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

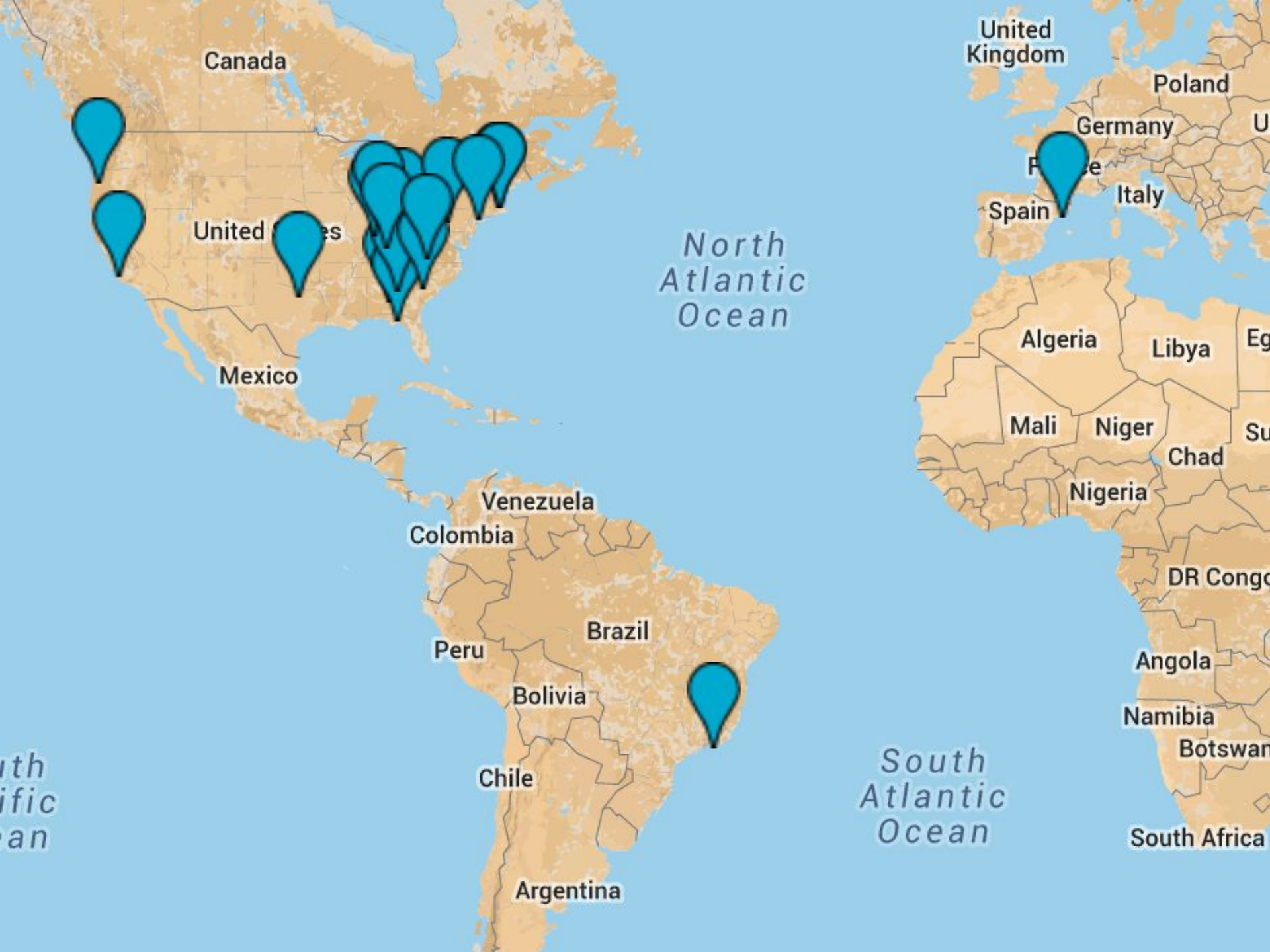
Hallmarks

- 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

Meta Archive
COOPERATIVE


Canada
United States
Mexico
North Atlantic Ocean
United Kingdom
Poland
Germany
Spain
Italy
Algeria
Libya
Mali
Niger
Chad
Nigeria
DR Congo
Angola
Namibia
South Africa
Venezuela
Colombia
Brazil
Peru
Bolivia
Chile
Argentina
South Atlantic Ocean

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

Meta Archive COOPERATIVE

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

MetaArchive COOPERATIVE

Philosophy in Practice

- Compatible with any repository system
  - E.g., Dspace, Fedora, Archivalware, ETDb, CONTENTdm, BePress, Digital Commons, etc
- Member institutions determine their own curatorial practices
- MetaArchive is a community of support to help them make informed decisions

