METAARCHIVE
NDIIPP Partnership

Robert H. McDonald
Associate Director of Libraries for Technology & Research
Florida State University
rmcdonal@mailer.fsu.edu

Tyler Walters
Associate Director for Technology & Resource Services
Georgia Institute of Technology Library & Information Center
Tyler@gatech.edu
MetaArchive Project Summary

• Six partner institutions:
  – Emory - Georgia Tech - Florida State
  – Virginia Tech – Auburn - Louisville

• Collaborate w/ LC/NDIIPP – $1.4M initial effort to develop cooperative for preservation of digital content, 2004-2010

• Origins/Precursors - MetaScholar.org:
  AmericanSouth.org
  • Harvest metadata from cultural memory insts. pulling into central location to aggregate, index, search, and discover. 50,000 records, +30 archives
University of Louisville

Emory University

Ga Tech

Florida State University

Auburn University

Va Tech

MAX Connection to Va Tech

FL Lambda Rail

Abilene Network

MAX Network

SOX Network
MetaArchive Project Goals

1) Create a conspectus of digital content held by the partner sites

2) Harvest a body of most critical content to be preserved (3 terabytes, w/ capability to expand)

3) Develop model cooperative charter for collaboration and sustainability

4) LOCKSS-based distributed preservation network infrastructure
   Address geographic and institutional heterogeneity needed to safeguard each institution’s digital collections

- Distribution/Communication: Internet2 Abilene network and local connections of the Southern Crossroads (SoX) network consortium and the Mid-Atlantic Crossroads (MAX) network consortium
MetaArchive Deliverables to Date

• **2004-2005**
  – Create Collection Conspectus
  – LOCKSS Customization
  – Network Setup

• **2005-2006**
  – Network Launch/Initial Harvest
  – Charter Cooperative Agreement

• **2006-2007 (projected)**
  – Non-Profit Setup (Educopia Institute)
  – Extend partnership beyond original 7 partner institutions
  – Finish Current Collection Ingest
MetaArchive Technology

- LOCKSS Scalability
  - 2.2 Terabytes Storage Space Available
  - Proven Security (SE-Linux) Across Multiple Networks
  - Enterprise Installation
    - RedHat
    - CentOS

- Database Driven Conspectus
  - Manages Collections within LOCKSS
  - Network Administrative Management Tool

- Collaboration Tools
  - MediaWiki, VideoConference, Teleconference
MetaArchive Preservation Network

Internet2 Abilene Network

RedHat OS LOCKSS

Switch to FSU Network

Dell Poweredge 1850 - Gateway

Brocade Fibre Channel Switch

Dell/EMC AX 100 - SAN

APC UPS - Power Supply

FSU

Emory

GaTech

Auburn

VaTech

Louisville
MetaArchive Conspectus Database

Conspectus Database – Ties in with LOCKSS Central Plugin Repository

Log In Here

Login: 
Password: 
Log In

If you've lost or forgotten your password, click Here.

Login Attempt Failed

Actions

View All RDF Data
View By Institution
View All Valid RDF Data
View Selected Collections
View Conspectus Schema

View Existing Descriptions
**Example Conspectus Entry – GaTech ETDs**

- **Collection ID:** 32
- **Alternate Title:** Georgia Tech Electronic Thesis and Dissertation Collection
- **Description:** The Electronic Thesis and Dissertation Collection at Georgia Tech covers Fall 2003 through the present. This program is jointly sponsored by the Graduate Studies and Research Office and the Library and Information Center. Georgia Tech is a member of the Networked Digital Library of Theses and Dissertations and participates in the Open Archives Initiative.

<table>
<thead>
<tr>
<th>ESC Heading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Transportation</td>
</tr>
<tr>
<td>Art and Architecture</td>
</tr>
<tr>
<td>Law and Politics</td>
</tr>
<tr>
<td>Geography</td>
</tr>
<tr>
<td>Business, Agriculture, and Industry</td>
</tr>
<tr>
<td>Social Class</td>
</tr>
<tr>
<td>Environment</td>
</tr>
<tr>
<td>Recreation</td>
</tr>
<tr>
<td>History, Manners, and Myth</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LCSH Heading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissertations, Academic -- Georgia -- Atlanta</td>
</tr>
<tr>
<td>Electronic dissertations -- Georgia -- Atlanta</td>
</tr>
</tbody>
</table>
## LOCKSS ADMIN INTERFACE for METAARCHIVE NETWORK

