organization's decision-making body. It is comprised of one representative from each of the Sustaining Members. Further policy, strategy, and operational

[3] Walters, Tyler O., *"Digital Sustainability: Weaving a Tapestry of Interdependency to Advance Digital Programs,"* in Strategies for Sustaining Digital Libraries. (Atlanta: Emory University) Martin Halbert and Katherine Skinner, editors. 2008. http://www.e-booksdirectory.com/details.php?ebook=1223

decisions are carried out in three other committees: Content, Preservation, and Technical. Individuals representing all member institutions may serve terms on these committees. Member services include digital preservation network management, digital collection disaster recovery, digital preservation network consulting and training, and a future service for format migration.

A significant characteristic of the MetaArchive Cooperative's organizational sustainability strategy is its management by an independent non-profit organization called the Educopia Institute, Inc. The Cooperative decided that the "single host institution" model left it vulnerable to the changes in direction a single institution might make. Also, no single consortium currently serves all the MetaArchive's institutions. Educopia provides administrative and fiscal oversight to the Cooperative and consists of a Board of Directors that draw upon professionals from other institutions that are not members of the Cooperative, which further diversifies its base of management and input.[4] The intention is to provide a low-cost, low-overhead entity that oversees cyberinfrastructure and cyberlearning projects to advance digital learning, scholarship, and research in higher education and with cultural memory organizations. The emerging organizational model has also led to a solid base for economic sustainability, namely, a diverse set of revenue streams for the Cooperative. To date, these have been grants and contracts from federal funding agencies, member dues, individual institutional contributions, and consulting/training fees.

**Technical Achievements.** There has been much adaptation of LOCKSS to optimize the performance of PLNs. The MetaArchive Cooperative has worked with the LOCKSS development team at Stanford on new tools such as the MetaArchive's Conspectus database and a new and improved LOCKSS Cache Manager. The Conspectus is a database product that sits in front of a LOCKSS network and manages a robust collection of metadata for the network's digital collections. It was developed by the MetaArchive Cooperative and is available as an open source offering to the LOCKSS community. The new LOCKSS Cache Manager tool seeks to accommodate the particulars of a Private LOCKSS Network (as distinct from the public LOCKSS network model for which the LOCKSS software was originally designed). The Cache Manager is the network monitoring tool used in the LOCKSS software and is comprised of four main network monitoring components: 1) caches, 2) collections, 3) archival units, and 4) disks. Some of the desired improvements are to sort archival units by institution (not a normal LOCKSS feature), improve the display of collections and archival units, and to provide reporting features by Archive and by Institution. Essentially, these are useful tools in a PLN which is managing a wide array of digital materials, but done so by and for a specific institution. Such institution-specific issues are generally not of concern in the original general/public network, yet are of central concern in a PLN framework.

Content transfer involving LOCKSS is another technological issue that the MetaArchive has been investigating. The focus has been to improve content transfer to PLNs from digital asset management systems (DAMS) like DSpace and between preservation systems like LOCKSS and grid-based systems using SRB and/or iRODS. Training LOCKSS to harvest data from DAMS is problematic because it is designed to harvest content from static web sites, not dynamic content. Georgia Tech and the MetaArchive have adopted an approach to harvesting source content and metadata using DSpace's native OAI and METS export functionality. LOCKSS pulls data from DSpace by making OAI requests for METS XML, which delivers URIs for associated content and metadata. The LOCKSS daemon and generic plugin reads DSpace OAI/METS output. Testing has proven successful, although Georgia Tech anticipates the possibility of needing custom Java plugins for specific types of complex, multi-object content. This harvesting approach provides a reliable data source for reconstructing repositories.

There also has been ongoing work supported by the National Historical Publications and Records Commission (NHPRC) in data transfer between preservation systems with the MetaArchive and the San Diego Supercomputer Center's Chronopolis digital preservation program, which is currently based on SRB. The BagIt technology for content transfer, developed by the Library of Congress and the California Digital Library, has been the focus of the studies thus far. Recently, a joint grant proposal from the MetaArchive and Chronopolis has been submitted to the NSF INTEROP program to build transfer tools in support of BagIt, the native LOCKSS technologies, and the data transfer approaches of iRODS, which Chronopolis plans to use in the future. Moreover, projects like those described above demonstrate that there is much activity in place with the aim being to improve distributed digital preservation.

