

Turning Silos into Bridges: Optimizing Local Digital Library Programs for Collaboration and Beyond

Robert H. McDonald; Florida State University; Tallahassee, Florida/USA

Charles F. Thomas; Florida Center for Library Automation; Gainesville, Florida/USA

Abstract

Florida State University Libraries' Digital Library Program is one of six founding members of the MetaArchive of Southern Digital Culture <<http://www.metaarchive.org>>. This partnership with the research libraries of Auburn University, Emory University, the Georgia Institute of Technology, the University of Louisville, and the Virginia Polytechnic and State University is funded under the auspices of the Library of Congress National Digital Information Infrastructure and Preservation Program to build a collaborative digital preservation network of Southern cultural heritage materials. The partnership was established to explore a new model for distributed, collaborative digital archiving and preservation. The new model spans multiple state jurisdictions, and includes public, private and governmental participants. Joining this very large and distributed partnership made the FSU Digital Library realize that with collaboration comes growth. Since 2003, the FSU Libraries have considered the questions, "What must be in place for a prototypical digital library program to engage in multiple large-scale collaborations?" and "How can new and or established digital library programs create a model workflow that will support collaborative efforts with as little added work as possible?" This paper considers those questions in the context of the FSU-NDIIPP partnership, and suggests evaluative criteria for building bridges and opening silos.

FSU as Prototypical Digital Library Program

Building upon earlier small digital initiatives and collaborations with statewide partners, the Florida State University Libraries established a formal digital library program in 2003. The earlier initiatives had produced a legacy of several thousand electronic files, each with its accompanying demands for description, management, storage and nurturing that are necessary to make digital content usable and persistent. The very earliest months of the FSU Digital Library were dedicated, primarily, to establishing intellectual and physical control of these legacies, and turning them into collections that later could be leveraged as institutional digital assets. Much of the activity during those early months was exploratory and experimental, and focused on very narrow local objectives.

Since then, FSU's Digital Library program has progressed well beyond the "skunk works" stage described in Greenstein & Thorin's (2001) *The Digital Library: A Biography*. The program has (a) secured dedicated staffing and budgeting, (b) established and disseminated policies on services, standards, and cost recovery mechanisms, (c) purchased and implemented enterprise systems to support digital collections, a campus institutional repository, an ETD program for the campus, and other forms of digital publishing, (d) instituted, refined and documented workflows for digital content and metadata production, and (e) made activities such as

grant-writing and training a routine, predictable part of its annual operation cycle. All of these accomplishments are key indicators of progress along a continuum of growth. In fact, the FSU Digital Library is a typical, or *prototypical*, example of a digital library program establishing and solidifying itself.

Growth and engagement with a wider world of other digital library programs has resulted in several collaborations and partnerships. One sphere of collaboration is an increasingly sophisticated relationship with statewide partners who employ a mixture of shared technical infrastructure and wholly local systems. Together, these statewide partners in Florida have created large shared digital collections such as the Florida Heritage Project, and are working together on other joint ventures. These state partnerships are centrally coordinated and have required that all participants support some core standards for metadata and digital formats. From the perspective of one local participant like the FSU Digital Library, this has required supporting those shared standards and interoperability, while also continuing to develop systems and standards that are beyond the scope of the statewide partnership, yet reflect the emerging and changing needs of the FSU campus, which also is yielding some new partnerships for the FSU Libraries.

Another important partnership for the FSU Digital Library is its involvement in the collaborative MetaArchive of Southern Digital Culture. This partnership, described below, is important to the FSU Libraries because it focuses on the more advanced and substantially more worrisome challenges that a growing digital library program encounters. Digital preservation is foremost among these challenges, and is unfortunately so intimidating that the FSU Libraries, as a digital library, had fallen into the "procrastination as digital preservation strategy" mindset that plagues many other programs. Fortunately, the MetaArchive partnership presented new opportunities for a more active role in addressing digital preservation and other next-generation challenges for digital libraries. As this paper demonstrates, the MetaArchive experience also has encouraged the FSU Libraries to become more proactive in some other areas, including adopting a more comprehensive planning and evaluation process for entering into other, future partnerships.

The MetaArchive Partnership: Post Digital Library Challenges

The MetaArchive digital preservation network and partnership represents one of the most technically sophisticated solutions developed for digital preservation, yet is characterized by its emphasis on easy replication. This section will explain how the partners achieved an equitable balance between the technical complexity required for reliable digital archiving, and the competing needs of each institution for ease of control and support for its part of the distributed storage network.

