
NETWORK PLAN AND HARVESTING SCHEDULE

MetaArchive Project, Extension Phase



2008-02-08

Summary

This document describes the plan for 1) conducting extension phase harvesting activities for the MetaArchive of Southern Digital Culture; 2) adding new networks to the MetaArchive Cooperative's network system; and 3) conducting initial harvesting activities for new networks. Harvesting activities include 1) preparatory work on collections (identification, evaluation, AU designation, any necessary "data wrangling" and mounting, and conspectus entry creation); 2) creation of required LOCKSS plugins and manifests and signaling of harvest readiness; and 3) the activation of the harvesting process in every individual site installation. This document references the project extension plan (A.6.) and other project documents which provide more details about other aspects of the project.

What

In order to complete technical activities for the extension phase, we will undertake overlapping phases of work in four main areas: A) harvesting content, B) network expansion, C) network testing, and D) network reporting.

A. Harvesting Content

- 1) Readyng Final Phase 1 Collections for harvest
 - a. Collections identified but not yet harvested include the following:
 1. Digitized Juvenile Literature, FSU
 2. FSU Historic Photograph Collection, FSU
 3. International Archive of Women in Architecture Biographical Database, VT
(determined out of scope for project phase 1; will be harvested in extension)
 4. Kentucky Quilt Project image masters, Louisville
 5. Sam Nunn Constituent Files, Emory
 6. Southern Changes, Emory
 7. Special Collections and Archives Digital Image Master Files, Emory
 8. The Civil War in America from the Illustrated London News, Emory
 9. ACES Photographs, Auburn
 10. SMARTech (one AU), GA Tech
 - b. Collections whose harvest was incomplete in phase 1 include the following:
 1. BiblioTech, VA Tech
 2. Bugle Archive (all AUs), Va Tech
 3. History of Architecture Catalogue for Hypertext, VA Tech
 4. Library Friends, VA Tech

5. Mountain Slavery, VA Tech
6. Tin Horn, VA Tech
7. Spectrum, VA Tech
8. Virginia Libraries, VA Tech
9. Virginia Libraries, VA Tech
10. Papers of Judge Wm. M. Harris, VA Tech
11. My Precious Loulie, VA Tech
12. Conspectus Database, Emory
13. Southern Changes Masters, Emory
14. Southern Spaces-0, Emory
15. Southern Spaces-1, Emory
16. Southern Spaces Media Files, Emory
17. A Photographic Atlas of Selected Regions of the Milky Way, GA Tech
18. Georgia Tech Advertisements, GA Tech
19. SMARTech Collection (AUs 4, 12, 43, 346), GA Tech
20. Encompass, Auburn (moved and still unresolved)
21. Jean Thomas, Louisville
22. Bernheim Foundation interviews, Louisville

2) Harvesting Final Phase 1 Collections (see above—both new collections and misconfigured collections): **Complete by 02/29/08.**

3) Identifying Phase 2 Collections for Southern Digital Culture (SDC) Network

a. Initial Emory:

1. Southern Spaces (ongoing publication)
2. Southern Spaces Media (ongoing publication)
3. Lincoln Sermons
4. Dawson Collection
5. Highlander
6. AmRoutes
7. SouthComb
8. Emory ETDs
9. Emory Southern texts (Mass Digital Publications)

b. Initial FSU

1. FSU ETDs (11810 electronic theses and dissertations - pdf) – currently available online in DigiTool 3.1
2. FSU Undergraduate Honors in the Major Theses Collection (1927 electronic theses - pdf) – currently available online in DigiTool 3.1
3. FSU “Flying High” Circus Collection (155 individual B&W images (tiff)) - currently available online in DigiTool 3.1
4. FSU Special Collections EAD Finding Aids (70 EAD with PDF, HTML, and raw XML manifestations)
5. FSU Dept. of Oceanography technical reports (21 pdf) - currently available online in DigiTool 3.1
6. FSU Heritage Protocol (20 images jpg/tiff) - currently available online in DigiTool 3.1
7. Napoleon Collection – will be digitized in the future (METS, jpgs, jpeg2000)