## The Council of Prairie and Pacific University Libraries (COPPUL) Private LOCKSS Network

**Organizational Management.** Another major PLN effort is the COPPUL PLN. COPPUL is a consortium of 21 university libraries located in Manitoba, Saskatchewan, Alberta and British Columbia who participate in resource sharing, collective purchasing, reciprocal document delivery, and other activities. As of September 2009, eight

[4] See http://educopia.org

of these institutions are participating in the COPPUL Private LOCKSS Network.

One organizational challenge that the COPPUL PLN faces is that some of its members are very large by Canadian standards, while others are quite small. Large in this context includes the University of British Columbia (with a student population of over 50,000) and the University of Alberta (student population of over 35,000); the smallest is University of Winnipeg (student population of approximately 9,000). The most obvious implication of having members that differ so widely in size is ensuring that the cost of membership in the PLN remains equitable. The principal cost is annual membership in the LOCKSS Alliance (there is no membership fee for the COPPUL PLN itself), which is based on institution size and therefore mitigates this problem to a certain degree. Other costs incurred by participating in a LOCKSS PLN, however, tend to be roughly the same for all members, so smaller institutions pay more in relation to their overall budgets. These other costs include the hardware used for the LOCKSS servers, staff time required to participate in the PLN (for selection of material, integrating LOCKSS into existing or new preservation strategies, etc.), and miscellaneous costs such as travel to meetings.

Another organizational challenge particular to COPPUL is the PLN's relationship with Synergies, a "not-for-profit platform for the publication and the dissemination of research results in social sciences and humanities published in Canada."[5] Although Synergies' use of LOCKSS is currently (as of September 2009) still in the planning stages, it is likely that Synergies journals will be preserved using some combination of the public LOCKSS network and a Synergies PLN. Any institution that belongs to the LOCKSS Alliance may be a member of multiple networks, which means that institutions will not have to pay a separate Alliance membership fee to belong to both the COPPUL and Synergies PLNs or the public LOCKSS network. Nonetheless, questions of overlap between the networks remain outstanding and are being studied.

**Sustainability.** The main issue surrounding the sustainability of a PLN in a library consortium the size and nature of COPPUL is ensuring that enough members participate in it. The technical architecture of a PLN requires that it contains at least six nodes, since the integrity of the preserved content is verified using a voting mechanism based on this minimum. If two of the eight current members were to drop out for any reason, and were not replaced by new members, the PLN would be operating at its minimum size. In order for the COPPUL PLN to be sustainable, between a third and a quarter of COPPUL members need to participate. Given frequent cutbacks to university library budgets, this is a relatively high proportion.

**Technical Achievements.** The COPPUL PLN is the first PLN to automate the harvesting of OJS (Open Journal Systems) content for preservation. OJS is an open-source journal management platform developed and supported by the Public Knowledge Project.[6] Currently there are over 3000 journals using OJS. Many of the participants in the COPPUL PLN host OJS journals in support of scholarly publishing on their campuses or for scholarly societies, professional associations, and other academic publishers. The libraries that host these journals feel obligated to ensure that the content remains accessible if their OJS servers go offline for any reason. Preserving their journals in the PLN was an obvious strategy. The development of the LOCKSS plugin that harvests OJS content not only allows COPPUL libraries to preserve locally hosted OJS journals in their PLN, it expands the capabilities of the network. If an OJS journal is nominated for inclusion in the public network, the necessary plugin will have been created.

The COPPUL PLN, like most PLNs, preserves content that is unique to the network members and is not preserved in the public LOCKSS network. In COPPUL's case, this local content currently includes the electronic theses collection from the University of Saskatchewan, the Grande Prairie Historical Photos Collection from the University of Calgary, and the Editorial Cartoons Collection from Simon Fraser University. Determining the best way to preserve local content in the PLN offers some interesting opportunities for developing institutional preservation practices. For example, in the case of SFU's Editorial Cartoons collection, local staff refined their practice of putting the master TIFF file and various types of metadata for each cartoon into a BagIt package[7] that is optimized for local digitization workflow and submission to, and extraction from, the PLN. This technique could be applied to any institution's digital assets management workflow, however. As mentioned earlier in regards to the MetaArchive, there have been many important technology developments that have advanced the functionality and overall success of PLNs.

