Preservation as a Process

MetaArchive and Distributed Digital Preservation

Sam Meister
Deanna Ulvestad

OhioDIG Meeting
March 9, 2016
MetaArchive History

- Founded 2004
- Distributed digital preservation cooperative
- Preservation aims: prevent loss and corruption from human malice/error or from a disaster
- First (known) preservation network to preserve special collections/unique materials
Hallmarks

- Distributed digital preservation
- Institutions maintain control over their own content
- Preservation as a process, not a push-button exercise
- Simplicity in ingest, management
Membership

- Auburn University
- Boston College
- Cal Poly San Luis Obispo
- Consorci de Biblioteques Universitaris de Catalunya
- Florida State University
- Isabella Stewart Gardner Museum
- Greene County Public Library
- HBCU Library Alliance
- Indiana State University
- Oregon State University
- Penn State University
- Pontificia Universidade Catolica do Rio de Janeiro
- Purdue University
- Rockefeller Archives Center
- University of Louisville
- University of North Texas
- University of South Carolina
- Virginia Tech University
MetaArchive Practices

- Basic processes
  - “Producer” (OAIS) determines curation practices; brings SIPs to MetaArchive
  - Multiple copies of AIPs dispersed across geographical, political, and environmental lines
  - Checks and repairs automated across network
  - Deaccession cycle versus data deletion
MetaArchive Pricing

- Three membership levels
  - Collaborative members: $2.5K/year
  - Preservation members: $3K/year
  - Sustaining members: $5.5K/year
- Server cost: <$5K/term
- Storage cost: $585/TB/year
Membership Responsibilities

● Undertake a 3-year membership term
● Take responsibility for content preparation, evaluation, staging, and ingest testing
● Monitor collections to ensure accurate long-term preservation
● Host and maintain a MetaArchive cache (server) or pay in a technology support fee
● Consider contributing to Committees!
Cooperative Preservation

- MetaArchive is a cooperative, not a vendor:
  - All hardware and software assets are owned by members
  - Membership fees and storage fees go to a central pool of support for members’ co-op activities
Philosophy in Practice

- Compatible with any repository system
  - E.g., Dspace, Fedora, Archivalware, ETDdb, CONTENTdm, BePress, Digital Commons, etc
- Member institutions determine their own curatorial practices
- MetaArchive is a community of support to help them make informed decisions
Ingest Demo / Overview
Prepare SIP
Stage Collection

- Collections consist of Archival Units (one or many)
Stage Collection

AU
AU
AU
AU
AU
AU
Stage Collection

- Collections consist of Archival Units (one or many)
- Archival Units contain content and metadata
Stage Collection

ARCHIVAL UNIT

Content + Metadata
Stage Collection

- Collections consist of Archival Units (one or many)
- Archival Units contain content and metadata
- Collections organized to be able to restore collections later
- Include documentation on restoration procedures
- Make collection web accessible at URL
Stage Collection

http://metaarchive-staging.lib.calpoly.edu
Create Collection

- Create collection level metadata for collection in Conspectus management tool
  - Title
  - Archive
  - Description
  - Base URL
# Create Collection

## Collection Details

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Cal Poly University Archives digitized masters</td>
</tr>
<tr>
<td>Archive</td>
<td>GEN</td>
</tr>
<tr>
<td>Plugin</td>
<td>edu.calp.bagitplugin2 by California Polytechnic</td>
</tr>
<tr>
<td>Base URL</td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
</tr>
<tr>
<td>Description</td>
<td>Digital masters of digitized documents, photographs, sound recordings, and video from the Robert E. Kennedy Library's University Archives</td>
</tr>
<tr>
<td>Repository Type</td>
<td>Not specified</td>
</tr>
</tbody>
</table>
Create Manifest Page

- Simple HTML page with basic collection description information and links to collection content for LOCKSS crawlers
- LOCKSS Crawlers MUST find permission statement to be able to harvest content
Create Manifest Page

COLLECTION_NAME

Content preserved by MetaArchive Network.

LOCKSS system has permission to collect, preserve, and serve this Archival Unit.