c. Initial Louisville

1. Newton Owen Postcard Collection (18.2 GB)
2. Claude C. Matlack Collection (15.6 GB)

3. Kate Matthews Collection (18 GB)
 4. Kentucky Maps (19 GB)
 5. African-American Oral History Collection (74 GB)
 6. Macauley's Theatre Collection (20 GB)
 7. Arthur Younger Ford Photograph Albums (3.12 GB)
 8. Herald-Post photograph collection (19 GB)
- d. Initial GA Tech
 1. Aardvark (852 GB)
 2. Fulton Bag and Cotton Mills Digital Collection (11 GB)
 - e. Initial VA Tech
 1. Prevail Archives (25 GB)
 2. ETDs (100 GB)
 3. International Archive of Women in Architecture (IAWA) (2 GB)
 4. South Atlantic Humanities Center (SAHC) (.02 GB)
 5. VT IImageBase (220 GB)
 - f. Initial Auburn
 1. Auburn Sesquicentennial Lecture Series (8 GB)
 2. Eugene B. Sledge Papers (2.8 GB)
 3. Glomerata, continued (??)
 4. Auburn University Numbered Photographs (3.5 GB)
- 4) Reading Phase 2 Collections for SDC Network **(09/07 – 01/09)**
 - 5) Harvesting Phase 2 Collections for SDC Network (Staggered harvest with 3 collections for each institution by **03/31/08**; half of all collections for each institution by **06/30/08**; and all collections by **02/27/09**)
 - 6) Identifying Phase 2 Collections for additional Networks **(09/07 – 10/08)**
 - 7) Reading Phase 2 Collections for additional Networks **(09/07 – 01/09)**
 - 8) Harvesting Phase 2 Collections for additional Networks **(completed by 02/09)**

B. Network Expansion

- 1) Identifying Network topical needs of all members; deciding on new networks
 - a. Identified: Maps, Teaching Resources, ETDs, History of the Slave Trade **(09/07)**
 - b. Chosen: ETDs, History of the Slave Trade **(09/07)**
- 2) Determining best network structure for multi-“distributed archive” system **(by 03/08)**
 - a. Single LOCKSS daemon or multiple?
 - b. Single title database or multiple?
 - c. Single conspectus and cache manager or multiple?
 - d. All members storing all data in network or portions of data in network?

- 3) Improving Conspectus Database tool for use in multi-network system
 - a. Initial evaluation completed by Catherine Jannik, **09/07**
 - b. Secondary evaluation completed by Steering Committee, **02/08**
 - c. Final evaluation completed by Conspectus/Cache review committee, **02/08**
- 4) Establishing additional Networks
 - a. ETD network, by **05/08**
 - b. History of the Slave Trade, by **07/08**

C. Network Testing

- 1) Testing a single-node failure and full rebuild: Network Test #1 (**Emory, 02/08**)
- 2) Complete a formal testing design for ongoing 6-month tests and reports (**03/08**)
- 3) Network Test #2 (**07/08**)
- 4) Network Test #3 (**01/09**)

D. Network Reporting Tools

- 1) Evaluate existing tool (Cache Manager)
 - a. Initial evaluation completed by Catherine Jannik, 2007
 - b. Second evaluation completed by Monika Mevenkamp, 12/07-02/08
 - c. Final evaluation completed by Steering Committee, 02/08
- 2) Coordinate with LOCKSS team to evaluate their current cache manager and their ongoing development
- 3) Co-develop with LOCKSS a new Cache Manager that better serves the needs of the expanded systems

Harvesting

Harvesting activities will be conducted during Project Phase A1 (Content Harvest of collections not completed in the first project phase) and during Project Phase A5 (Subsequent Content Harvest). harvesting *must* be conducted at all six primary preservation sites, and *may* be conducted at the Library of Congress and other potential sites that may join the cooperative over time.

Next Steps Targets

Originally, we intended to complete all Phase 1 content harvest between 9/16/07 and 11/30/07. Due to the departure of the MetaArchive System Administrator from Emory in November and the Emory Node failure and rebuild later that month, we will now complete all Phase 1 content harvest by **02/29/08**.

Simultaneously, we will begin harvesting Phase 2 collections immediately. All partners should have at least three collections prepared and loaded (including plugins and manifest pages) no later than **03/31/08**. At least half of the collections that are intended to be contributed by each partner should be prepared and loaded by each partner (including plugins and manifest pages) by the mid-point of this project, or **06/30/08**. All Phase 2 collections should be harvested by **02/27/09**.

Notes on this document

This document was drafted by Katherine Skinner on 2008-01-17 and was completed at the second MetaArchive All-Project Meeting held at University of Louisville on Friday 2008-02-08, by the Steering Committee and additional attending project participants.