## The Alabama Digital Preservation Network (ADPNet)

The Alabama Digital Preservation Network (ADPNet),[8] a program of the Network of Alabama

[5] See http://www.synergiescanada.org/page/publishers

[6] See http://pkp.sfu.ca

[7] http://tools.ietf.org/html/draft-kunze-bagit-04

[8] See http://www.adpn.org/

Academic Libraries, is a distributed digital preservation network for locally created digital content. It represents a low-cost digital preservation solution for academic institutions, state agencies, and cultural heritage organizations in Alabama. ADPNet is another PLN, meaning that the archived content is accessible only to the members and only if it is needed to restore lost content. Any Alabama-based cultural heritage organization with publicly-available digital assets whose activities and objectives are consistent with ADPNet's mission and principles may join the network.

**Organizational Management.** Originally funded in 2006 by a two-year National Leadership Grant from the Institute of Museum and Library Services (IMLS), ADPNet is now self-sustaining. The current membership includes: Alabama Department of Archives and History, Auburn University, Spring Hill College, Troy University, University of Alabama, University of Alabama at Birmingham, and University of North Alabama. Members are required to maintain an appropriately configured LOCKSS server, contribute digital content to the network and harvest digital content from other member institutions, join the LOCKSS Alliance, determine their rights to preserve content prior to submitting it to network, and hold the network and other members harmless. The costs for selecting and digitizing material, systems administration, and equipment upgrades are borne by the members individually. There is no ADPNet membership fee.

ADPNet is governed by a steering committee that represents the members, oversees the management and operation of the network, sets general policy, reviews and approves requests to expand the network's storage capacity, and reviews and approves applications for membership. The committee consists of one voting representative appointed by each member organization. The ADPNet Technical Policy Committee reviews the network's capacity and technical specifications and makes recommendations related to the network's hardware and software. The participants represent widely variant organizations in size, type, and governance. This attribute is at once a strength and a challenge. The diversity of participating organizations adds to the value and vitality of the network. At the same time, widely varying needs and resources present an opportunity to explore avenues of mutual benefit that neither inhibit the members who move forward at a rapid pace, nor present barriers to members who have fewer assets to preserve.

ADPNet is guided by the following operating principles:

- Mutual commitment to long-term preservation of critical cultural heritage content;
- Collaboration to adopt policies and procedures that will sustain the Network to the mutual benefit of its partners and content contributors;
- Commitment to keeping overhead low and achieving low-cost preservation strategies;
- A cooperative, robust, and decentralized peer-to-peer approach to selecting content of shared value, and mutual support of content with a particular value to individual institutions;
- Application of LOCKSS software, as the principal system for distributing copies of replicated content in secure, distributed locations over time;
- Wide applicability to a range of institutions and digital content;
- Commitment to storage and maintenance in migratable formats and data structures;
- Commitment to high standards for metadata and content; and
- Ongoing exploration of projects to advance digital preservation.

**Sustainability and Technical Achievements.** Because the bar for participation is relatively low and the costs for ongoing membership are reasonable, the long-term viability of ADPNet looks promising. The network provides a useful and proven option for Alabama institutions to entrust with important digital collections. The geographic distribution of participants from around the state reduces concerns regarding the reach of natural disasters.

From ADPNet's perspective, the LOCKSS software places few demands on server hardware. Very low-cost hardware can be used and ADPNet members, thus far, have chosen to purchase similar hardware for each site, though this is not a requirement. ADPNet is designed to be simple (technically and administratively) and cheap. Currently, the network preserves 46 digital collections totaling about 400 GB of harvested data. The current network capacity is 1.5 TB, but will soon be raised to 8 TB.

There are several future challenges to address. These include items like network scalability, preservation metadata types and formats, the varying needs and resources of the membership, expanding the membership during economically challenging times, and finding partners with greater geographic diversity. The ADPNet currently is in conversations with COPPUL and the

Arizona-based PeDALS PLN[9] to achieve greater geographic dispersion.

## The Challenge of Expanding Distributed Digital Preservation Federations: The *MetaArchive West* Project

Now that several PLN organizations have been established, the next step coming in their maturation process is their expansion as they invite and bring in new institutional members. Cultural heritage institutions in the American West have been digitizing content since the mid-1990's, and therefore, have been seen as likely candidate institutions to join a PLN to preserve the digital collections that they have been building so robustly. Today, libraries, archives, museums and historical societies provide access to digital objects including millions of pages of historic newspapers; hundreds of thousands of historic photographs, maps, and text; and terabytes of digital audio and video. Much of this content has been funded by state and federal grants, and millions of dollars in local match. Those who created the content for access however have paid little attention to the issue of assuring long term accessibility.