Collection Info:

- Conspectus Collection(s): COLLECTION_NAME
- Institution: INSTITUTION
- Contact Info: CONTACT_PERSON

COLLECTION_DESCRIPTION

Links for LOCKSS to start its crawl:

- LINK_NAME - LINK_INFO
Create Manifest Page

- Simple HTML page with basic collection description information and links to collection content for LOCKSS crawlers
- LOCKSS Crawlers MUST find permission statement to be able to harvest content
- Place Manifest page on same host as content
Create Manifest Page

http://metaarchive-staging.lib.calpoly.edu/mabagitmanifest.html
Develop Collection Plugin

● Plugins tell member caches where to find a designated Manifest page and how far to follow the links to harvest collection content
Develop Collection Plugin

Plugin:
edu.calp.bagitplugin2

Member Cache

http://metaarchive-staging.lib.calpoly.edu
Develop Collection Plugin

- Member creates new plugin via Conspectus based on existing plugin, or uploads custom plugin
Develop Collection Plugin
Develop Collection Plugin

- Member creates new plugin via Conspectus based on existing plugin, or uploads custom plugin
- Member gives plugin a unique name
- Member defines plugin rules to determine which files will be harvested
Develop Collection Plugin

```xml
<map>
  <entry>
    <string>plugin_config_props</string>
    <list>
      <org.lockss.daemon.ConfigParamDescr>
        <key>start</key>
        <displayName>Start</displayName>
        <type>1</type>
        <size>20</size>
        <definitional>true</definitional>
        <defaultOnly>false</defaultOnly>
      </org.lockss.daemon.ConfigParamDescr>
      <org.lockss.daemon.ConfigParamDescr>
        <key>base_url</key>
        <displayName>Base URL</displayName>
        <type>3</type>
        <description>Usually of the form http://&lt;journal-name&gt;.com/</description>
      </org.lockss.daemon.ConfigParamDescr>
    </list>
  </entry>
</map>
```
Test & Review
Test Collection Plugin

- Member tests plugin locally and makes changes as needed
- Member defines Plugin name and Archival Units in Conspectus
# Test Collection Plugin

## Plugin Details

<table>
<thead>
<tr>
<th>Name</th>
<th>edu.calp.bagitplugin2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Provider</td>
<td>California Polytechnic</td>
</tr>
<tr>
<td>Plugin XML (Production)</td>
<td>No production version available</td>
</tr>
<tr>
<td>Plugin XML (Testing)</td>
<td>Plugin XML for edu.calp.bagitplugin2</td>
</tr>
<tr>
<td>Plugin XML History</td>
<td>View all Plugin XML revisions for this Plugin</td>
</tr>
<tr>
<td>Retired</td>
<td>False</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Name</th>
<th>Type</th>
<th>Definitional</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>start</td>
<td>String</td>
<td>Yes</td>
<td>Start</td>
</tr>
<tr>
<td></td>
<td>base_url</td>
<td>URL</td>
<td>Yes</td>
<td>Base URL</td>
</tr>
<tr>
<td></td>
<td>path</td>
<td>String</td>
<td>No</td>
<td>Path</td>
</tr>
</tbody>
</table>
Test Collection Plugin

Manage Archival Units for Collection

Collection Title | Cal Poly University Archives digitized masters
Base URL         | http://metaarchive-staging.lib.calpoly.edu
Plugin           | edu.calg.bagitplugin2 by California Polytechnic

<table>
<thead>
<tr>
<th>Plugin Parameters</th>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>start</td>
<td>String</td>
<td>Start</td>
</tr>
<tr>
<td></td>
<td>base_url</td>
<td>URL</td>
<td>Base URL</td>
</tr>
<tr>
<td></td>
<td>path</td>
<td>String</td>
<td>Path</td>
</tr>
</tbody>
</table>

Archival Units

<table>
<thead>
<tr>
<th>Parameter Values</th>
<th>Status</th>
<th>Copies</th>
<th>Up/Down</th>
</tr>
</thead>
<tbody>
<tr>
<td>base_url</td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
<td>Preserved</td>
<td>Site Up</td>
</tr>
<tr>
<td>path</td>
<td>dm-ua-sel_bag001_split</td>
<td></td>
<td></td>
</tr>
<tr>
<td>start</td>
<td>dm-ua-sel_bag001_0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Test Collection Plugin

<table>
<thead>
<tr>
<th>base_url</th>
<th><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>path</td>
<td>dm-ua-sel_bag003_split</td>
</tr>
<tr>
<td>start</td>
<td>dm-ua-sel_bag003__metadata</td>
</tr>
</tbody>
</table>

(Unsaved)

<table>
<thead>
<tr>
<th>start</th>
<th>Draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>start</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>path</th>
<th>Draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>path</td>
<td>-</td>
</tr>
</tbody>
</table>

[Add AU](#)  

Do nothing with selected AUs

[Save All Changes](#)
Review Plugin & Test Ingest

- Member requests plugin review and test by MetaArchive staff
- MetaArchive staff ingests collection to test network
Review Plugin & Test Ingest
Review Plugin & Test Ingest

