

June 9, 2009

The MetaArchive Cooperative is pleased to report to the NDLTD regarding the successful launch of our joint ETD pilot project begun one year ago. Since our last report in December 2008, we completed the Planning Phase and have made progress in the Curation Phase. We want to gather additional information about how ETDs are stored and learn more about the rate of growth of these collections before we move into the Preservation Phase.

As outlined in our Plan of Work, during this two year period (June 2008-June 2010) the MetaArchive Cooperative will, in cooperation with the NDLTD, accomplish the following goals:

1. Establish an ETD archive within the MetaArchive preservation network that ingests and preserves NDLTD members' electronic theses and dissertations
2. Document the best practices for ETD curation with the goal of long-term preservation
3. Scale up the preservation network by welcoming in new joint NDLTD/MetaArchive members

ETD Archive Task Force

- Diane Butler (Rice University)
- Bill Donovan (Boston University)
- Martin Halbert (Emory University)
- Gail McMillan (Virginia Tech), chair
- Beth Nicol (Auburn University)
- Katherine Skinner (Emory)
- Mark Stoffan (Florida St. U)
- Tyler Walters (Georgia Tech)

Curation Phase

The Curation Phase of the project enables the ETD Archive Task Force to study and document issues that arise as the Cooperative expands its network to include the new ETD Preservation Network. These include the following points:

1. ETD Curation for Preservation Toolkit: As we welcome new members into the ETD Archive, we will need a more specific set of documentation that informs NDLTD members of the best practices we recommend for preserving their ETD collections. Currently best practices are documented:

- *NDLTD Preservation Strategy*, 2nd ed, June 2009
<http://scholar.lib.vt.edu/theses/NDLTD/BoD200906/NDLTDPreservationPlan20090606.pdf>

- ETD 2009 presentations:

Avoiding the Calf-Path: Digital Preservation Readiness for Growing Collections and Distributed Preservation Networks. Martin Halbert, Katherine Skinner, and Gail McMillan. <http://conferences.library.pitt.edu/ocs/viewabstract.php?id=733&cf=7>






Distributed Digital Preservation for ETDs Workshop. Gail McMillan, Martin Halbert, Bill Donovan <http://www.library.pitt.edu/etd2009/ETDPreservationWorkshop200906.pdf>

- “Content Selection, Preparation, and Management,” Gail McMillan and Rachel Howard, forthcoming online book, *Guide to Distributed Digital Preservation Networks*. http://scholar.lib.vt.edu/theses/preservationGDDP_Chapter6DRAFT.pdf

2. Curation Tools: As we add a second archive to the network structure, we are enhancing our curation tools to support a multi-archive environment.

- Conspectus Database <http://www.metaarchive.org/conspectus/>
The MetaArchive network currently organizes its contents in one archive. Enhancements to the Conspectus Database will enable it to record what archive each collection belongs to, and what preservation nodes serve that archive. This tool is currently being tested (see examples below).

Conspectus archives: list			
	Title	Description	#Coll
<ul style="list-style-type: none"> archives list content_providers collections plugins 	ETD	Electronic Theses and Dissertations	4
	PREP	The Preparation Archive contains Collections/Archival Units that are prepared by content owners but are not yet ready to be tested Once a plugin is available and has been tested with the plugin tool and a locally running LOCKSS daemon (run_one_daemon) move them to the TEST archive.	0
	SDC	Southern Digital Culture Collections preserved in the MetaArchive production network -- do not move archival units/collections with preservation copies on production caches to the PREP Archive.	26
	TEST	The Test Archive contains Collections/Archival Units that are tested for the production network Once thoroughly tested move them to an appropriate production archive.	0

Conspectus content_providers: list					
	Name	PluginPrefix	#Plugins	#Collections	
<ul style="list-style-type: none"> archives content_providers list collections plugins 	 aub	Auburn University	edu.auburn	5	6
	 fsu	Florida State	edu.fsu	2	2
	 gt	Georgia Tech	edu.gatech	3	4
	 lou	University of Louisville	edu.louisville	3	3
	 vt	Virginia Tech	edu.vt	5	15

- Cache Manager: Enhancements will enable the Cache Manager to track the ingest and preservation procedures across archives and report to members. For example,

Collection: ETDs@VT

Plugin: edu|vt|library|theses

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

Base Uri2:

#Archival Units: 13

<u>Name</u>	<u>Base URL</u>	<u># Preserving</u>	<u>Preserving Caches</u>
ETDs@VT: year=1997	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=1998	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=1999	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2000	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2001	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2002	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2003	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2004	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2005	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2006	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2007	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2008	http://scholar.lib.vt.edu/theses/	8	aub-cap, em-cap, fsu, fsu-cap, gt-cap, hull, lou-cap, vt-cap
ETDs@VT: year=2009	http://scholar.lib.vt.edu/theses/	2	fsu, fsu-cap

Red shows that too few caches are preserving this collection. But FSU tried to ingest the 2009 ETDs before the collection was complete.

- Repository Tools: Through the ETD Preservation Survey (2008) we identified the range of repository software systems used by the ETD community. In 2009 we will study how to port content out of these major repository systems and into the ETD Archive and how to replace content into these repositories. These currently include,

- ETD-db: VT, FSU
- DSpace: Georgia Tech
- Fedora: Emory
- CONTENTdm: Auburn

In addition to MetaArchive activities specifically related to the ETD Pilot Project, it is benefiting from the continued maturation of the Cooperative, as seen in these activities from the last six months.

a. Market Analysis:

- Concentrated on the MetaArchive's primary market (university libraries) and two additional target markets (museums and archives)
- *Internet-based survey distributed through museum and archive listservs. Confirmed ETD Preservation Survey findings:
 - Degree of interest in community-based digital preservation services
 - Lack of digital preservation policies and readiness. KD/GMc to present at Library of Congress NDIIPP Partners on June 25

b. Systems development: Hub-and-Spoke joined by Cloud Computing (documented)

- No one institution should be a point of failure for entire network

c. Documentation: <http://www.metaarchive.org/resources.html>

d. Funding: Sustaining funds from new members. Awards from Library of Congress, NHPRC (IMLS, NEH)

ETD Preservation in the MetaArchive Cooperative

