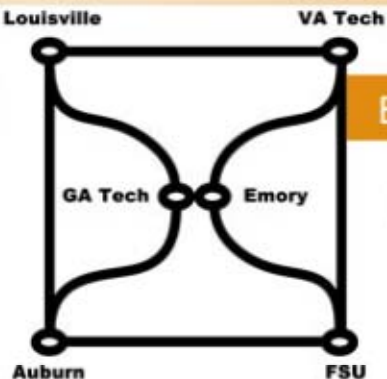


The MetaArchive of Southern Digital Culture: Building a Collaborative Digital Preservation Network

Martin Halbert, Emory University
Robert H. McDonald, Florida State University
Beth Nicol, Auburn University
Vicky Reich, LOCKSS Program
Tyler Walters, Georgia Institute of Technology

CNI Fall 2004 Task Force Meeting
December 7, 2004



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

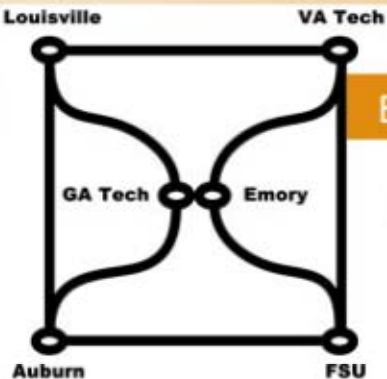
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Project Summary

- Six partner institutions will collaborate with LoC on a three year \$1.3M effort to develop a cooperative for the preservation of digital content with a particular content focus:
Southern culture and history



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

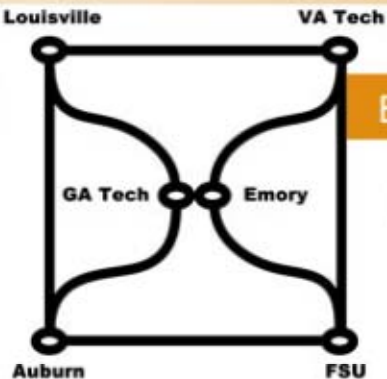
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Project Goals

- Conspectus of digital content held by the partner sites
- Harvested body of the most critical content to be preserved (**3 terabytes**)
- Model cooperative agreement for ongoing collaboration
- Distributed preservation network infrastructure based on the LOCKSS software.



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

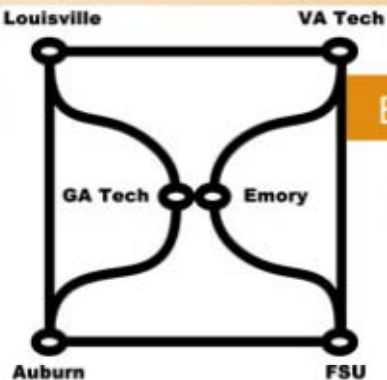
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Project Timeline

- Feb 2005: Conspectus complete
- May 2005: SW/HW tests complete
- Aug 2005: Initial archiving complete
- Feb 2006: Cooperative model analysis complete
- Aug 2006: All subsequent project harvests complete
- Jan 2007: All project goals complete



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

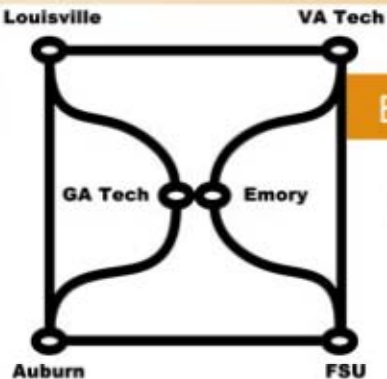
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Preservation Network

- Effective digital preservation succeeds by distributing copies of content in secure, distributed locations over time
- This preservation network is based on a leading preservation software for distributed archiving (LOCKSS), establishing from the beginning a distributed means of replicated archives



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

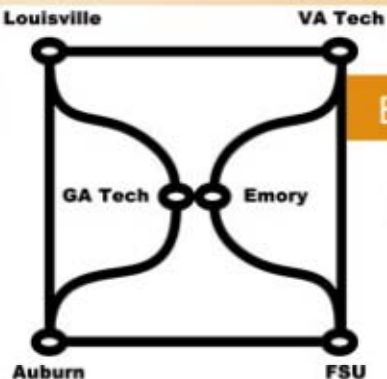
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Adapting LOCKSS Software

- Allows Cooperative to practice its “distributed digital replication” approach. Replicate archival material, not just published material, as originally designed:
 - data integrity checks
 - rigorous security checks
 - focused web crawls to gather/ingest digital content
 - problem of dynamically constructed content (to be studied)