The MetaArchive Partnership formed out of work done at Emory University for its MetaScholar Initiative

<<http://www.metascholar.org/>>. Through MetaScholar, research libraries in the Southeastern United States began sharing metadata via the OAI protocol for metadata harvesting within the MetaCombine Project <<http://www.metacombine.org/>>. This led to an interest among partners in establishing a partnership for collaborative digital preservation.

At the same time, many of these potential partners had heard of the emerging National Digital Information Infrastructure for Preservation Program (NDIIPP) that the Library of Congress was allocated funding to develop. Using partnership development funding from the Library of Congress, in 2004 the MetaArchive Digital Preservation Partnership was born from the MetaArchive of Southern Digital Culture.

NDIIPP and MetaArchive

In October 2004, the MetaArchive of Southern Digital Culture was funded by the Library of Congress's NDIIPP initiative to create a collaborative digital preservation network over a three year period. The network is comprised of six nodes, one hosted at each partner institution to automatically and redundantly ingest and preserve content and metadata from each of the partner institutions.

The overall goals of the NDIIPP MetaArchive Partnership are to create the following four deliverables:

- A. Content Conspectus – A comprehensive survey of at-risk digital content relating to Southern Cultural Heritage and housed in the collections of each partner library.
- B. Content Harvest – Each of the partners will contribute close to three terabytes of digital objects and metadata designated as a local priority for preservation.
- C. Cooperative Agreement – Another important deliverable will be the framework for the cooperative agreement among partners. This agreement will cover a wide variety of partnership areas including a charter and a document describing the character and context of the partnership arrangement.
- D. Preservation Network - This will be the functioning network of modified LOCKSS server nodes that collectively will act to preserve digital content over time.

Post Digital Library Challenges

The challenges posed by the MetaArchive Partnership are varied and yet echo many of the same issues that came out of the NSF-sponsored report by Larsen and Wachtlar (2004) on post digital library futures. Of these the MetaArchive Partnership has made headway in the following areas:

- Redundant Digital Archiving Options
- Persistent Identifiers
- Shared Metadata Standards
- Inter-institutional Cooperative Agreements
- Copyright

Rights expression and enforcement are areas of particular interest to the MetaArchive Partnership. Having participated in

recent round tables on orphaned works, the partnership has an approach to rights issues that is defined within the following parameters:

All digital materials that will be ingested within the MetaArchive Network will generally:

- fit within an exception to the exclusive rights of owners, such as the “fair-use” doctrine or other provisions relating specifically to library copying and other activities
- undergo an investigation to determine whether the work still enjoys protection or has lapsed into the public domain due to notice or renewal defects
- occur as a result of valid permission from the copyright owner(s)
- constitute an acceptable risk for the institution in the potential absence of “clear” resolution

MetaArchive Technical Infrastructure

The chief preservation strategy that is being implemented in the MetaArchive Project is that of redundant data storage. The redundancy is spread out over six different institutions utilizing the backbone of the Internet2 Abilene network and the local connections of the Southern Crossroads (SoX) network consortium and the Mid-Atlantic Crossroads (MAX) network consortium. The geographic area extends between Florida, Alabama, Georgia, Kentucky and Virginia. The process for ingesting the digital material and storing it across the different server nodes is automated and managed via the LOCKSS (Lots of Copies Keep Stuff Safe) software architecture. The MetaArchive Network is intended to be a private dark archive of digital objects and is only open to the servers that are part of the MetaArchive Network. Thus if any one server node fails, it can be restored either from the primary source (access copy) of the digital object or from any of the other five server nodes. The restoration process would be automatic once the server node was restored to an online status and could reconnect with the network.

It is impossible to speak of the Metaarchive technical infrastructure without a cursory glance at the overall hierarchy of the preservation collaborative. The MetaArchive framework is comprised of four layers: Consortial Administration, Archival Storage, Content Ingestion and Replication (LOCKSS Software and Hardware and Network Connectivity), and Shared Collection Description. These layers are mapped whenever possible to the Open Archival Information System (OAIS) Framework for digital preservation, which is defined as consisting of Ingest, Archival Storage, Data Management, Administration, Preservation Planning, and Access components.

Consortial Administration

Consortial Administration is at this point completely driven by human administration and oversight and is analogous with the Administrative layer of the OAIS model. This layer provides the common meeting point for the decisions made about content ingestion, archival planning, and inter-institutional long-term storage and access agreements.

Archival Storage

The Archival Storage Layer is the OAIS layer that will be investigated least by the current MetaArchive Partnership, as the partners are relying upon the robust capabilities already available through the LOCKSS system. As shown in Figure 2 below, the MetaArchive Network would like to add a bridge between the content ingestion and replication system with a modular component that will enforce archival storage.

