New Cooperative Strategies for Distributed Digital Preservation

MetaArchive
COOPERATIVE

TYLER WALTERS
UBC Vancouver, March 12, 2010

Presentation Overview

- **The MetaArchive Cooperative**
  - MetaArchive Components and Lessons Learned
    - Technical Infrastructure
    - Curating Collections
    - Organizational Infrastructure
- Future Directions and Initiatives…

So, what is a “Cooperative” anyway?

- From *Dictionary.com*:
- *“a jointly owned enterprise engaging in the production or distribution of goods or the supplying of services, operated by its members for their mutual benefit…”*

MetaArchive: Basic Facts

- Established in 2004, began preserving content for 6 member institutions
- Uses LOCKSS software to provide long-term management for materials in a distributed digital preservation network
- Sustainable organizational framework: Membership organization with a 501c3 host (Educopia Institute)
- 254 TB network capacity (20-50 TB per member, adding more as new members join)
- OAIS Compliant, and as a Trustworthy Digital Repository (2009 TRAC audit available on our site next week)
- *Guide to Distributed Digital Preservation* – published!
  - 755 downloads since February 23!

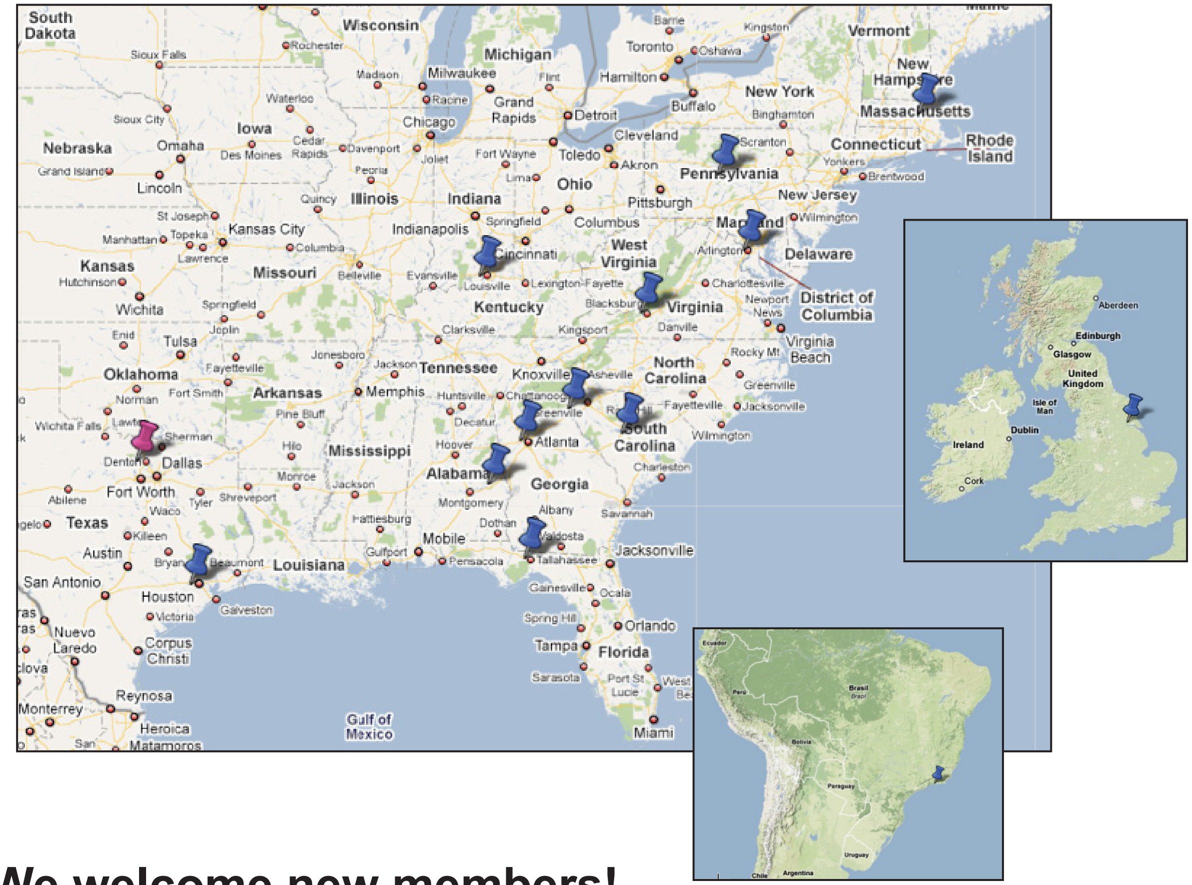
MetaArchive Cooperative www.metaarchive.org