<table>
<thead>
<tr>
<th>Status</th>
<th>Status Last Updated</th>
<th>Institution Name</th>
<th>IP Address</th>
<th>Reverse DNS Entry</th>
<th>Log Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>![ ]</td>
<td>September 20, 2005, 12:02 pm</td>
<td>Louisville University</td>
<td>meta-vault.library.louisville.edu</td>
<td><a href="http://meta-vault.library.louisville.edu">http://meta-vault.library.louisville.edu</a></td>
<td></td>
</tr>
<tr>
<td>![ ]</td>
<td>September 20, 2005, 12:02 pm</td>
<td>Auburn University</td>
<td>meg.lib.auburn.edu</td>
<td><a href="http://meg.lib.auburn.edu">http://meg.lib.auburn.edu</a></td>
<td></td>
</tr>
<tr>
<td>![ ]</td>
<td>September 20, 2005, 12:02 pm</td>
<td>Emory University</td>
<td>ndiap.library.emory.edu</td>
<td><a href="http://ndiap.library.emory.edu">http://ndiap.library.emory.edu</a></td>
<td></td>
</tr>
<tr>
<td>![ ]</td>
<td>September 20, 2005, 12:02 pm</td>
<td>Florida State University</td>
<td>clockss.lib.fsu.edu</td>
<td><a href="http://clockss.lib.fsu.edu">http://clockss.lib.fsu.edu</a></td>
<td></td>
</tr>
<tr>
<td>![ ]</td>
<td>September 20, 2005, 12:02 pm</td>
<td>Florida State University</td>
<td>clockss2.lib.fsu.edu</td>
<td><a href="http://clockss2.lib.fsu.edu">http://clockss2.lib.fsu.edu</a></td>
<td></td>
</tr>
<tr>
<td>![ ]</td>
<td>September 20, 2005, 12:02 pm</td>
<td>Georgia Institute of Technology</td>
<td>ndiiplockss.library.gatech.edu</td>
<td><a href="http://ndiiplockss.library.gatech.edu">http://ndiiplockss.library.gatech.edu</a></td>
<td></td>
</tr>
</tbody>
</table>
EVO Web Collaboration Tools
Main Page

Contents [hide]

1 MetaArchive of Southern Digital Culture Working Documents
   1.1 Conspectus
   1.2 Cooperative Agreement Analysis Plan (CAAP)
   1.3 Technology Documents
   1.4 NEH Grant Rough Draft
   1.5 IMLS Technology Training Grant Rough Draft
   1.6 Security and Continuity Plan Rough Draft

2 Wiki Editing Help

MetaArchive of Southern Digital Culture Working Documents

Contents of the old wiki can be found here.

Conspectus

- ScopeDocument
- DescriptiveSchema
- DescriptiveSchemaElements
- MaterialsDocument
- SubjectsDocument

Cooperative Agreement Analysis Plan (CAAP)

MediaWiki Portal – Document Collaboration
Organizational Issues

“Sustaining the Cooperative”

Advancing our models:
• Organizational
• Technology
• Service

Organizational Model of Sustainability:
• Individual institutions: Emory, etc.
• Consortia-building: Cooperative Charter
• Non-profit mgmt. entity: Educopia Institute
• National strategic partnerships: LC/NDIIPP
MetaArchive Digital Sustainability Model

**Economic**
- General Resource Inputs:
  1) Financial
  2) Infrastructure
  3) Expertise

- #4 Inputs: National/International Government Agencies & Organizations
  - More Resources, Strategic Alignment
  - High-Level Communication/Coordination
  - New Economic Models
  - International Collaboration
  - NDIIPP Funding
  - DPC Projects

- #3 Inputs: Organizations, Projects, Technologies
  - Oversee, Guide, Coordinate
  - More Strategy / More Resources
  - Sustain Innovation (Ideas/Technologies)
  - Mellon Funding

- #2 Inputs: Add & Combine
  - Resource inputs from six universities
  - New Infrastructure
  - New Ideas / Technologies begin and broaden NDIIPP Funding

- #1 Base of Inputs:
  - Universities' Infrastructure, People & Expertise,
  - Initial Technology

**Organization**
- LC / NDIIPP DPC
  - Educorpie Institute
  - MetaArchive Cooperative (GT, FSU, Auburn, VT, Louisville, with Emory)
  - Emory University

**Technology**
- LC / NDIIPP DPC
  - Educorpie Institute
  - MetaArchive Cooperative, i.e. Georgia Tech, other developers
  - Stanford / LOCKSS Development Group