Content Ingestion and Replication

Content Ingestion and Replication is the layer that will be described in detail further in this section and consists of the LOCKSS content ingestion and replication software component as well as a hardware component that makes use of Linux systems administration tools for allocating disc space among institutional nodes.

Shared Collection Description

Shared Collection Description is the infrastructure that was developed to enable and manage collection description for the purposes of identifying *at-risk* digital content among institutional partners. Further description of the collection description metadata schema is found online at:
<http://www.metaarchive.org/pdfs/conspectus_md_2005.html>.

LOCKSS: The Core of MetaArchive

Core to the MetaArchive Technical Architecture is the use of the LOCKSS software platform. The installation that is currently running in the test network is completely closed except to the nodes housed at the member institutions. This is different from the standard LOCKSS installation and offers more control over where digital content is stored and accessed. This enables the local MetaArchive group to set policies on access and control via institutional hierarchy rather than being dependent on all LOCKSS caches. This also limits the number of copies of a work that are available.

The IP control mechanisms being used are currently built into the LOCKSS software. However, the group is also using standard systems administration tools in Linux to create disc space allocation based on an institutional identity in order to govern the amount of disc space used by any one member institution's collections. Currently the amount of disc space per institution is expected to be about 400 GB now that the production system has been put into place (fall 2005).

Optimizing a Local Digital Library Program to Fit Most Needs

Like many other digital library programs, accommodating demands of external partnerships did not occur seamlessly for the FSU Digital Library. As the above description of the MetaArchive partnership makes clear, a great deal of negotiation and fine-tuning was required even after the six partner institutions had agreed to work together. The challenges of social, legal, technical and semantic interoperability are extensive in such ambitious collaborations, even with the existence of protocols and standards like OAI-PMH and SOAP. At the present time, the FSU Digital Library is considering steps that could be taken to make accommodating new partnerships easier. Two possible solutions

that would help are 1) to create an automated mediation and transformation service that could move local digital content and metadata into a variety of manifestations suited to different partners' needs, and 2) to develop an audit checklist of considerations that should be discussed regarding every possible partnership.

Mediation and Transformation Services

A need that currently remains unfilled is an application for moving content and metadata from one environment to another. One promising model for this is the EAI (Enterprise Application Integration) "hub and spoke" model defined by Joshua Fox in his whitepaper entitled *Central Information Models for Data Transformation*. A diagram of a digital library hub and spoke model appears in Figure 3 below.

This type of central information model creates a hub and spoke system similar to modes used in the commercial air transport industry. The hub in this case provides a central location for communication between and among schemas. While this is similar in many aspects to what Tim Berners-Lee and the World Wide Web Consortium (W3C) describe as the semantic web, it allows for transformation of disparate schemas which will be necessary for some time to come at least in the realm of distributed digital libraries.

Elaborating upon the hub and spoke model, the FSU Digital Library has identified the following areas that should be accommodated in a much more consistent and predictable manner when moving and sharing content and metadata for any purpose, including partnerships such as the MetaArchive: Multiple Archiving Options, Persistent Identifiers, Mappings, Protocols, and Discovery Services.

Multiple Archiving Options

Different local and collaborative platforms for archival storage present a variety of issues for digital libraries. At FSU we have internal fibre storage solutions as well as statewide digital repository solution options (DAITSS) and collaborative storage solutions with the MetaArchive Partnership (LOCKSS). Having a median between what is stored at the local level (FSUMD) and at the collaborative level (DAITSS and LOCKSS) is quite substantive and depends on a variety of attributes including format, technical metadata expression, preservation option and local or collaborative policy agreement. A next-generation mediating hub architecture should be able to understand what preservation processes and standards are employed by the "home" system of content, by the destination system, and supplement or modify formats and metadata as necessary.

Persistent Identifiers

Persistent identifiers, such as CNRI's Handle System are necessary for collaboration across institutional digital libraries. Having unique identifiers which can be used both internally in digital library workflow and externally for web navigation is a critical function of any digital library. Thinking institutionally about identifiers is at this point something that all libraries should be thinking about as a service which offers much more granularity than just DNS entries. This area is often overlooked within institutions as is shown in the article by Shreeves, Kaczmarek and

Cole, particularly when content and metadata is moved outside a local environment to a shared environment.

Mappings

Ideally, a mediation and transformation hub could maintain awareness of a variety of schema, namespaces, and encoding standards for metadata, and automatically map and transform metadata as it moved from one application to another.

Protocols

In addition to mediating formats and encodings of data, a robust hub service would be able to express and control the transmission via multiple protocols the same information in different “languages” such as Z39.50, OAI-PMH, etc.