Current Members/Contributors

Auburn University
Boston College
Clemson University
Emory University
Florida State University
Folger Shakespeare Library
Georgia Tech
Indiana State University
Penn State University
PUC Rio de Janeiro
Rice University
University of Hull
University of Louisville
University of North Texas
University of South Carolina
Virginia Tech

Current Affiliates

Library of Congress
NDLTD
SDSC Chronopolis



We welcome new members!

The Distributed Digital Preservation Process

1. Assessing DP needs
2. Developing DP plan and policies
3. Selecting content
4. Preparing contracts / MoAs
5. Implementing preservation technology/service
6. Submitting content
7. Monitoring content

Examples of MetaArchive's materials

- Born digital and digitized collections
- Digital image, sound, and video files
- Datasets and Databases
- GIS Collections
- Websites
- Email correspondence
- E-journals
- Electronic Theses and Dissertations (ETDs)
- Encoded texts

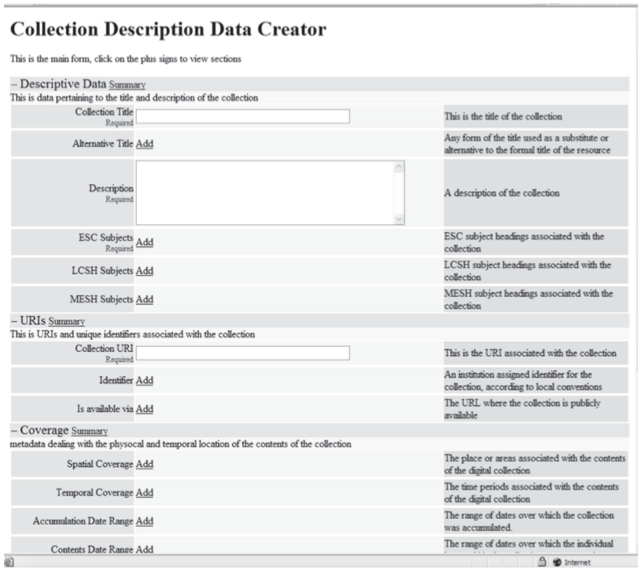
Technical Framework

- LOCKSS-based Distributed Digital Preservation Network
  - Robust, distributed network launched 2004
  - Open Source
    - For digital objects, not just journals
    - Working with larger file sizes
    - Working with many varied collections
  - Fully replicable technical model
  - Others now founding private LOCKSS networks:
    - Alabama/ADPNet, Arizona/PeDALS, ICPSR/DataPASS, western Canada/COPPUL, USGovDocs mirror sites, CLOCKSS, KOPAL (Germany)

Technical Framework

Created software tools to curate its collections:

- Conspectus schema
    - (Fixity, provenance, content, context)
  - Webform
  - Based on DC, MODS, DC/CLD,
    - UKOLN RSLP
  - PREMIS (future)



Collection Description Data Creator

This is the main form, click on the plus signs to view sections

– Descriptive Data Summary

This is data pertaining to the title and description of the collection

Collection Title Required		This is the title of the collection
Alternative Title	Add	Any form of the title used as a substitute or alternative to the formal title of the resource
Description Required		A description of the collection
ESC Subjects Required	Add	ESC subject headings associated with the collection
LCSH Subjects	Add	LCSH subject headings associated with the collection
MESH Subjects	Add	MESH subject headings associated with the collection