MetaArchive COOPERATIVE

Ingest Demo / Overview

Meta Archive
COOPERATIVE

Prepare SIP

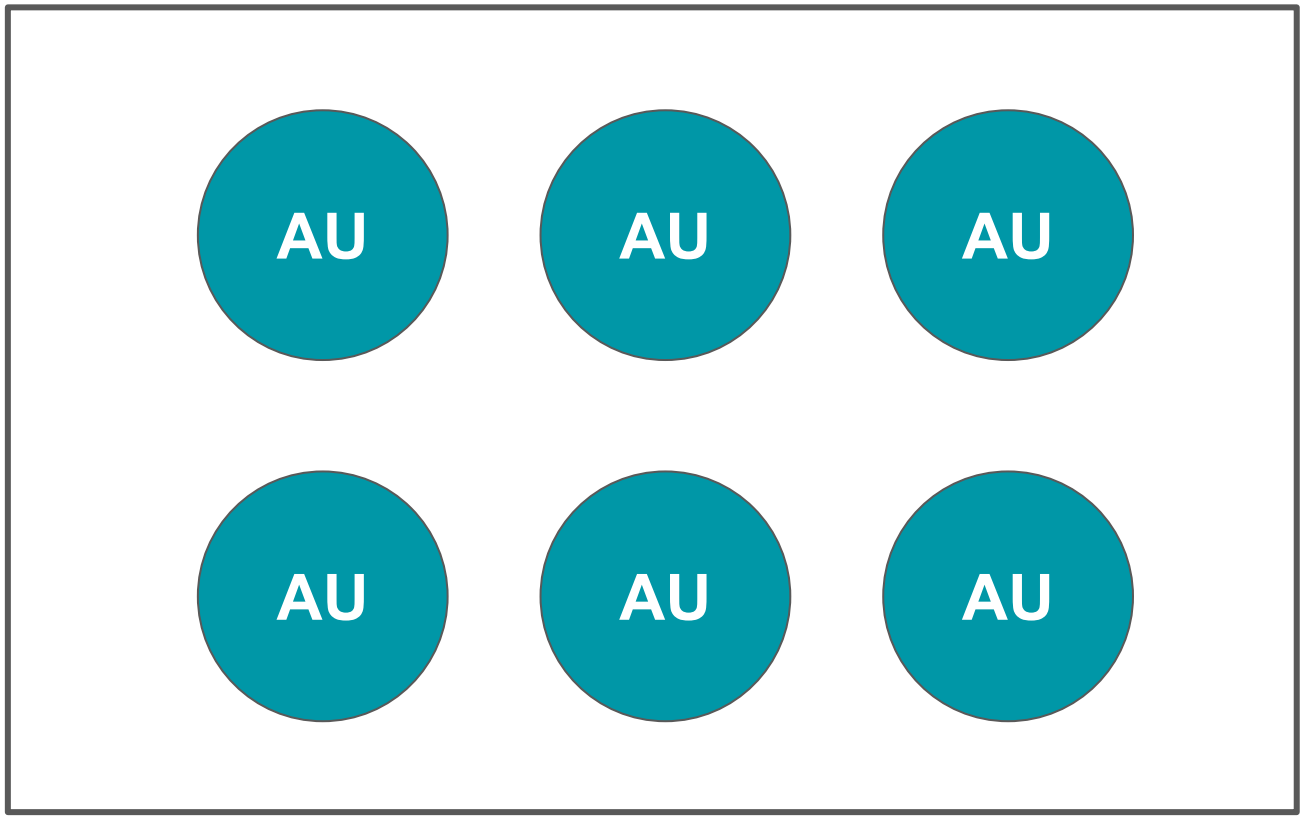
Meta Archive
COOPERATIVE

Stage Collection

- Collections consist of Archival Units (one or many)

