Distributed Digital Preservation: The MetaArchive Model

ACRL
Poster 18

Safeguarding Culture and Identity

Content Owner
Web Content: any format

Restricted Access:

ETDs
Data Sets
Oral Histories

Public Access:

ETDs
Faculty Archives
Images, Maps
Music, Narratives
Exhibits
University Publications
University Archives

MetaArchive

Conspectus Database:

Collection Descriptions
Harvest Location
Accrual rules

Plugin Repository:

Plugin rules for locating, harvesting, and verifying content

Collection Preparation

Conspectus Database: Describe collections in a bibliographic-like database, including how frequently to copy new material (crawl rules).

In a Nut Shell

MetaArchive uses the free open source LOCKSS software to operate **dark archives** in a private network of preservation servers.

Content is stored in and restored to its original format thereby avoiding data loss.

Participating institutions identify their digital resources to be preserved and make them available ONLY to the MetaArchive for dark archiving.

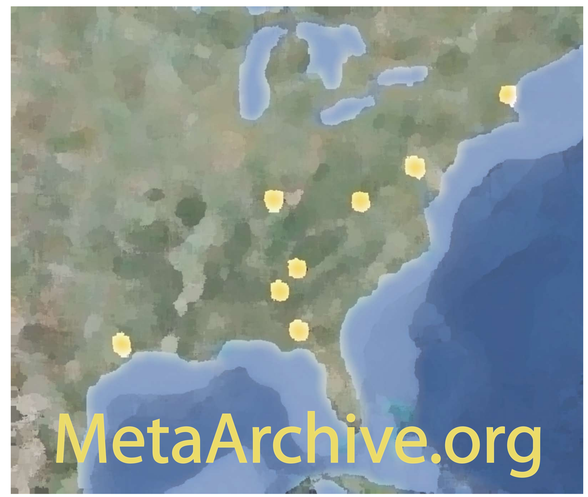
Membership Categories

Affordable for libraries of all sizes.

• Contributing :	$300/year	5 GB
• Preservation:	$2,000/year	20 GB
• Sustaining:	$5,000/year	40 GB
Additional Storage		$2/GB

5% discount for joining both NDLTD (Networked Digital Library of Theses and Dissertations) and the MetaArchive.

Copies in Multiple Dark Archives



MetaArchive

Conspectus Database:

Collection Descriptions
Harvest Location
Accrual rules

Plugin Repository:

Plugin rules for locating, harvesting, and verifying content

Content Owner

Restricted Access:

ETDs
Data Sets
Oral Histories

Public Access:

ETDs
Faculty Archives
Images, Maps
Music, Narratives
Exhibits
University Publications

Ingest Content

At least 6 caches replicate content.

LOCKSS Software running on each cache ingest content based on crawl rules .

Cache VA Tech | Cache Emory | Cache FSU | Cache Rice | Cache Boston College | Cache Auburn

Automated Ingest, Replicate, and Restore

Content Monitoring and Recovery

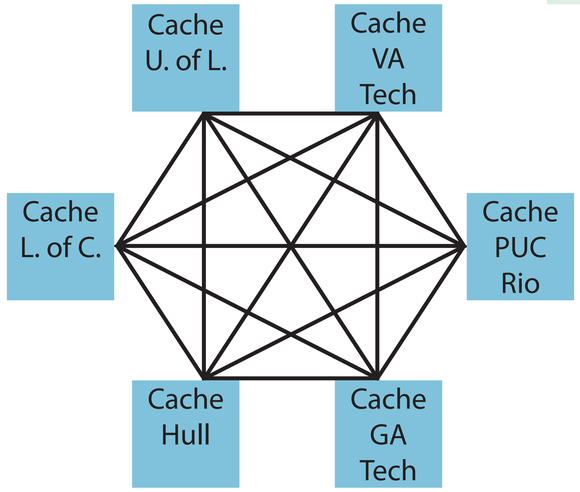
LOCKSS daemons on all caches periodically update content from the web, validate content via polls and votes, and repair content when necessary.

The cache manager monitors all LOCKSS caches for uptime, availability of disk space, and sufficient replication of content.

Cache Contents: Dark Archive

ETDs
Data Sets
Oral Histories

Faculty Archives
Images, Maps
Music, Narratives
Exhibits
University Publications
University Archives



Preservation Scenario

Librarian: Let's preserve this web resource and its metadata.

Techy: Please describe it in the Conspectus Database.

Plugin Developer: Do I harvest all or parts?

Librarian: Everything!

Plugin Developer: develops rules (plugins) that locate the requested content, e.g., master TIFFs, but not JPGs

Techy: I'll tell the network (publicize plugin, activate content, ...)

Librarian: Let me see content in the network (AuditProxy)

Together: Verify that preserved content is sufficient to restore.

AFTER A FATAL CRASH

Techy: I restored your content!