– URIs Summary

This is URIs and unique identifiers associated with the collection

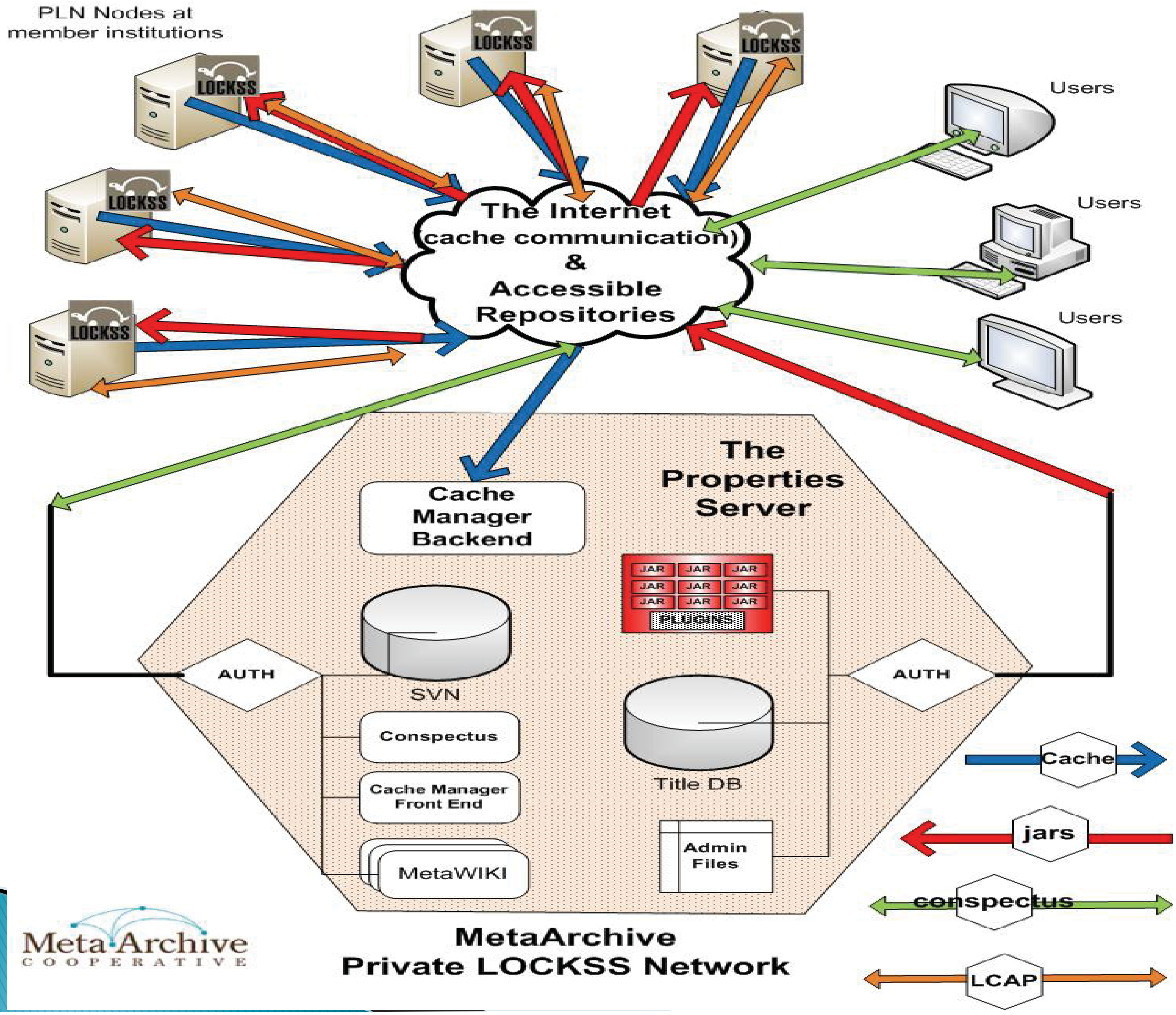
Collection URI Required		This is the URI associated with the collection
Identifier	Add	An institution assigned identifier for the collection, according to local conventions
Is available via	Add	The URL where the collection is publicly available

