Collaborative Preservation of ETDs: The MetaArchive Cooperative and LOCKSS

Gail McMillan
gailmac@vt.edu
Digital Library and Archives, Virginia Tech

Canadian ETD and Open Repositories Workshop
Atelier sur les TME et les dépôts à libre accès
I met a traveller from an antique land
Who said:—Two vast and trunkless legs of stone
Stand in the desert. Near them on the sand,
Half sunk, a shatter'd visage lies, whose frown
And wrinkled lip and sneer of cold command
Tell that its sculptor well those passions read
Which yet survive, stamp'd on these lifeless things,
The hand that mock'd them and the heart that fed.
And on the pedestal these words appear:
"My name is Ozymandias, king of kings:
Look on my works, ye mighty, and despair!"
Nothing beside remains: round the decay
Of that colossal wreck, boundless and bare,
The lone and level sands stretch far away.
Digital Preservation

- Systematic management of digital works over an indefinite period of time
  - Processes and activities that ensure the continued access to works in digital formats
  - Requires ongoing attention--constant input of resources: effort, time, money
- IR ≠ Digital Preservation
- Backups ≠ Digital Preservation
Backups ≠ Digital Preservation

• Backups are tactical measures
  • Make copies to restore originals after data loss event.
  • Typically stored in a single location
    • Often nearby
    • Collocated with the servers backed up
• Backups address short-term data loss with minimal investment resources
**Digital Preservation is Strategic**

- Long-term, error-free storage and for the entire time span the information is required.
- Realistically address issues in preserving information over time
  - Affordable ongoing investment
  - Geographically dispersed set of secure caches
  - Multi-institutional collaboration through formal agreements
- DDPN: Distributed Digital Preservation Network
What the LOCKSS software does:

• **Gathers** web content into (6) caches as a journal issue is completed/published.

• **Compares** content among caches and detect/repair discrepancies.

• **Preserves** the contents of each cache for posterity by never flushing it.

• **Serves** content to readers from the publisher, or, if necessary, from the cache.
THE GREATEST THREAT

to digital assets is not fire, flood or theft. It’s the assumption that cultural memory organizations have taken the requisite steps to preserve them.

Too often, we haven’t.

Which is why the MetaArchive Cooperative is fostering a community-based approach to digital preservation. Our members participate in a secure, cost-effective solution to the threat of data loss. As a community, we are building our own trustworthy digital repository and becoming skilled cultural stewards of the digital age.
Distributed Digital Preservation Network

- **Secure**
  - Reduces the likelihood that any single cache will be compromised.

- **Distributed Geographically**
  - Reduces likelihood that loss of any single cache will lead to loss of the preserved information.
  - A single organization is unlikely to have the capability to operate several geographically dispersed and securely maintained servers.
  - Inter-institutional agreements will ensure commitment to act in concert over time.
MetaArchive Cooperative: DDPN

- Library of Congress, supported since 2003
- LOCKSS without public access
  - Separated preservation from access
- Bit-level preservation
  - Format agnostic
  - Images, text, multimedia, datasets, program executables, etc.
- 16 members (150 collections): US, UK, Brazil
The Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations.

This website contains information about the initiative, how to set up Electronic Thesis and Dissertation (ETD) programmes, how to create and locate ETDs, and current research in digital libraries related to the NDLTD and ETDs.

- for Researchers: Find ETDs
- for Authors: Submit ETDs
- for Institutions: Manage ETDs

News
- NDLTD ETD Awards 2010 Winners Announced 2010-03-29
- ETD 2010 Blog 2010-03-23
- ETD 2010 Conference - Last Call for Papers 2010-03-09
- Open Access Week - October 18-24, 2010 2010-02-03
- ETD 2010 Conference - Call for Papers 2010-01-22

Upcoming Events
- ETD 2010 - 13th International Symposium on Electronic Theses and Dissertations Austin, Texas, 2010-06-16
- Previous events...
- Upcoming events...

Website development funded by UNESCO, Adobe Systems Inc. and NDLTD members.
NDLTD/MetaArchive Alliance
ETD Preservation Survey

• Dec. 2007-April 2008
• 95 institutions responded
• 80% have ETDs
• 27% have a preservation plan
• 92% interested in DDPN

http://lumiere.lib.vt.edu/surveys/results/
NDLTD Preservation Guidelines


- Hardware, software
- Metadata: Conspectus Database of Collections
- Organizing ETD collections—best practices and data wrangling
- Institutional Workflow
- Personnel
- Training Opportunities
- Documentation and Reports
- Retrieving from the ETD Archive
NDLTD Preservation Strategy: MetaArchive


• Preliminaries
  • Join, training, installation

• Collection readiness
  • Manageable units, permission, define path

• Harvest/ingest/cache
  • Continual comparisons, repairs

• Dark archive: PPDN
MetaArchive Membership Levels

- Preservation Members
  - Fundamental activity: network node server
  - Preserve their own and others ETDs
- Sustaining Members
  - Preservation member responsibilities
  - Steering Committee, leadership, and technical development
Technical & Organizational Solutions

http://www.metaarchive.org/resources

Cooperative Charter
Membership Agreement
Technical Specifications
MetaArchive Trusted
Repository Audit
Market Analysis
Management Plan
Operations Plan

Risk Assessment
Financial Plan
Extension Harvest Plan
Outreach Program
Implementation Plan
Data and Node Recovery
Conspectus Database
Conspectus Schema
The MetaArchive is a Cooperative.

• Not a vendor.
  
  A *cooperative* is an organization of individuals or institutions who have joined together to perform a function more efficiently than each individual could do alone. The purpose of a cooperative is not to make profits, but to improve each member's situation and the situation of the surrounding society.

• A collaborative association of cultural memory organizations with a nonprofit administration.

• Membership fees go to a central pool of support for members’ co-op activities.

• All hardware and software assets are owned by the members.
ETD Preservation

• Plan for it; don’t put off for another day
• Separate issues: access and preservation
• Collaboration: everybody works and everybody benefits