Meta Archive
COOPERATIVE

Stage Collection



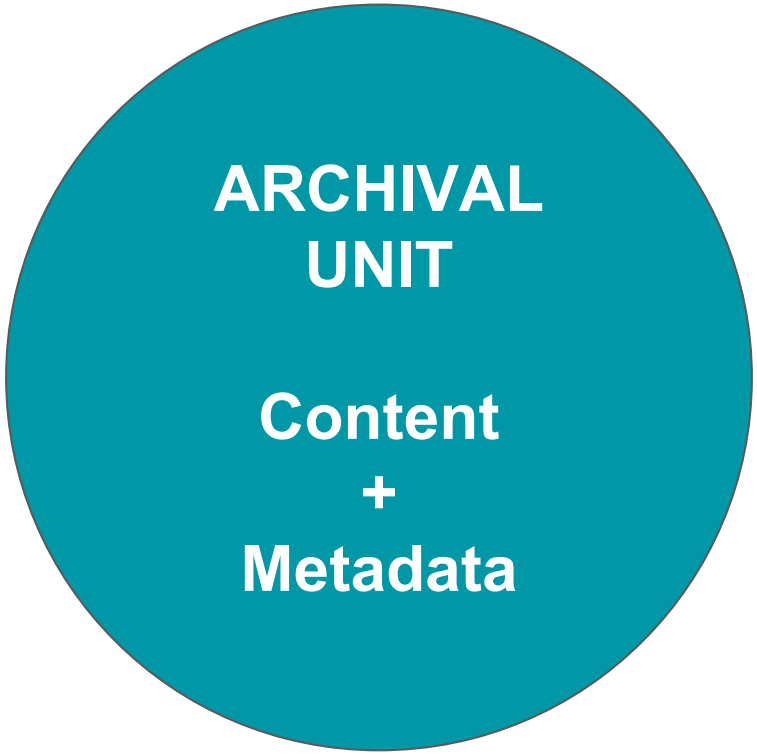
Meta Archive
COOPERATIVE

Stage Collection

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

Meta Archive
COOPERATIVE

Stage Collection



Meta Archive
COOPERATIVE

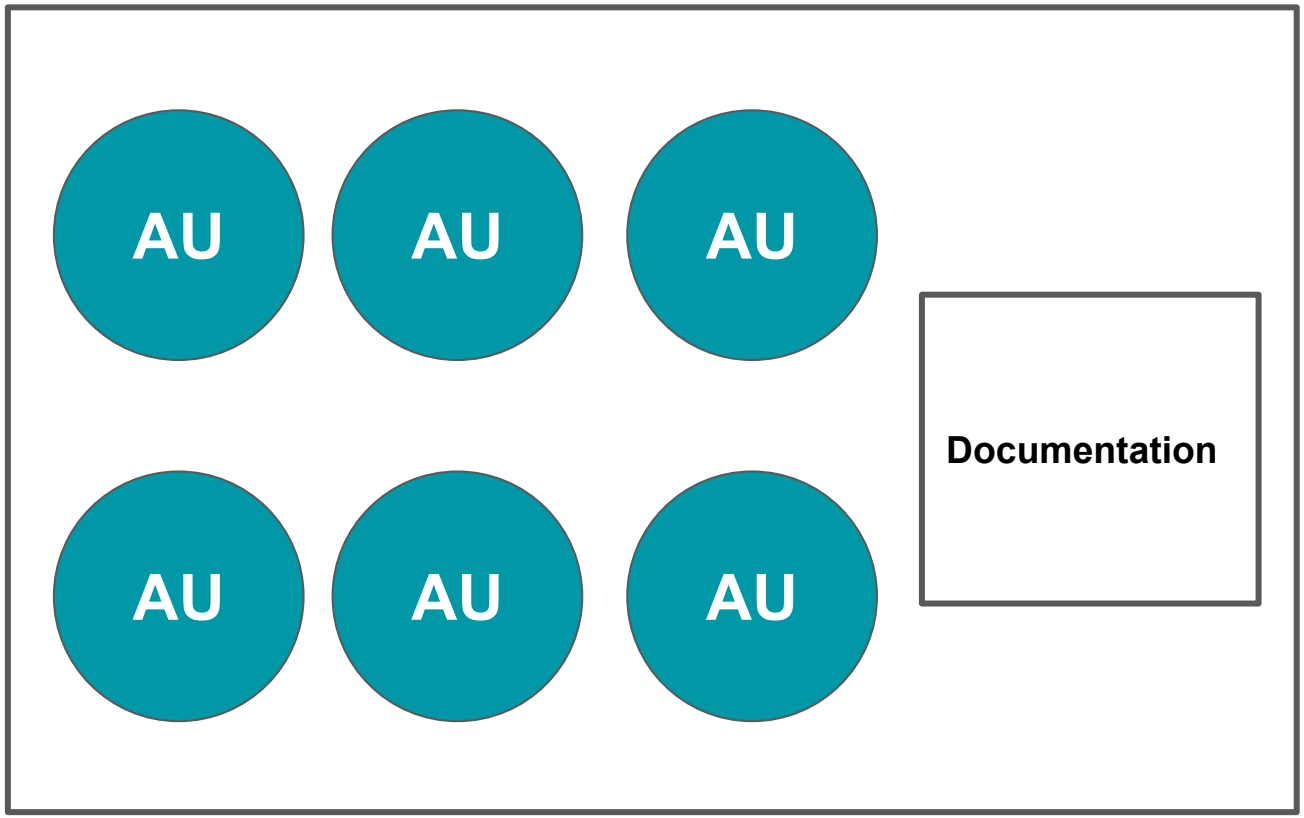
Stage Collection

- 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

Meta Archive
COOPERATIVE

Stage Collection

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



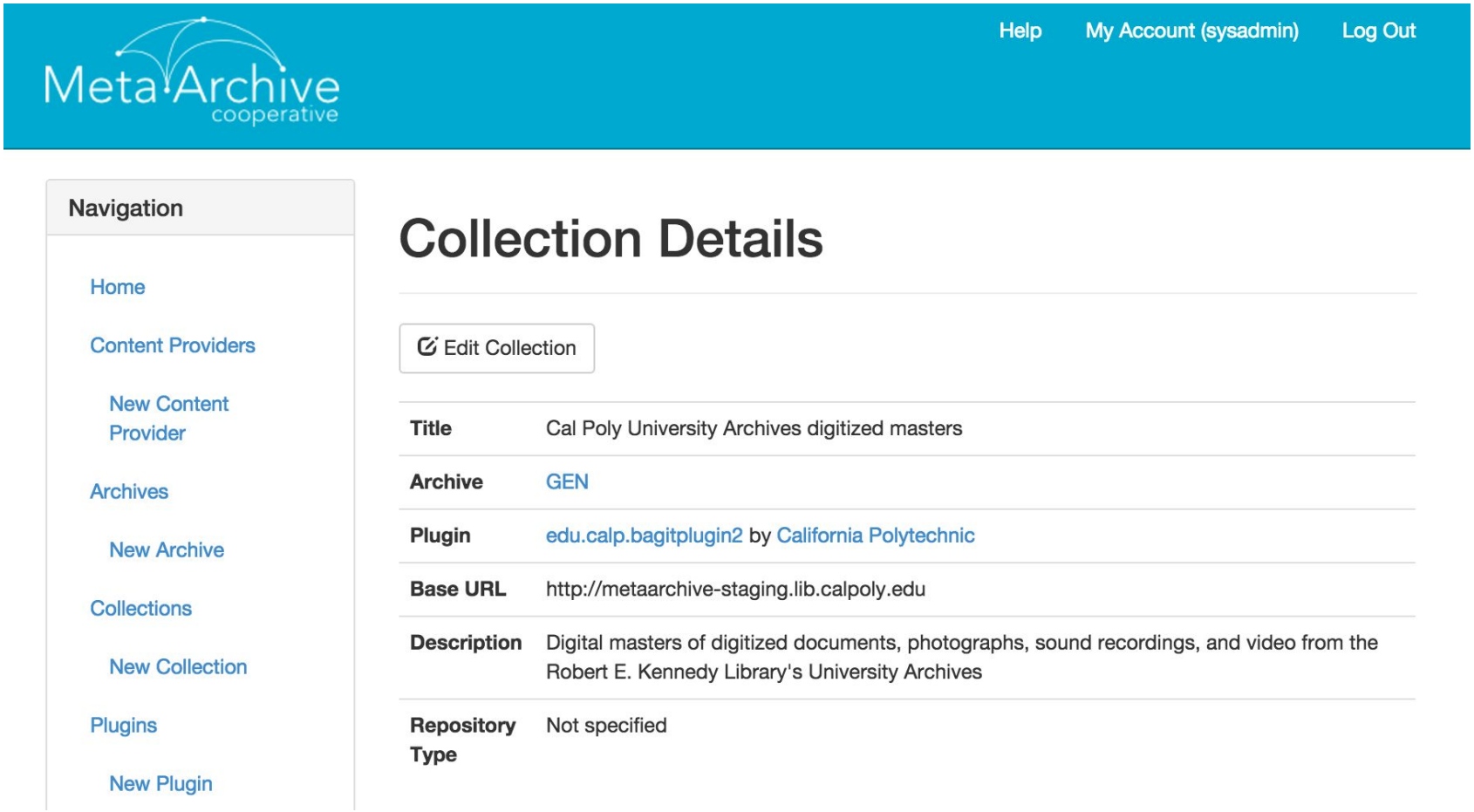
MetaArchive COOPERATIVE

Create Collection

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

Meta Archive
COOPERATIVE

Create Collection



Meta Archive COOPERATIVE

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

Meta Archive