– Coverage Summary

metadata dealing with the physocal and temporal location of the contents of the collection

Spatial Coverage	Add	The place or areas associated with the contents of the digital collection
Temporal Coverage	Add	The time periods associated with the contents of the digital collection
Accumulation Date Range	Add	The range of dates over which the collection was accumulated.
Contents Date Range	Add	The range of dates over which the individual

- Cache manager
  - Monitors network (AUs, collections, inst.-level)
  - Generates human-readable reports



PLN Nodes at member institutions

LOCKSS

The Internet (cache communication) & Accessible Repositories

Users

The Properties Server

Cache Manager Backend

JAR

PLUGINS

AUTH

SVN

Conspectus

Cache Manager Front End

MetaWIKI

Title DB

Admin Files

Cache

jars

conspectus

LCAP

MetaArchive Cooperative

MetaArchive Private LOCKSS Network

Lessons Learned: Technical Framework

- LOCKSS applies well to non-serialized content
- Guaranteeing authenticity of digital resources may be best accomplished with a distributed digital preservation system approach
- Technical infrastructure requires systems administration attention at each Preservation site
- Sustainability requires having multiple people experienced with the management of the system at all times

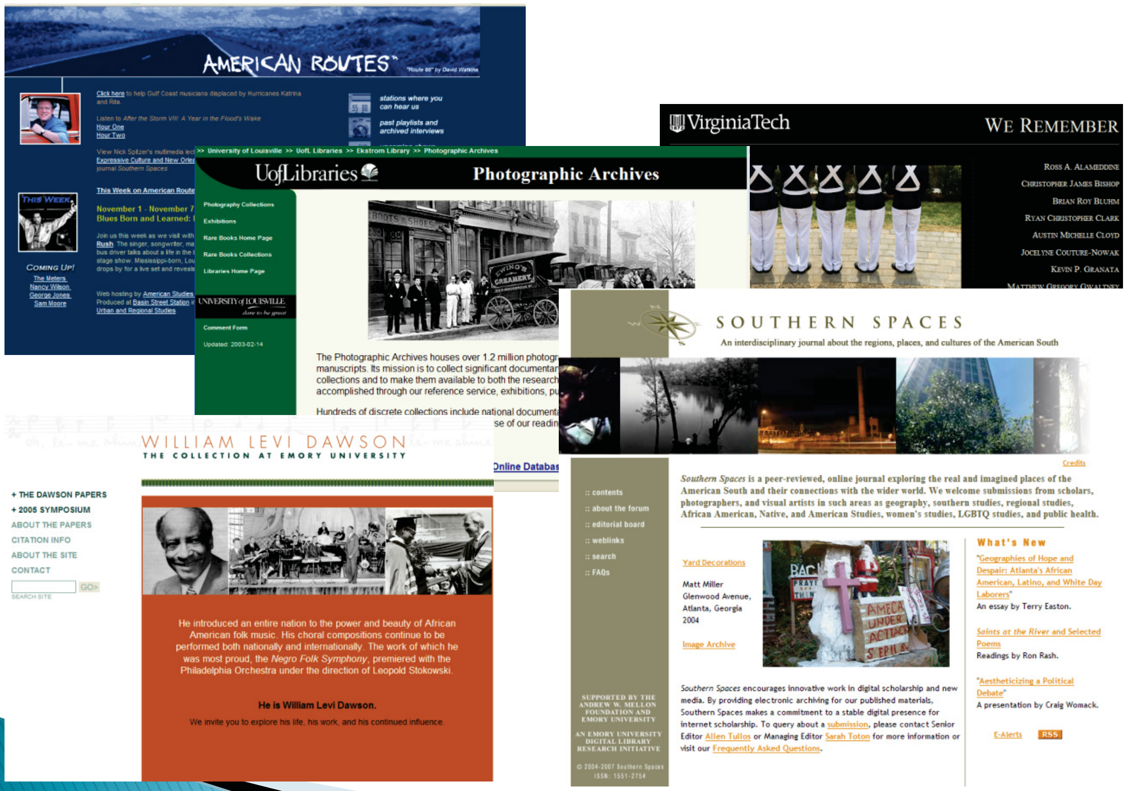
Curating Collections

- Three “archives” to date:
  - Southern Digital Culture,
  - Electronic Theses and Dissertations
  - History of the Slave Trade