Western cultural heritage institutions have expertise in physical preservation practices, but they have yet to develop policies, procedures and solutions to preserve their digital collections. Heritage Preservation's *A Public Trust at Risk: the Heritage Health Index Report on the State of America's Collections* noted that on average less than 30% of institutions have an executable emergency plan and only 17% of the institutions have emergency plans with staff trained to carry out the plan.[10] The Northeast Document Conservation Center's 2005 Survey found that while institutions are actively creating and acquiring digital content, they are not addressing the issue of long term digital preservation.[11] While the Heritage Preservation survey didn't analyze efforts in *digital* preservation, one can expect that efforts have been no better in digital than in traditional preservation. The situation documented in these reports shows that digital collections are vulnerable to mismanagement, file corruption, disc failure, and natural and man-made disasters.

The *MetaArchive West* project is designed to create a culture of digital preservation among the partners and will demonstrate a model for collaborative digital preservation that can meet the goal of long term access. In this project, Bibliographical Center for Research and the MetaArchive Cooperative will provide a PLN solution[12] for libraries, archives and museums in the West. *MetaArchive West* will implement seven new MetaArchive nodes over a two year grant period and grant support for the initiative is being sought from the Institute of Museum and Library Services (IMLS). The seven partners (BCR, Oregon State University Library, Montana State Library, Colorado Alliance for Research Libraries (Alliance), University of Utah Library, Rice University Library and Georgia Institute of Technology Library) will run MetaArchive nodes (i.e. servers) to extend the MetaArchive preservation network. Two additional partners, Denver Museum of Nature and Science Library and the National Center for Atmospheric Research Library, will contribute Western digital culture content to the network.[13]

*Why the MetaArchive Cooperative?* Project participants planned and prepared for the organizational and technological challenges of distributed digital preservation. They evaluated several options for preservation solution development including initiating a local solution, choosing a not-for-profit or commercial vendor solution, implementing a state or regional Private LOCKSS Network, or building on an existing Private LOCKSS network. After the analysis, BCR in working with the Western cultural heritage institutions chose to join an existing distributed digital preservation -- the MetaArchive Cooperative. The Western consortium came to this decision because partnering with the MetaArchive Cooperative met the following assessment criteria:

- enables active participation in the management and implementation of a digital preservation solution
- engages with an established preservation solution, including existing preservation policies and plans

[9] See http://www.pedalspreservation.org/Default.aspx)

[10] http://www.heritagepreservation.org/HHI/HHIsummary.pdf

[11] Clareson, Tom. "NEDCC Survey and Colloquium Explore Digitization and Digital Preservation Policies and Practices" *RLG DigiNews*, 10:1 (February 2006). Available at: http://www.rlg.org/en/page.php?Page_ID=20894#article1.

[12] MetaArchive's (http://metaarchive.org) LOCKSS-based network (http://lockss.org) enables institutions to create, monitor, and maintain their own preservation service by 1) making copies of files and distributing them across multiple, geographically distinct servers, or "MetaArchive nodes"; 2) constantly monitoring each file across those distributed MetaArchive nodes for any change in status that may arise due to natural disasters, technical failures, or bit rot; and 3) repairing damaged files within the network.

[13] The Denver Museum of Nature and Science and NCAR are joining MetaArchive as "Contributing Members." They will contribute content *without* operating MetaArchive nodes.

- provides a technical infrastructure that is readily extensible
- uses *distributed* digital preservation, maintaining a geographically dispersed solution
- is deliberately decentralized in its technical and organizational frameworks
- is a network that runs its own technical infrastructure in a self-sufficient manner[14]
- has a scalable organizational structure that allows for the incorporation of new partners and new collections, and
- provides a low-cost solution that meets the financial needs of most cultural heritage institutions.

The *MetaArchive West* project will become a testbed for a significant expansion of an existing distributed digital preservation federation. It will introduce new components to the MetaArchive Cooperative to scale up its operations and enhance its sustainability, including a training program, additional tools to ingest content from leading repository platforms, and a digital preservation readiness assessment program. The *MetaArchive West* project also will test a variety of components of the MetaArchive's existing preservation plan by expanding the plan to incorporate non-research libraries, consortia, independent science libraries, and museums. The MetaArchive network is viewed by the Western consortium as a trustworthy, reliable, and sustainable repository capable of handling the full range of digital materials held by large and small cultural heritage institutions.[15] The project will use the *Trustworthy Repositories Audit & Certification: Criteria and Checklist* (TRAC)[16], to guide its development. The project has identified specific components of TRAC as a measurement of success, including ones in the area of professional development, repository scalability, documentation, assessment, and software technology.