- Member requests plugin review and test by MetaArchive staff
- MetaArchive staff ingests collection to test network
- MetaArchive staff sends member test ingest report to review
SIP to AIP
Commit Plugin

- If test ingest is successful, MetaArchive staff commits plugin to production plugin repository
Commit Plugin

```
<map>
  <entry>
    <string>plugin_config_props</string>
    <list>
      <org.lockss.daemon.ConfigParamDescr>
        <key>start</key>
        <displayName>Start</displayName>
        <type>1</type>
        <size>20</size>
        <defaultOnly>false</defaultOnly>
      </org.lockss.daemon.ConfigParamDescr>
      <org.lockss.daemon.ConfigParamDescr>
        <key>base_url</key>
        <displayName>Base URL</displayName>
        <description>Usually of the form http://&lt;journal-name&gt;.&lt;gt;com/</description>
        <type>3</type>
        <size>48</size>
      </org.lockss.daemon.ConfigParamDescr>
    </list>
  </entry>
</map>
```
Complete Collection Metadata
Make collection available to network

- MetaArchive staff regenerate LOCKSS Title Database to expose collection to production network
- MetaArchive staff assigns six geographically distributed caches to crawl and harvest the collection
Replicate collection

### Related Items

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Metadata Import/Export</th>
<th>Plugin XML</th>
<th>Archival Units</th>
<th>Parameter Values</th>
<th>Status</th>
<th>Copies</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>base_url</td>
<td>preserved</td>
<td>7</td>
<td>Up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>path</td>
<td>dm-ua-sel_bag001_split</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>start</td>
<td>dm-ua-sel_bag001_0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>base_url</td>
<td>preserved</td>
<td>7</td>
<td>Up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>path</td>
<td>dm-ua-sel_bag001_split</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>start</td>
<td>dm-ua-sel_bag001_1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>base_url</td>
<td>preserved</td>
<td>7</td>
<td>Up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>path</td>
<td>dm-ua-sel_bag001_split</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>start</td>
<td>dm-ua-sel_bag001__metadata</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>base_url</td>
<td>preserved</td>
<td>7</td>
<td>Up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>path</td>
<td>dm-ua-sel_bag002_split</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>start</td>
<td>dm-ua-sel_bag002_0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>base_url</td>
<td>preserved</td>
<td>7</td>
<td>Up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>path</td>
<td>dm-ua-sel_bag002_split</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>start</td>
<td>dm-ua-sel_bag002_1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Replicate collection

<table>
<thead>
<tr>
<th>Status</th>
<th>Preserved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability</td>
<td>Site Up</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Values</th>
<th>base_url</th>
<th><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>path</td>
<td>dm-ua-sel_bag002_split</td>
</tr>
<tr>
<td></td>
<td>start</td>
<td>dm-ua-sel_bag002_2</td>
</tr>
</tbody>
</table>

## Preservation Status

<table>
<thead>
<tr>
<th>Cache</th>
<th>Size</th>
<th>DiskUsage (MB)</th>
<th>Last Crawl</th>
<th>Crawl Result</th>
<th>Last Poll</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>crmu</td>
<td>146.8 MB</td>
<td>154324992.0</td>
<td>Nov. 10, 2015, 4:06 p.m.</td>
<td>Successful</td>
<td>Jan. 4, 2016, 5:27 p.m.</td>
<td>100.0 %</td>
</tr>
<tr>
<td>sc-fe</td>
<td>146.8 MB</td>
<td>154169344.0</td>
<td>Nov. 10, 2015, 3:35 p.m.</td>
<td>Successful</td>
<td>Jan. 7, 2016, 5:15 a.m.</td>
<td>100.0 %</td>
</tr>
<tr>
<td>is-erack</td>
<td>146.8 MB</td>
<td>154169344.0</td>
<td>Nov. 10, 2015, 6:38 p.m.</td>
<td>Successful</td>
<td>Jan. 5, 2016, 11:17 p.m.</td>
<td>100.0 %</td>
</tr>
<tr>
<td>utk</td>
<td>146.8 MB</td>
<td>154169344.0</td>
<td>Nov. 10, 2015, 4:46 p.m.</td>
<td>Successful</td>
<td>Jan. 5, 2016, 3:35 a.m.</td>
<td>100.0 %</td>
</tr>
<tr>
<td>unt</td>
<td>146.8 MB</td>
<td>154324992.0</td>
<td>Oct. 12, 2015, 7:46 p.m.</td>
<td>Successful</td>
<td>Jan. 9, 2016, 12:01 p.m.</td>
<td>100.0 %</td>
</tr>
<tr>
<td>cal-poly</td>
<td>146.8 MB</td>
<td>154169344.0</td>
<td>Nov. 10, 2015, 6:37 p.m.</td>
<td>Successful</td>
<td>Dec. 31, 2015, 1 p.m.</td>
<td>100.0 %</td>
</tr>
<tr>
<td>puc-ibm</td>
<td>146.8 MB</td>
<td>154173440.0</td>
<td>Nov. 10, 2015, 7:20 p.m.</td>
<td>Successful</td>
<td>Jan. 7, 2016, 4:18 p.m.</td>
<td>100.0 %</td>
</tr>
</tbody>
</table>
Auditing Control