COOPERATIVE

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

MetaArchive COOPERATIVE

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

Meta Archive
COOPERATIVE

Create Manifest Page

http://metaarchive-staging.lib.calpoly.edu/**mabagitmanifest.html**

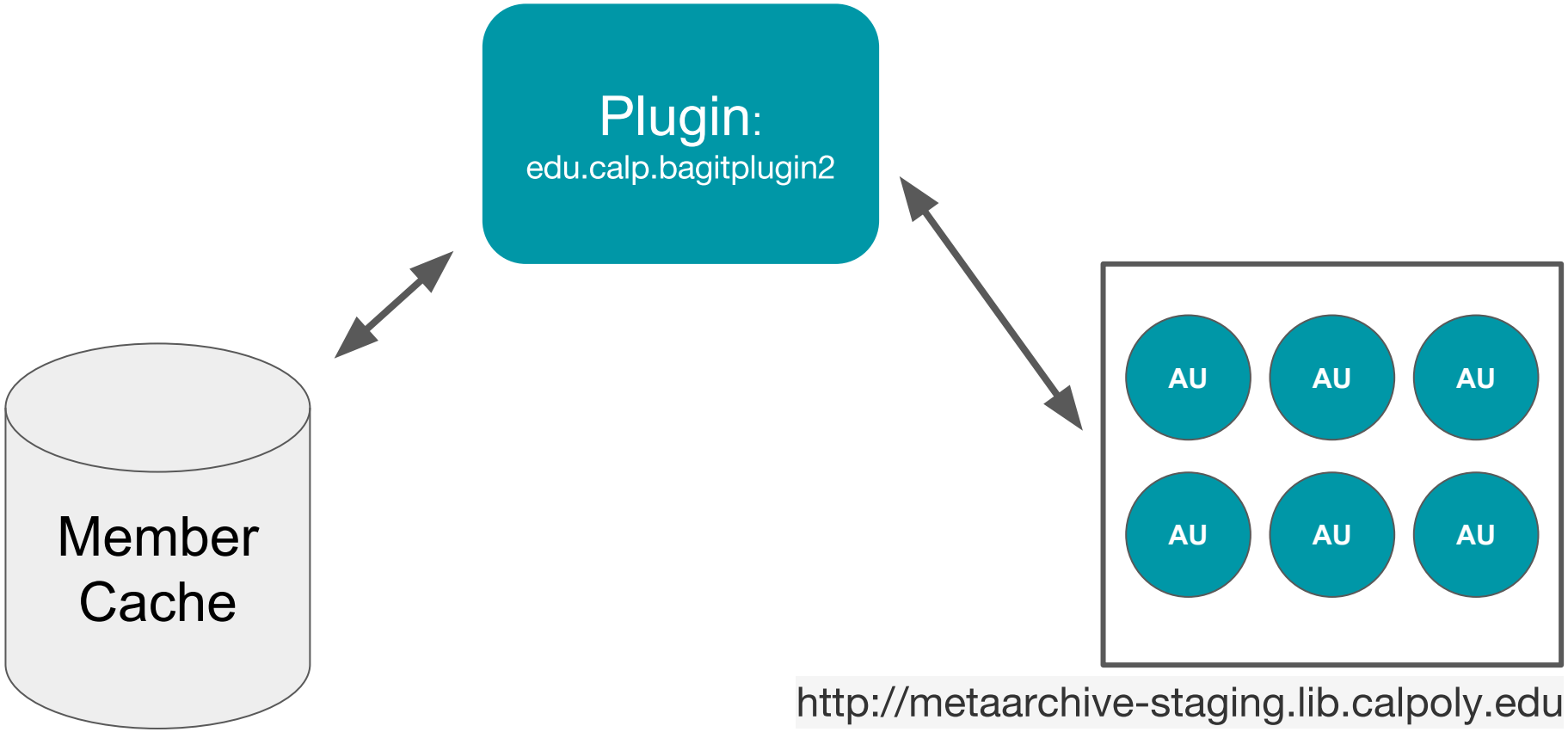
Meta Archive
COOPERATIVE

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

Meta Archive
COOPERATIVE

Develop Collection Plugin



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

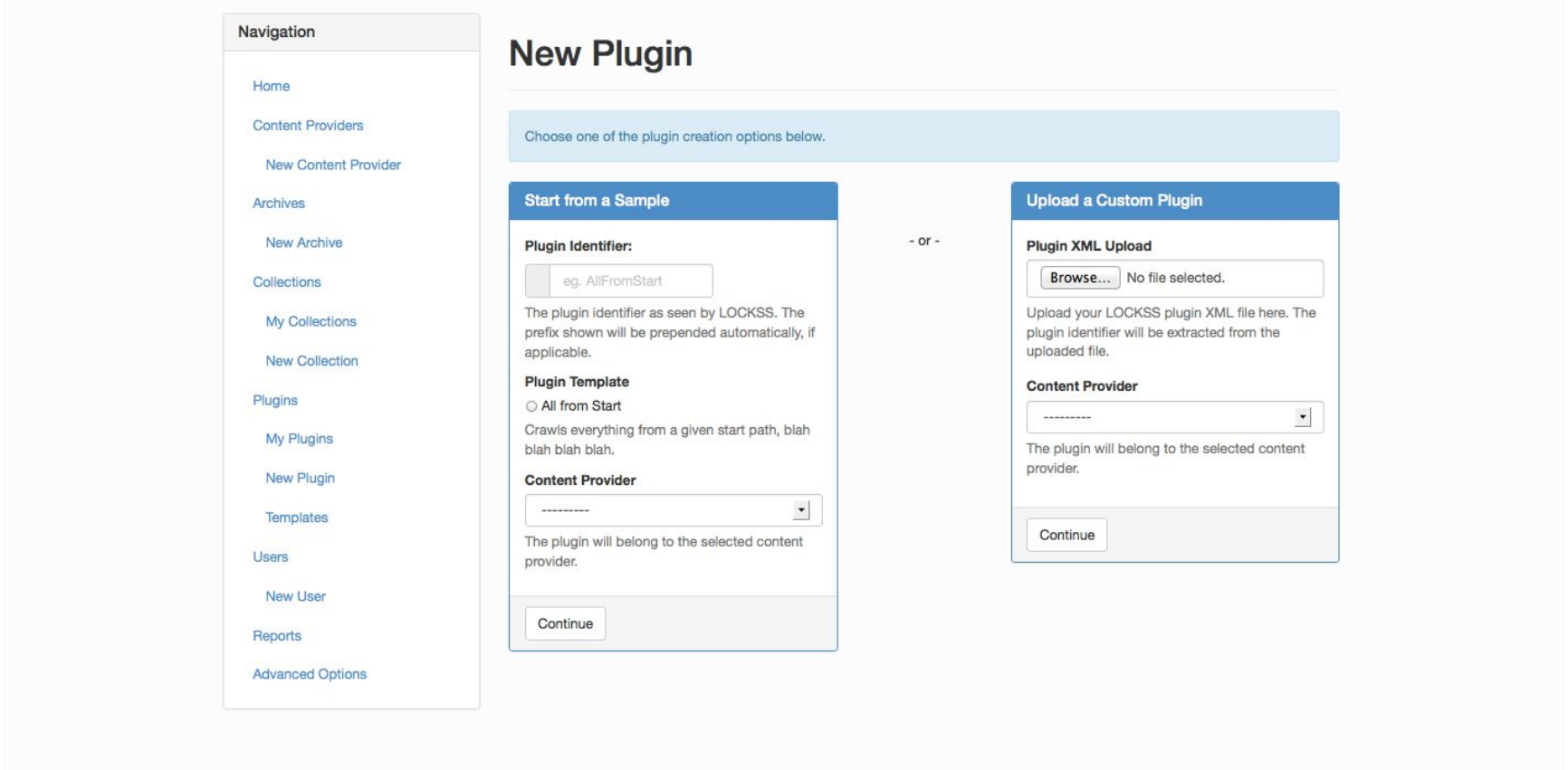
MetaArchive COOPERATIVE

Develop Collection Plugin

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

Meta Archive
COOPERATIVE

Develop Collection Plugin



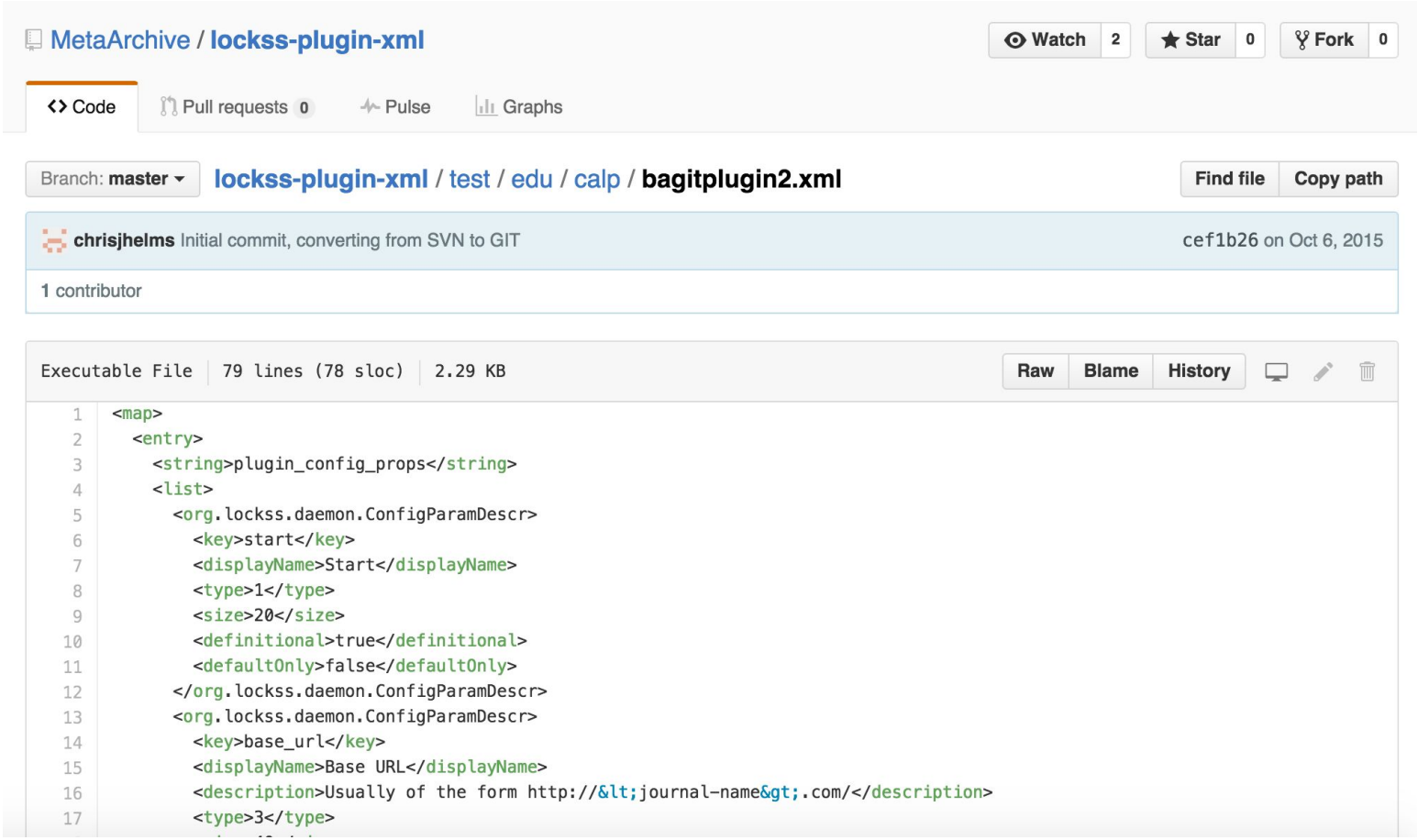
MetaArchive COOPERATIVE