[14] Several other known PLNs, including DataPass, the Arizona PeDALs network, and the Alabama ADPnet, rely on the Stanford University LOCKSS team to operate major components of their networks. These networks are highly dependent on LOCKSS for their daily operations.

[15] MetaArchive's network is format agnostic, it can provide at least bit-level preservation for any existing file type.

[16] *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Chicago, IL. Center for Research Libraries, March, 2007. http://www.crl.edu/content.asp?l1=l3&l2=58&l3=162&l4=91

## Expanding Distributed Digital Preservation Federations and Institution-level Decision-making: Clemson University

The Clemson University Libraries started its digital initiatives program only recently, in late 2007. However, this late start date has allowed Clemson the chance to learn from the other institutions coming before it who have built active and successful digital initiatives programs. Perhaps the most important lesson learned has been that successful digital initiative programs should include a sound digital preservation program to provide for the long-term access of its digital objects.

With the lessons in mind, a plan and the search for a systems solution to preserve Clemson's growing digital collection commenced in 2008. The Clemson Libraries was already a member of the LOCKSS Alliance which supports the public LOCKSS network, and hence, was aware of the concept of distributed preservation and the LOCKSS implementation of it. When Clemson began researching further the digital preservation solutions available that supported a similarly distributed approach, it found the MetaArchive Cooperative, which was founded in the U.S. Southeast by neighboring institutions only a few hours drive away. The original MetaArchive network focuses on Southern Digital Culture as a subject theme, which relates to U.S. Southern and South Carolina cultural themes found in many of Clemson's digital collections. Given these factors, it was logical for Clemson to discuss the parameters of joining the MetaArchive Cooperative.

There were other factors taken into consideration when the decision was made to join the MetaArchive Cooperative. Two of them were the research and development focus of the Cooperative and its successful organizational structure. It was important to Clemson to join an organization where it would become an active partner by being a decision-maker in the endeavor and build its own skills and experiences in digital preservation, as opposed to purchasing vendor-based services where the skill- and experience-building reside mostly on the vendor side of the relationship. Moreover, Clemson made the early decision to not only preserve digital collections, but also be part of a development effort to forge potential partnerships to further affordable, manageable digital preservation solutions. Since joining in 2009, Clemson has participated as a full member in shaping the MetaArchive Cooperative's use of the Amazon cloud environment to host its central resources. Clemson also is helping the original members build an effective training program to transfer knowledge of its LOCKSS-based practices to new member institutions. Part of the new training program's improvements will be to communicate as effectively as possible about the required knowledge of staff to implement and maintain a LOCKSS node. Requirements about staff knowledge and

time commitments will be updated based upon the last few years of experience and members in training will examine their capacity to apply the resources required to manage an active PLN node. Clemson feels it is meeting its early digital preservation objectives by choosing the cooperative approach, sharing in the decision-making, ownership, and development of a multi-institutional digital preservation solution.

## Conclusion

Distributed digital preservation federations such as private LOCKSS networks have been developing steadily over the past five years as a stable and economically viable digital preservation strategy. Each of the PLNs examined have initiated somewhat different approaches to their management structure, sustainability planning, and technology development activities. However, they all share the same mission of building successful means to preserve digital assets of scholarly, research, and cultural value. Major trends in PLN-related work have been to affect their interoperability with other preservation systems, digital asset management systems, and content production systems. Creative, yet manageable organizational structures and flexible, fiscally diverse sustainability models have been advanced as well. PLNs are growing in number, international in reach, and foreshadow a significant impact on the shape of future digital preservation organizations.

## References


[1] MetaArchive Cooperative website. http://www.MetaArchive.org

[2] Alabama Digital Preservation Network [ADPNet] website. http://www.adpn.org/

[3] Council of Prairie and Pacific University Libraries website.

http://www.coppul.ca/index.html
http://coppullockssgroup.pbworks.com/

[4] Persistent Digital Archives and Library System (PeDALS) Project website. http://rpm.lib.az.us/pedals/

[5] Data-PASS website. http://www.icpsr.umich.edu/DATAPASS/

[6] U.S. National Digital Information Infrastructure and Preservation Program [NDIIPP] website. http://www.digitalpreservation.gov