Discovery Services

The area of discovery services is quite new but is one which will remain relevant to digital libraries for some time to come as this type of aggregated search and or personalized information discovery service will be dependent on good documentation and easy transformation of digital library metadata for use with such products. In Figure 3 these services are shown as part of the spokes of the collaborative data views of the digital library content. Another aspect of these services will be the availability of metadata for creating combined or collaborative search indices as well as indices for recommender systems and other enriched discover offerings.

Audit Checklist RLG

Though the mediation and transformation hub described above would require substantial resources, a much easier and immediate step is also available to help the FSU Digital Library accommodate potential partnerships. This solution is a checklist for evaluating both our own program and the program of potential partners.

Choices For Collaboration

In the following chart the authors have taken the pertinent larger evaluative categories from the RLG/NARA *Draft Checklist for the Certification of Trusted Digital Repositories* and have adapted it for use in evaluating digital libraries for collaborative partnerships. The scale at this stage is shown by a checkmark which indicates current awareness of issue and or plan. Lack of a checkmark indicates that this is still an urgent need for the DL. In this case the table shows the collaborative capability of the FSU Digital Library and in comparison the overall capabilities for collaboration of the MetaArchive Partnership.

Checklist for Digital Library Collaborative Readiness	FSU DL	Meta Archive DL
Organization	X	X
Digital Library Functions, Processes, and Procedures		X
Designated Community and Usability of Information		X
Technologies and Technical Infrastructure	X	X

Future Directions

While the future of libraries seem at best a torrent of change and new directions, it is clear from works like DeRosa that users are expecting new services, features, and a level of transparency and access that is unprecedented in the history of cultural heritage organizations. Digital library programs should be aware of this shift and seek to enable interoperability and compatibility among their various systems. While this challenge is somewhat driven by technology and its uses by commercial firms such as Google or Amazon, this is not all to blame on the 90s dot-com explosion. It is a generational statement that comes from a new generation of library and digital library users known as *millennials*. The *millennials* are already having a tremendous effect on all areas of society and as Thomas and McDonald have made clear, these expectations create disconnect between traditional and emerging practices of the entire cultural heritage industry. Thus libraries and in particular digital libraries take on a corporate effect that can not be mitigated by a *wait and see* attitude but must be embraced with new ideas for both market share and iconic placement.

Another clear example of this new direction for cultural heritage institutions and in particular libraries is the study by DeRosa which challenges the traditional roles of libraries and suggests new opportunities. From this study it is easy to see trends in user-centricity – self-service – personal knowledge – personal involvement – online resources – better fit the lifestyle of the majority of users but in particular those who fit a *millennial* demographic. This personal involvement is a new phenomenon in commercial advertising as well. Companies like Toyota through its Scion brand and licensed brands such as Pabst Blue Ribbon Beer have found an anti-advertising niche yet is this really anti-advertising or user-centered advertising that overlaps with what we know about *millennial* user needs and wants. Can libraries learn from this technique and enhance our relationships with this demographic. Can libraries learn from their users about their users? If so then as we optimize our digital libraries for collaboration how can we optimize our digital libraries for user manipulation (mashing) and the creation of new uses for digital libraries that lies in relationships among information across the bulk of all open access digital library content.