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

Meta Archive
COOPERATIVE

Develop Collection Plugin



MetaArchive COOPERATIVE

Test & Review

Test Collection Plugin

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

Meta Archive
COOPERATIVE

Test Collection Plugin

Navigation

- Home
- Content Providers
  - New Content Provider
- Archives
  - New Archive
- Collections
  - New Collection
- Plugins
  - New Plugin
  - Templates

Plugin Details

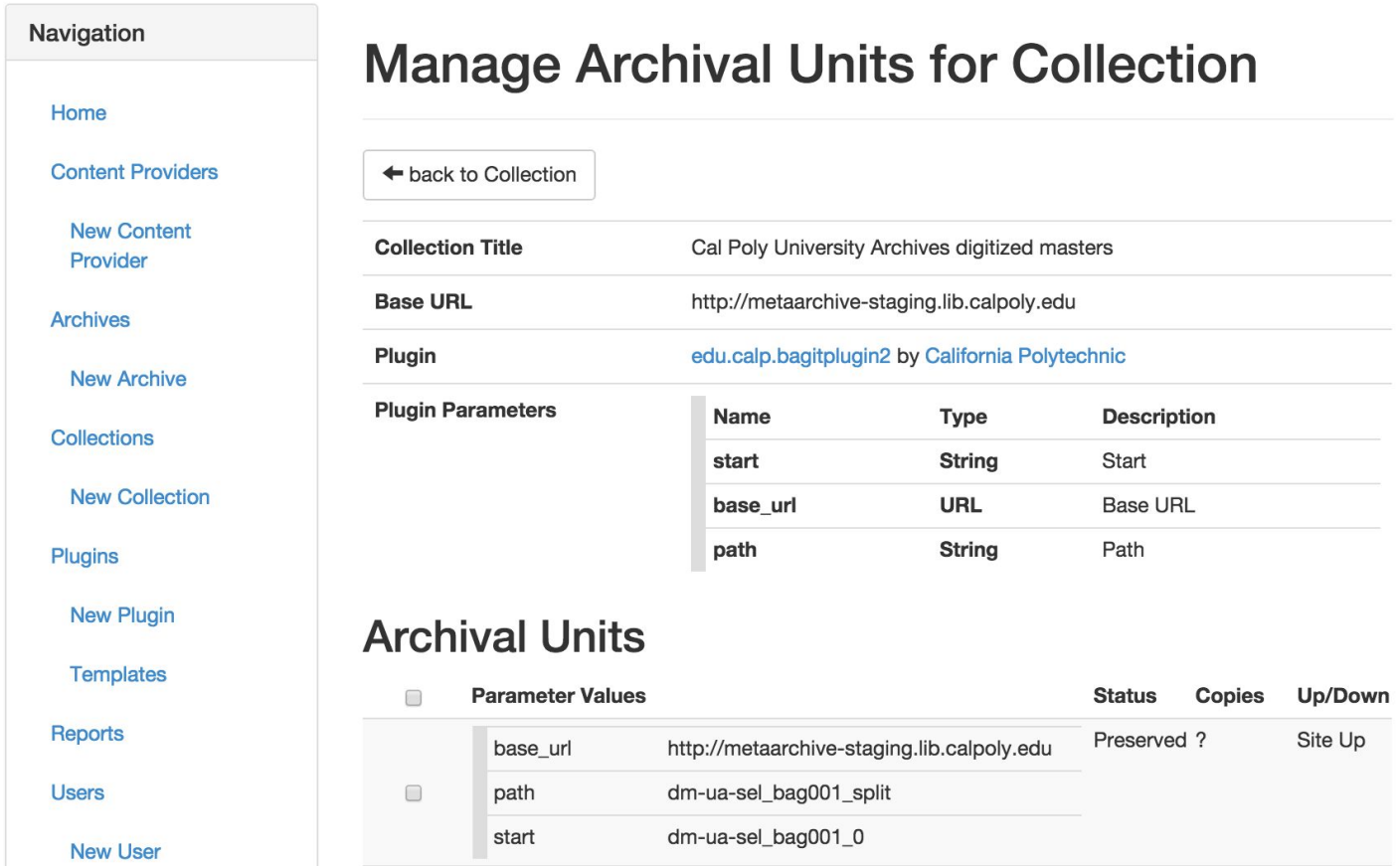
Name	edu.calp.bagitplugin2
Content Provider	California Polytechnic
Plugin XML (Production)	No production version available
Plugin XML (Testing)	Plugin XML for edu.calp.bagitplugin2
Plugin XML History	View all Plugin XML revisions for this Plugin
Retired	False