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

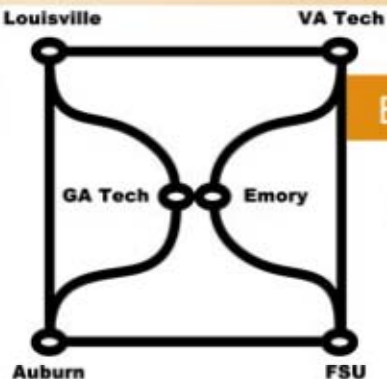
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Dark Archiving

- Advantage: many preservation efforts mix high accessibility online with long-term access (preservation). High accessibility equals high costs
- Content in preservation network exchangeable via OAI-PMH and LOCKSS manifest. Download available for replacement only
- Processes will be developed to retrieve items from the preservation network



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

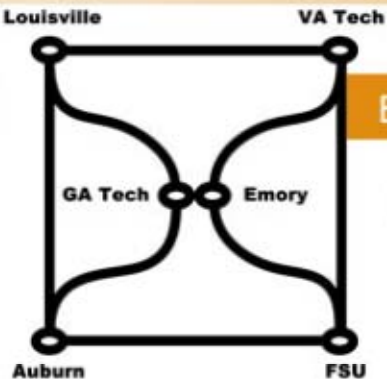
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Low Cost

- Designed for minimal expenditures, low barriers to adoption, for medium-sized institutions. Runs on inexpensive computers, modest degree of systems administration for ongoing maintenance
- The Cooperative will develop a freely available, open source adaptation of the LOCKSS software



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

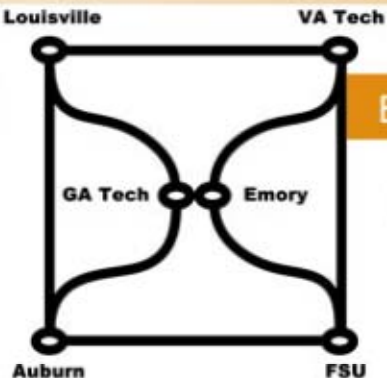
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Network Hardware (all institutions)

- SAN array:
 - Dell/EMC AX100 Array (single processor)
 - 3TB storage space (Four 3X250 GB 7200 rpm serial ATA hard disk drives)
- SAN server:
 - Dell PowerEdge 1850 (2 processors)
 - 3.0GHz/1MB Cache, Xeon 800MHz
- Firewall
 - Dell PowerEdge 1850 (1 processor)
 - 3.0Ghz/1MB Cache, Xeon 800MHz
 - Dell PowerConnect 2616 Unmanaged Switch



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

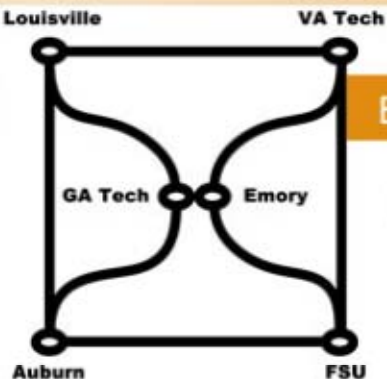
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

LOCKSS Program and MetaArchive

- MetaArchive Work is Important:
 - Extending community: open source software & collection practices
 - More developers welcome and needed:
 - More e-journals (write Plug-Ins)
 - Develop and maintain Java code base
 - New applications: Newspapers, E-thesis, Web sites, Gov Docs



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

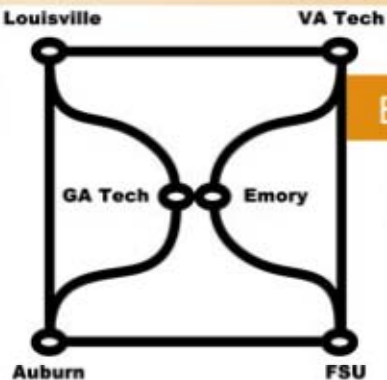
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

LOCKSS Program and MetaArchive

- MetaArchive Work is Important:
 - Building software diversity
 - Building hardware diversity
 - Building collection diversity
- **Preservation system requirement:
no single point of failure**



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

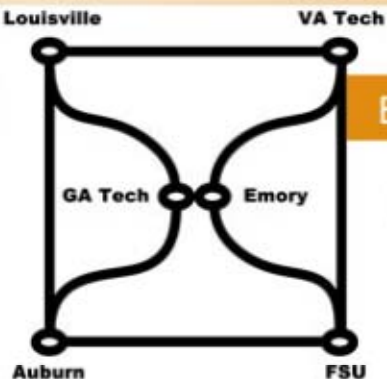
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Content Selection

- Defining the Scope of the Content
 - What is “Southern Culture”?
 - What is meant by “at risk”?
 - Classification and organization of intellectual content based on the Encyclopedia of Southern Culture chapter headings
 - The selection of specific materials is left to the cooperating institutions