- Establishing new archives at member requests
  - e.g. Newspapers, Research Data, Univ. e-Records/Archives
- Curatorial decisions made by the contributing institutions, *not* by MetaArchive
- Can ingest digital objects and their metadata
- Require collection-level metadata for retrieval

Curating Collections

Ingest from web, OAI, CONTENTdm, DSpace, Fedora, and more / Preserving ca. 1,000 Collections to date


AMERICAN ROUTES
"Route 66" by David Watkins
stations where you can hear us
past playlists and archived interviews
This Week
Coming Up!
The Meters
Nancy Wilson
George Jones
Sam Moore

Click here to help Gulf Coast musicians displaced by Hurricanes Katrina and Rita.

Listen to *After the Storm VIII: A Year in the Flood's Wake*
Hour One
Hour Two

View Nick Spitzer's multimedia lect
Expressive Culture and New Orle
journal *Southern Spaces*

This Week on American Route

November 1 - November 7
Blues Born and Learned:

Join us this week as we visit with
Rush. The singer, songwriter, ma
bus driver talks about a life in the
stage show. Mississippi-born, Lou
drops by for a live set and reveals

Web hosting by American Studies
Produced at Basin Street Station i
Urban and Regional Studies

The Photographic Archives houses over 1.2 million photogr
manuscripts. Its mission is to collect significant documentar
collections and to make them available to both the research
accomplished through our reference service, exhibitions, pu

Hundreds of discrete collections include national documenta
se of our readin

Ross A. Alameddine
Christopher James Bishop
Brian Roy Bluhm
Ryan Christopher Clark
Austin Michelle Cloyd
Jocelyne Couture-Nowak
Kevin P. Granata
Matthew Gregory Gwaltney

He introduced an entire nation to the power and beauty of African American folk music. His choral compositions continue to be performed both nationally and internationally. The work of which he was most proud, the *Negro Folk Symphony*, premiered with the Philadelphia Orchestra under the direction of Leopold Stokowski.

He is William Levi Dawson.

We invite you to explore his life, his work, and his continued influence.

Southern Spaces is a peer-reviewed, online journal exploring the real and imagined places of the American South and their connections with the wider world. We welcome submissions from scholars, photographers, and visual artists in such areas as geography, southern studies, regional studies, African American, Native, and American Studies, women's studies, LGBTQ studies, and public health.

Yard Decorations

Matt Miller
Glenwood Avenue,
Atlanta, Georgia
2004

Image Archive

"Geographies of Hope and Despair: Atlanta's African American, Latino, and White Day Laborers"
An essay by Terry Easton.

Saints at the River and Selected Poems
Readings by Ron Rash.

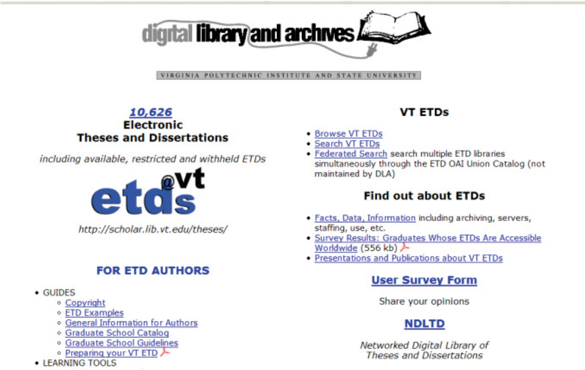
"Aestheticizing a Political Debate"
A presentation by Craig Womack.