Parameters

Name	Type	Definitional	Description
start	String	Yes	Start
base_url	URL	Yes	Base URL
path	String	No	Path

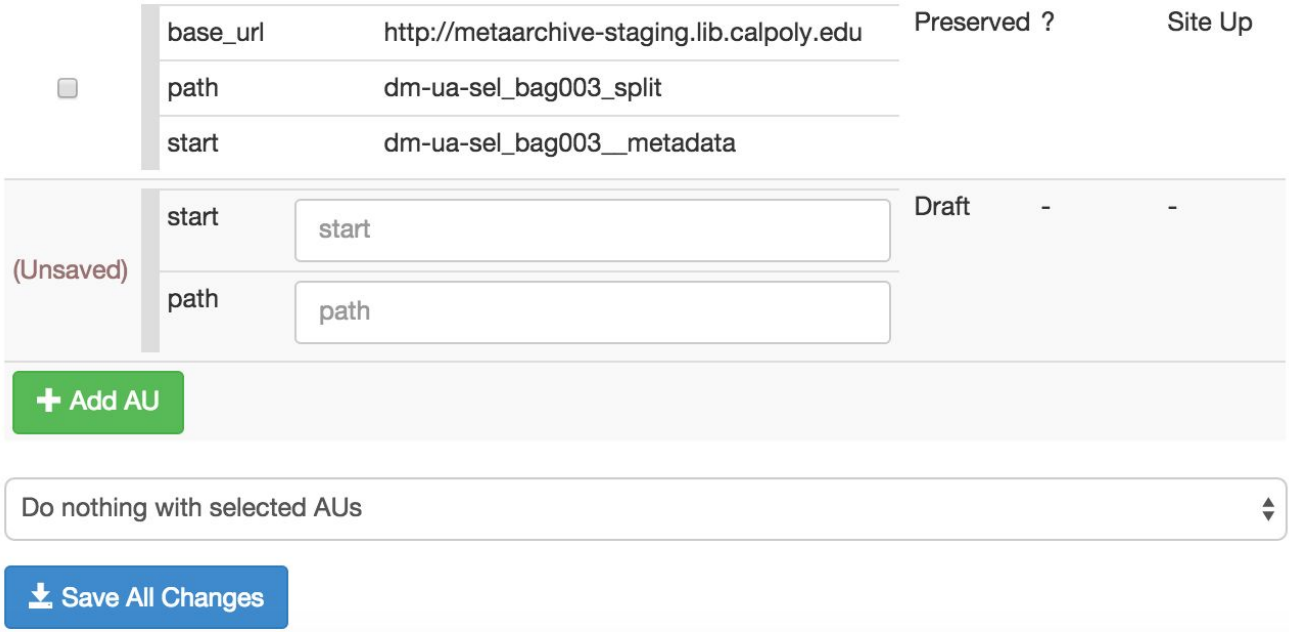
Meta Archive COOPERATIVE

Test Collection Plugin



MetaArchive COOPERATIVE

Test Collection Plugin



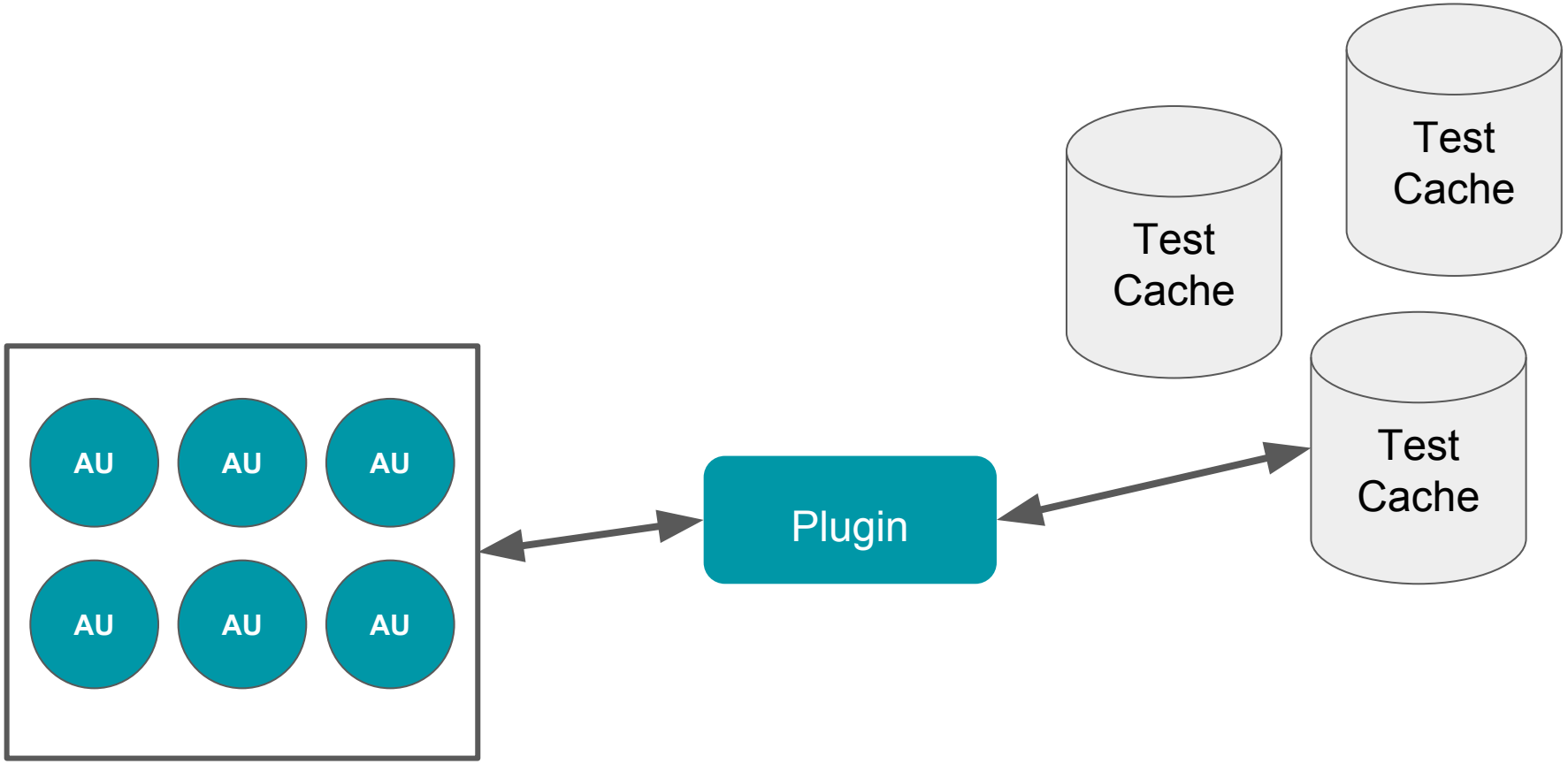
MetaArchive COOPERATIVE

Review Plugin & Test Ingest

- Member requests plugin review and test by MetaArchive staff
- MetaArchive staff ingests collection to test network