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

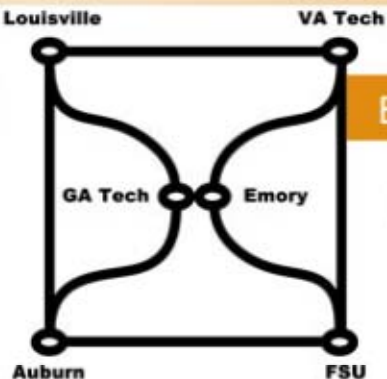
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Content Selection

- Developing a Conspectus
 - Materials to be preserved
 - Format of preserved materials
 - Insuring that we have appropriate rights to harvest and redistribute the content
 - Can the materials be made available for collection?



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

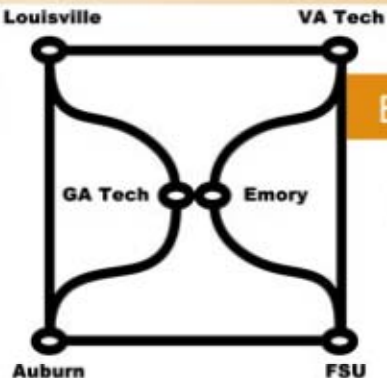
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Metadata

- Gathering Stage (What can We Build On?)
 - Dublin Core Collection Description Application Profile
 - UKOLN/RSLP Collection Description Metadata Schema
 - Western States Dublin Core Metadata Best Practices
 - IMLS/NLG/UIUC Collection Description Metadata Schema
 - LOCKSS Program Proposed Metadata
 - NISO Metasearch Collection Description Group



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

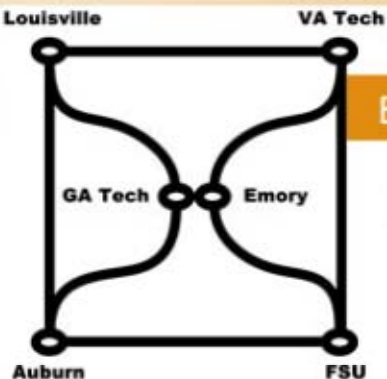
M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Metadata

- Deliverables
 - Conspectus (Database format) will include metadata (only available to network members).
 - Will publish online the adapted way in which we use metadata – showing any unique or qualified tags that are used (Storage & Use MD and metadata that is adapted for LOCKSS of interest).
 - All original tags will be created in subschema.



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

Metadata - Strategy

GROUP METADATA CHOICES

SCHEMA

Emphasis
On
Collective
Description

CONTENT

Normalize
Into One
MetaArchive
Vocabulary

STORAGE & USE

Area of
Project
Exploration

SHARING

Less focus
on sharing
what is in
LOCKSS

LOCAL METADATA CHOICES

SCHEMA

Emphasis
On
Item-
Level

CONTENT

LCSH
AAT
GETTY

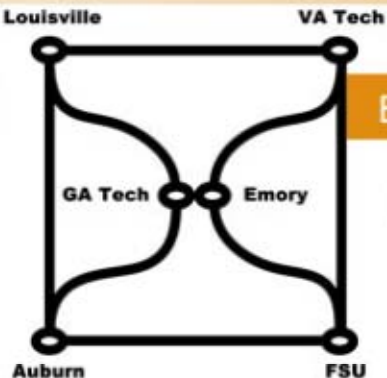
STORAGE & USE

XML
RDBMS
METS

SHARING

OAI-PMH
Access/Use
Licensing
Integration

MA MD
Record



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

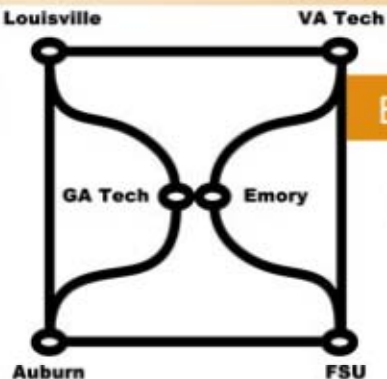
METAARCHIVE

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>

QUESTIONS ?

- Contact Info.
 - Martin Halbert - mhalber@emory.edu
 - Robert McDonald – rmcdonal@mailier.fsu.edu
 - Beth Nicol – nicollb@auburn.edu
 - Vicky Reich – vreich@stanford.edu
 - Tyler Walters – tyler.walters@library.gatech.edu
- WebSite – <http://www.metaarchive.org>



Emory Univ. | Ga. Tech. | Va. Tech | Florida State Univ. | Auburn Univ. | Univ. of Louisville

M E T A A R C H I V E

Supported by the Library of Congress

<http://www.metaarchive.org> ■ <http://www.digitalpreservation.gov>