Southern Spaces encourages innovative work in digital scholarship and new media. By providing electronic archiving for our published materials, Southern Spaces makes a commitment to a stable digital presence for internet scholarship. To query about a submission, please contact Senior Editor Allen Tullos or Managing Editor Sarah Toton for more information or visit our Frequently Asked Questions.

SUPPORTED BY THE ANDREW W. MELLON FOUNDATION AND EMORY UNIVERSITY

AN EMORY UNIVERSITY DIGITAL LIBRARY RESEARCH INITIATIVE

© 2004-2007 Southern Spaces
ISSN: 1551-2754


digital library and archives
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
etds @vt

10,626
Electronic
Theses and Dissertations

including available, restricted and withheld ETDs

http://scholar.lib.vt.edu/theses/

FOR ETD AUTHORS

- GUIDES
  - Copyright
  - ETD Examples
  - General Information for Authors
  - Graduate School Catalog
  - Graduate School Guidelines
  - Preparing your VT ETD
- LEARNING TOOLS

VT ETDs

- Browse VT ETDs
- Search VT ETDs
- Federated Search search multiple ETD libraries simultaneously through the ETD OAI Union Catalog (not maintained by DLA)

Find out about ETDs

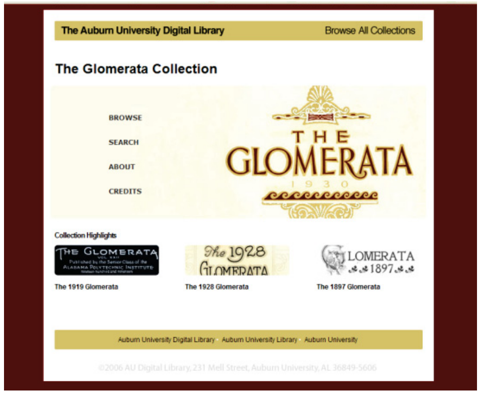
- Facts, Data, Information including archiving, servers, staffing, use, etc.
- Survey Results: Graduates Whose ETDs Are Accessible Worldwide (556 kb)
- Presentations and Publications about VT ETDs

User Survey Form

Share your opinions

NDLTD

Networked Digital Library of Theses and Dissertations


The Auburn University Digital Library
Browse All Collections
BROWSE
SEARCH
ABOUT
CREDITS
THE GLOMERATA 1930
THE GLOMERATA
The 1928 Glomerata
GLOMERATA 1897
Auburn University Digital Library · Auburn University Library · Auburn University

The Glomerata Collection

Collection Highlights

The 1919 Glomerata
The 1928 Glomerata
The 1897 Glomerata

©2006 AU Digital Library, 231 Mell Street, Auburn University, AL 36849-5606

Lessons Learned: Curating Collections

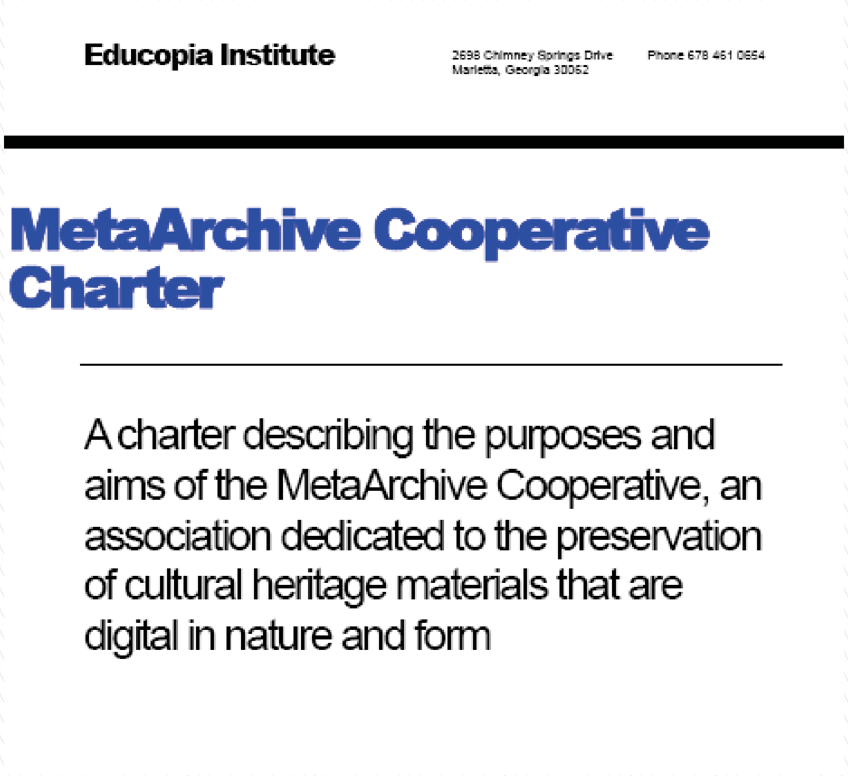
- We needed improved collection-level metadata as a basic tool and system component. (Conspectus DB)
- Just because an institution has content doesn’t mean that content is ready for ingest
- Preservation begins at creation: the organization of an institution’s collections can help or hinder its preservation readiness
- Preservation depends on internal institutional documentation as well