Graphics and Equations

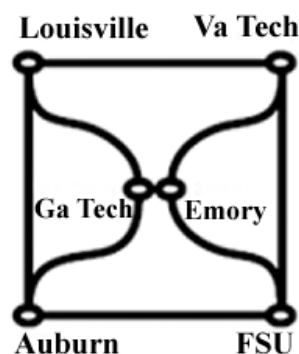


Figure 1. MetaArchive Logo

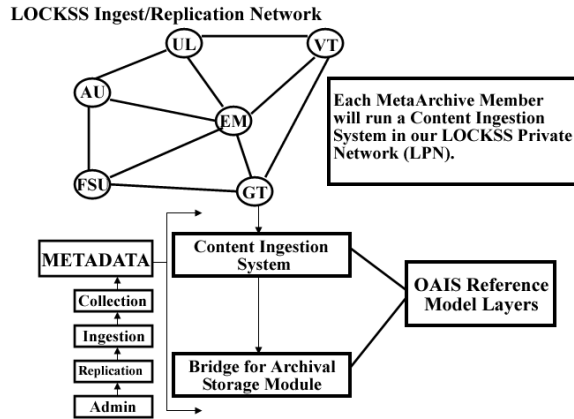


Figure 2. Diagram of Archival Storage Layer

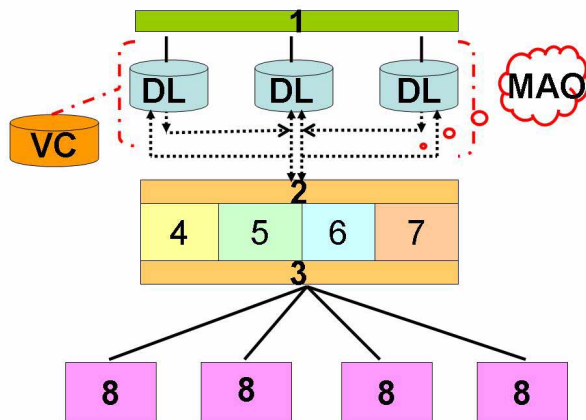


Figure 3. Diagram of Digital Library Hub and Spoke Model

1. Individual Digital Library User Interface
 2. Digital Library Collaboration Hub
 3. Dissemination Layer
 4. Protocols
 5. Identifiers
 6. Mappings
 7. Discovery
 8. Collaborative Data Views – ex. Course Management Systems, Collaborative Interfaces, OAI Broker, Enriched Discovery Tools (Metasearch)
- VC = Virtual Collections derived from discrete Digital Library Collections
 MAO = Multiple Archiving Options
 DL=Discrete Digital Library Collection

References

- [1] Brian Lavoie and Lorcan Dempsey, "Thirteen Ways of Looking at ... Digital Preservation." *D-Lib Magazine*. 10:7/8, (2004). Online at <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>.
- [2] Paul Miller, "Interoperability: What It Is and Why Should I Want It." *Ariadne* 24, (2000). Online at <http://www.ariadne.ac.uk/issue24/interoperability/>
- [3] Vicky Reich and David S.H. Rosenthal, "LOCKSS: A Permanent Web Publishing and Access System." *D-Lib Magazine* 7(6):. Online at <http://www.dlib.org/dlib/june01/reich/06reich.html>
- [4] David S.H. Rosenthal, Thomas Lipkis, Thomas S. Robertson, and Seth Morabito, "Transparent Format Migration of Preserved Web Content." *D-Lib Magazine* 11(1). Online at <http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>.
- [5] Joshua Fox, "Central Information Models for Data Transformation." *EAI Journal* (Fall 2002). Online at <http://www.wilshireconferences.com/EDF2002/Fox-WhitePaper.pdf>.
- [6] Cathy DeRosa et. al., *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*, OCLC, (2005). Online at <http://www.oclc.org/reports/2005perceptions.htm>.
- [7] Ronald L. Larsen and Harold D. Wachtlar, "Knowledge Lost in Information: Report of the NSF Workshop on Research Directions for Digital Libraries." National Science Foundation, (2004). Online at <http://www.sis.pitt.edu/%7Edlwkshop/report.pdf>.
- [8] Charles F. Thomas and Robert H. McDonald, "Net Value(s): Disconnects Between Libraries and the Information Age Mindset." Proceedings of the Free Culture and Digital Library Symposium, Atlanta, GA, (2005). Online at <http://dscholarship.lib.fsu.edu/general/4>.
- [9] Harvard University Libraries, *Global Digital Format Registry Project*. Online at <http://hul.harvard.edu/gdfr>.
- [10] Office of the Register of Copyright, *Report on Orphan Works*, Library of Congress, (2006). Online at <http://www.copyright.gov/orphan>.
- [11] Open Content Alliance, *Open Content Alliance FAQ*, Online at <http://www.opencontentalliance.org/faq.html>.
- [12] Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS)*, (2002). Online at <http://nssdc.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>
- [13] RLG and NARA, *Draft for Comment An Audit Checklist for the Certification of Trusted Digital Repositories*, (2005). Online at http://www.rlg.org/en/page.php?Page_ID=20769.
- [14] Sarah Shreeves, Joanne S. Kaczmarek, and Timothy W. Cole, "Harvesting Cultural Metadata Using the OAI Protocol." *Library Hi Tech* 21 (2003): 159-169.

Author Biography

Robert H. McDonald is the Associate Director for Technology & Research at the Florida State University Libraries. He is responsible for the Division of Library Technology & Research and has previously worked on digital library projects and network infrastructure issues. He holds an MLIS degree from the University of South Carolina and a MMus degree from the University of Georgia and is active in Educause, CNI, and ALA.

Charles F. Thomas is currently a member of the Digital Library Services Team at the Florida Center for Library Automation and was previously the Assistant Director for Scholarly Communication and Integrated Digital Services at the Florida State University Libraries. With over a decade of digital library program management experience, he is heavily involved in digital collections building and has worked previously on large-scale metadata solutions for diffuse digital libraries.