MetaArchive COOPERATIVE

Review Plugin & Test Ingest



Meta Archive
COOPERATIVE

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

MetaArchive
COOPERATIVE

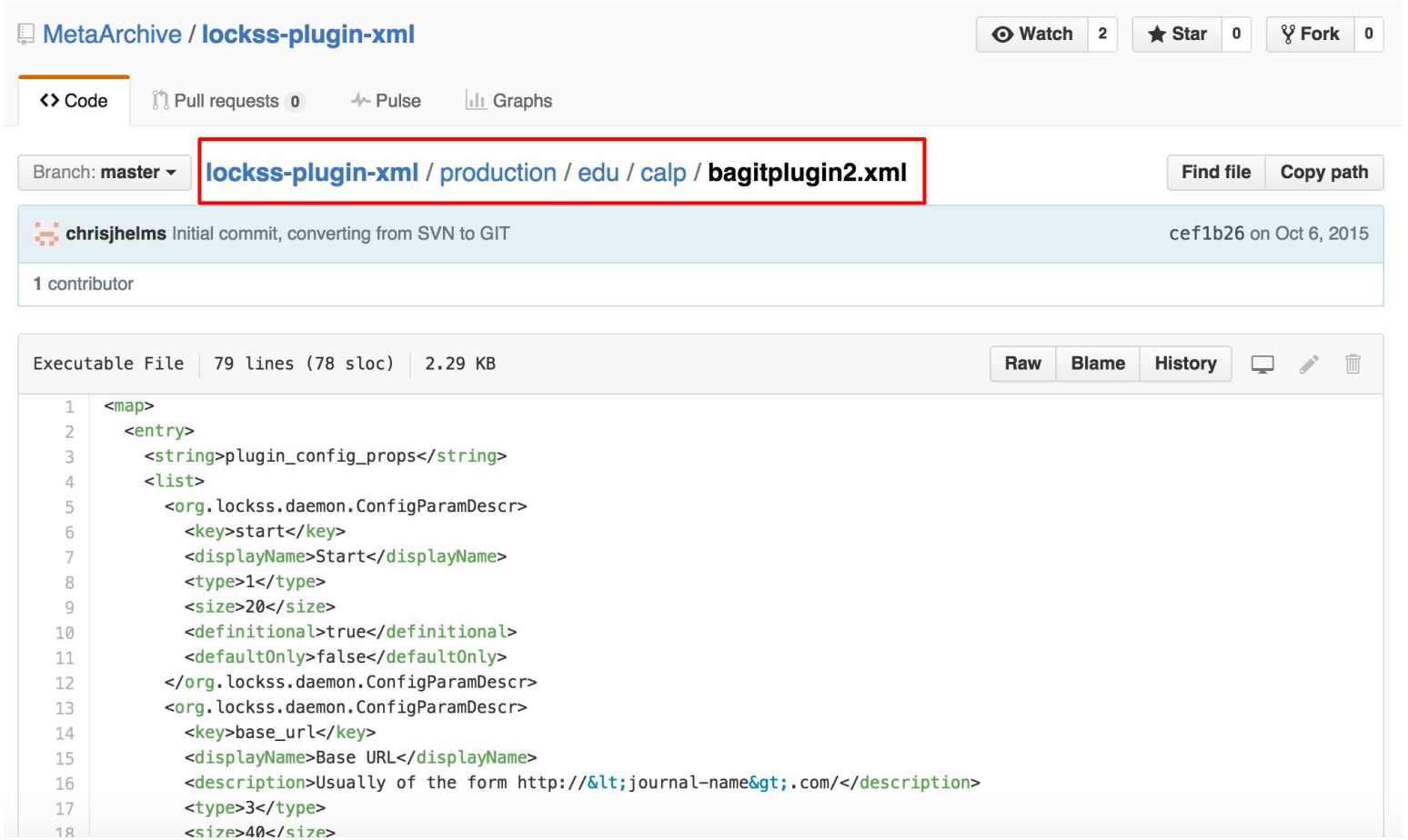
SIP to AIP

Commit Plugin

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

MetaArchive COOPERATIVE

Commit Plugin



MetaArchive COOPERATIVE

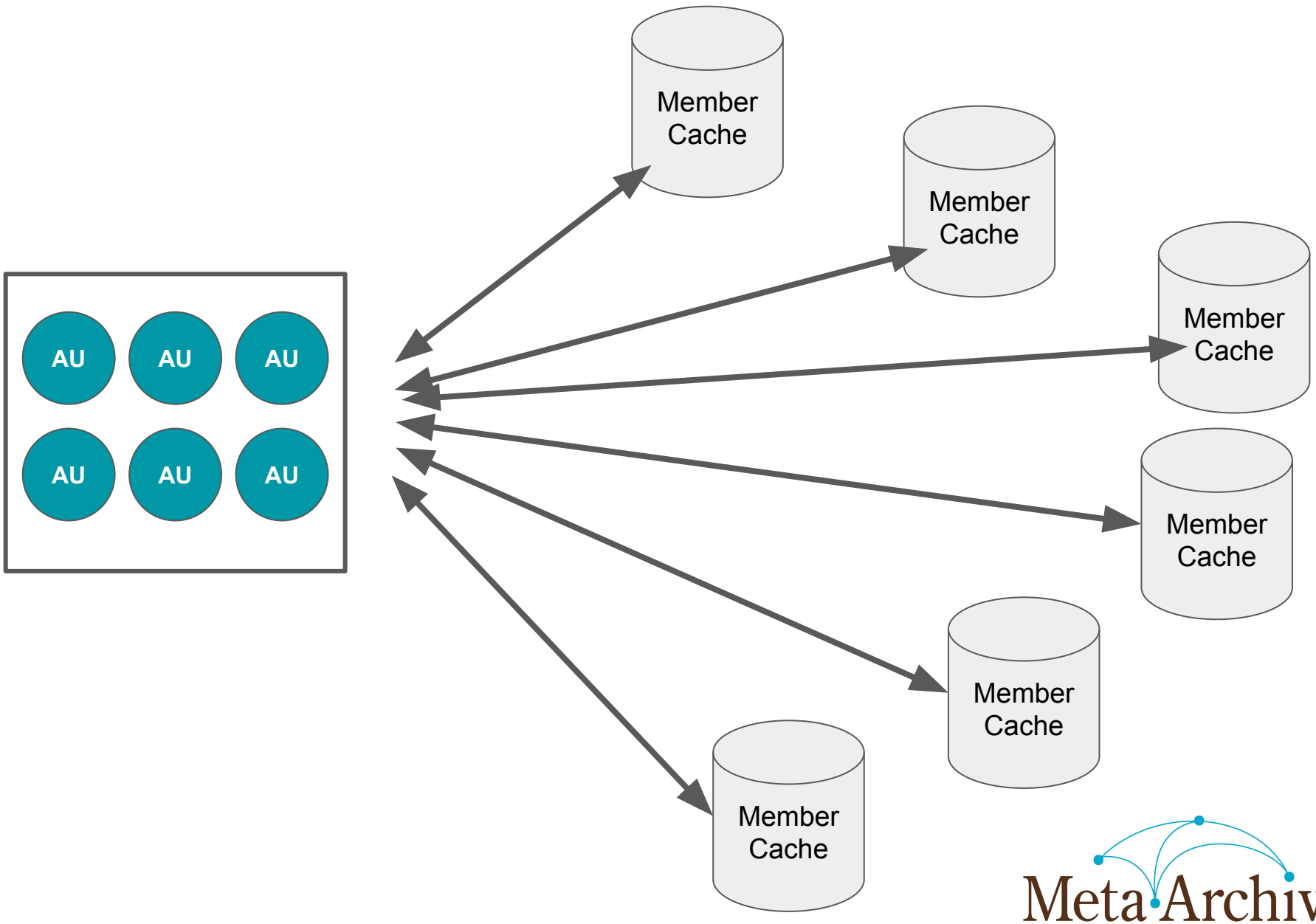
Complete Collection Metadata

Meta Archive
COOPERATIVE

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

MetaArchive COOPERATIVE


AU
AU
AU
AU
AU
AU
Member Cache
Member Cache
Member Cache
Member Cache
Member Cache
Member Cache
Meta Archive
COOPERATIVE

Replicate collection

Related Items

Metadata | Metadata Import/Export | Plugin XML | Archival Units 14

Parameter Values		Status	Copies	
base_url	http://metaarchive-staging.lib.calpoly.edu	preserved	7	Site Up
path	dm-ua-sel_bag001_split			
start	dm-ua-sel_bag001_0			
base_url	http://metaarchive-staging.lib.calpoly.edu	preserved	7	Site Up
path	dm-ua-sel_bag001_split			
start	dm-ua-sel_bag001_1			
base_url	http://metaarchive-staging.lib.calpoly.edu	preserved	7	Site Up
path	dm-ua-sel_bag001_split			
start	dm-ua-sel_bag001__metadata			
base_url	http://metaarchive-staging.lib.calpoly.edu	preserved	7	Site Up