**Collections**
- Sustain:
  - Content (Reliability)
  - Context (Provenance)
  - Structure
  - Authenticity

- Services:
  - Preservation (Content ingest / retrieval)
  - Format migration
  - Digital Collection Disaster Recovery
  - Digital Preservation Network Consulting
  - LOCKSS services
MetaArchive as Consortia: Cooperative Charter

- Cooperative Charter addresses four areas:

- A) Types of partnerships in the MetaArchive Cooperative:

  - Development Partners – responsible for developing and testing the preservation network technology and operating a preservation node

  - Preservation Partners – responsible for operating a preservation node, ingesting collections from partner institutions, and making the node available for testing

  - Contributing Partners – cultural memory institutions that possess digital content to preserve via the MetaArchive Preservation Network. They contribute fees for this service, do not operate a node

  - Sponsored Partners – cultural memory institutions that own or possess digital content they wish to preserve, but cannot contribute service fees. A Development Partner sponsors the preservation of their digital content
Cooperative Charter

B). How MetaArchive is organized, governed, and communications
   – Steering Cmte.
   – Content Cmte.
   – Preservation Cmte.
   – Technical Cmte.

• Partner institution reps can serve terms on the committees, ensuring a broad participation in governance and operations

C). Cooperative services to members
   – Digital preservation (network dev./maintenance, content ingest/retrieval)
   – Format migration
   – Digital collection disaster recovery
   – Digital preservation network consulting
   – LOCKSS services

D). 1. Technical specs for preservation network
    2. Copy of memo of understanding between the six universities
MetaArchive and Educopia Institute

- Non-profit management entity – Three issues:
  1) Continuing need for financial resources
  2) Expose MetaArchive to new digital projects to inform development
  3) Economically efficient, catalytic structure to bring these about

- Educopia Institute:
  - Provide oversight to MA Cooperative and other digital projects
  - Low-cost, low overhead conduit for digital library, scholarly communications technology projects
  - Advance cyberinfrastructure needed to drive research, teaching and learning in contemporary digital era

  NSF (2003) and ACLS (2006) Cyberinfrastructure reports:
  - Scholarly activity – teaching, research, learning, knowledge transfer via scholarly communications – need rational, strategic cyberinfrastructure

**EI: Generate DL technology projects to support mission and goals**
Nat’l Strategic Partnership: LC/NDIIPP

- **Access to resources, context of a nat’l digital preservation agenda**
  - LC’s digital preservation partners and their approaches
  - Expertise within LC itself
  - Funding - $650,000, 2004-2007 - servers, project staff
  - Driver to develop MA models: Organizational, Technology, Services

- **Facilitate communication on dig. sustainability, gather know-how**

- **Economic models for digital library activity**
  - Consultant, Paul Courant – Univ. of Michigan, participates in meetings and projects regarding economic sustainability issues for digital pres.

- **Align programs and projects with Nat’l/Int’l initiatives**
  - Increasingly important to generating revenues

- **Emerging Nat'l/Int’l approaches to guide funding**
  - UK: JISC and DCC
  - Australian Partnership for Sustainable Repositories [APSR]
  - Netherlands KB E-Depot
  - U.S.: NDIIPP
MetaArchive: Organizational Issues

International Strategy?

- Membership in Digital Preservation Coalition
  - Further embed MA’s activities in Nat'l/Int’l strategic initiatives

Summary: In two years, much work has taken place to grow the MetaArchive Cooperative into a sustainable organization:

- Four levels of organization building:
  - Institution
  - Consortium
  - Non-profit Management Entity
  - National/International Strategic Strategic Partner

...should result in a dynamic and developing organization that productively addresses distributed digital preservation issues
Digital Preservation of Interest

- **LC NDIIPP**
- **MetaArchive**
  [http://www.metaarchive.org](http://www.metaarchive.org)
- **KB e-Depot**
  [http://eprints.rclis.org/archive/00005795](http://eprints.rclis.org/archive/00005795)
- **JISC UK LOCKSS Pilot**
  [http://www.jisc.ac.uk/index.cfm?name=programme_lockss](http://www.jisc.ac.uk/index.cfm?name=programme_lockss)
- **Australian Partnership for Sustainable Repositories**
- **SDSC/NARA Chronopolis Project**
Collaboration Tools of Interest

- **EVO**
  [http://evo.caltech.edu/](http://evo.caltech.edu/)
- **VRVS**
- **ConferenceXP**
  [http://research.microsoft.com/conferencexp/](http://research.microsoft.com/conferencexp/)
- **Jabber**
- **PHPCollab**
- **MediaWiki**
  [http://www.mediawiki.org](http://www.mediawiki.org)