Parameter Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>base_url</td>
<td><a href="http://metaarchive-staging.lib.calpoly.edu">http://metaarchive-staging.lib.calpoly.edu</a></td>
</tr>
<tr>
<td>path</td>
<td>dm-ua-sel_bag002_split</td>
</tr>
<tr>
<td>start</td>
<td>dm-ua-sel_bag002_0</td>
</tr>
</tbody>
</table>

Preservation Status

Use the options below to request an audit on this archival unit. When the process completes, a new Report will be created.

Auditing Tools

Bag Checksum Audit

Descriptive text here

Cache

cmu

The LOCKSS cache on which to generate hashes.

Baglt Manifest

Browse… manifest-md5.txt

Upload a Baglt manifest for this AU. Must be named either manifest-md5.txt or manifest-sha1.txt. The hashing algorithm is detected based on the filename.

Request Audit
Auditing Control

Report Details

Created On: Jan. 12, 2016, 3:51 p.m.
Requested By: sysadmin
Related To: Archival Unit in Cal Poly University Archives digitized masters

Found 4 hashes in HashCUS file
Compared 4 records from uploaded manifest, encountered 0 mismatches.
Encountered 0 errors

Show Report Details
Voting and Polling

Cf2304e9b416e4c6e4d7a1bb22bf95e4  data/ua-sel_00000268-M.wav
046763d382e557359731edc1d5a8b821  data/dm-ua-sel_bag_002.MODSmetadata.xml
cf9beab2c63082d0d0b40ce9a8faa0a6  data/ua-sel_00000265_001-M.wav
733298738956be7ff4d9ed6b5d021e56  data/ua-sel_00000259-M.wav
Voting and Polling
Voting and Polling
Damage and Repair

Cf2304e9b416e4c6e4d7a1bb22bf95e4  data/ua-sel_00000268-M.wav
046763d382e557359731edc1d5a8b821  data/dm-ua-sel_bag_002_MODSmetadata.xml
cf9beab2c63082d0d0b40ce9a8faa0a6  data/ua-sel_00000265_001-M.wav
733298738956be7ff4d9ed6b5d021e56  data/ua-sel_00000259-M.wav

Cf2304e9b416e4c6e4d7a1bb22bf95e4  data/ua-sel_00000268-M.wav
046763d382e557359731edc1d5a8b821  data/dm-ua-sel_bag_002_MODSmetadata.xml
cf9beab2c63082d0d0b40ce9a8faa0a6  data/ua-sel_00000265_001-M.wav
733298738956be7ff4d9ed6b5d021e57  data/ua-sel_00000259-M.wav
Hi there. Remind me, have we talked before?
Damage and 

Yep. We go way back.
Thanks!

metaarchive.org

sam@educopia.org

@samalanmeister
November 2010
Attended 5-day workshop
“Digital Preservation Management”
University of Michigan

August 2011
Compared Digital Preservation
Repository options

April 2012
Joined MetaArchive as a
Preservation Member

January 2013
Started ingesting collections

Greene County Public Library was housed in the Carnegie building from 1906 – 1978. Xenia, Ohio.
Why MetaArchive

- Transparent
- Affordable
- Community-based
- Supportive
- Diverse

First bookmobile used by the Greene County Public Library from 1948 – 1958. Xenia, Ohio.
Modified Ingest

- MetaArchive Server
- MetaArchive Server
- MetaArchive Server
- MetaArchive Server
- MetaArchive Server

MetaArchive Server Replicates GCPL Archive Units
Cost of a Digital Time Capsule....

Library Paid in 2015

Greene County Courthouse Time Capsule of 1901 opened in 2001. Xenia, Ohio.

<table>
<thead>
<tr>
<th>Service</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservation Membership</td>
<td>$3,000</td>
</tr>
<tr>
<td>Technology Fee</td>
<td>1,000</td>
</tr>
<tr>
<td>Storage</td>
<td>1,800</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Total MetaArchive Fees 2015</td>
<td>$4,800</td>
</tr>
</tbody>
</table>