Organizational Infrastructure

- Began as one six-institution network as part of the Library of Congress NDIIPP MetaArchive project
    - Emory University, Georgia Tech, Virginia Tech, Auburn University, University of Louisville, Florida State
- Quickly realized that a preservation solution cannot be dependent on grant funding!
- Sustaining the network demanded longer-term relationship, not dependent on one institution
    - Cooperative Charter and Membership Agreement

Model: Cooperative Association

- Cooperative Charter Goals:
- 1. Define the mission and operating principles, membership responsibilities, governance structure, and services and operations of the Cooperative, and
- 2. Formalize relationships between member institutions as an effective consortium



Educopia Institute

2698 Chimney Springs Drive
Marietta, Georgia 30062

Phone 678 461 0654

MetaArchive Cooperative Charter

A charter describing the purposes and aims of the MetaArchive Cooperative, an association dedicated to the preservation of cultural heritage materials that are digital in nature and form

Organizational Infrastructure

- **New Host Institution: Educopia Institute, Inc.**

- In October 2006, we created the Educopia Institute, a 501(c)3 nonprofit organization to address the needs of cultural memory institutions for shared cyberinfrastructure

  - Work with prospective members;
  - Collect, maintain, and distribute funds;
  - Maintain documentation, website, listservs;
  - Organize and host meetings and workshops;
  - Hold members accountable for completing tasks;
  - Foster relationships with other consortia

Organizational Infrastructure

- Sustaining Members:
  - Pioneers. $5,000/year; 3-year term; host node for research, development, and preservation activities; representation on the Steering Committee; access to 40 GB space*

- Preservation Members:
  - Central preservation partners. $1,000/year, 3-year term, host node for preservation activities, access to 20 GB space*

*more space can be purchased by GB as needed ($2/GB)

Future: Chronopolis – MetaArchive

Improving Inter-Institutional Preservation

- From Silos to Interoperability…
- Two successful early approaches:
  - Integrated Rule-Oriented Data System (iRODS)
  - Lots of Copies Keep Stuff Safe (LOCKSS)
- Powerful technologies, currently isolated
- Seeking to bridge the gap and foster interoperability

Other Future Initiatives…

- PLN Conference in Boston, Oct. 25–26, 2010
  - Creating a community of distributed digital preservation practitioners

- “Super Node” Initiative under discussion

- Current / Future studies:
  - Understanding the Economics of Digital Preservation

Questions and Comments?

Tyler Walters

404 385 4489

Tyler@gatech.edu

TyWalters1 = Skype / ooVoo/ Gmail / Gtalk

NHPRC
DOCUMENTING DEMOCRACY
National Historical Publications and Records Commission

Digital PRESERVATION
National Digital Information Infrastructure & Preservation Program:
A Collaborative Initiative of the Library of Congress