path	dm-ua-sel_bag002_split			
start	dm-ua-sel_bag002_0			
base_url	http://metaarchive-staging.lib.calpoly.edu	preserved	7	Site Up
path	dm-ua-sel_bag002_split			
start	dm-ua-sel_bag002_1			

MetaArchive COOPERATIVE

Replicate collection

Status	Preserved
Availability	Site Up

Parameter Values

base_url	http://metaarchive-staging.lib.calpoly.edu
path	dm-ua-sel_bag002_split
start	dm-ua-sel_bag002_2

Preservation Status | Auditing Tools

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

MetaArchive COOPERATIVE

Auditing Control

Parameter Values		
	base_url	http://metaarchive-staging.lib.calpoly.edu
	path	dm-ua-sel_bag002_split
	start	dm-ua-sel_bag002_0

Preservation Status | Auditing Tools

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

Bag Checksum Audit

Descriptive text here

Cache

cmu

The LOCKSS cache on which to generate hashes.

BagIt Manifest

Browse… manifest-md5.txt

Upload a BagIt 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

MetaArchive COOPERATIVE

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

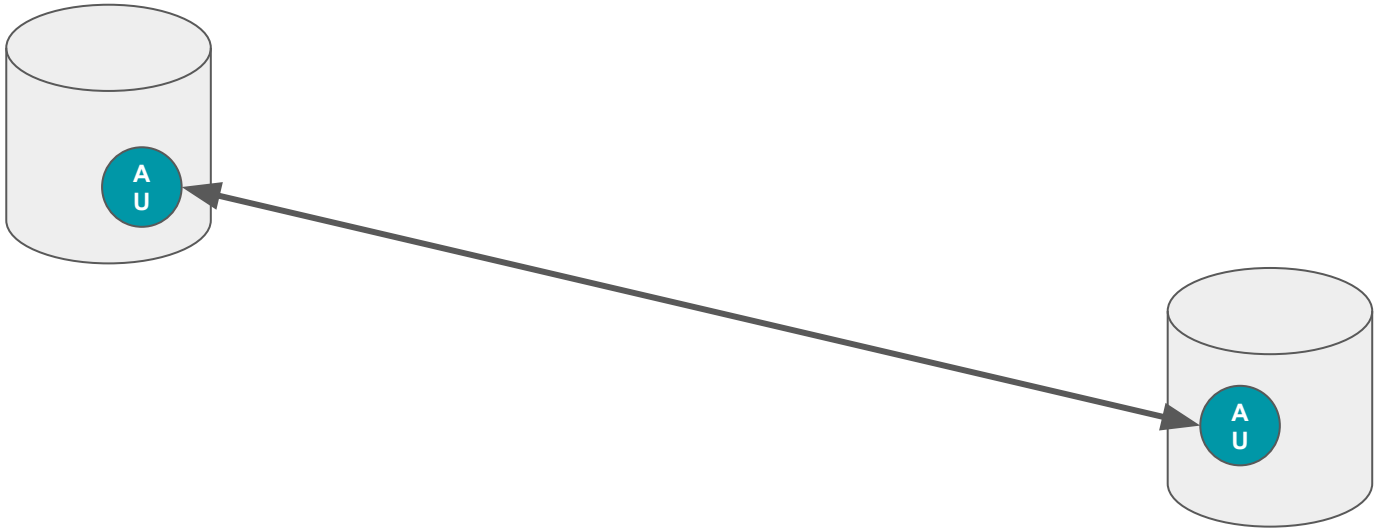
```
Found 4 hashes in HashCUS file
Compared 4 records from uploaded manifest, encountered 0 mismatches.
Encountered 0 errors
```

Show Report Details

MetaArchive COOPERATIVE

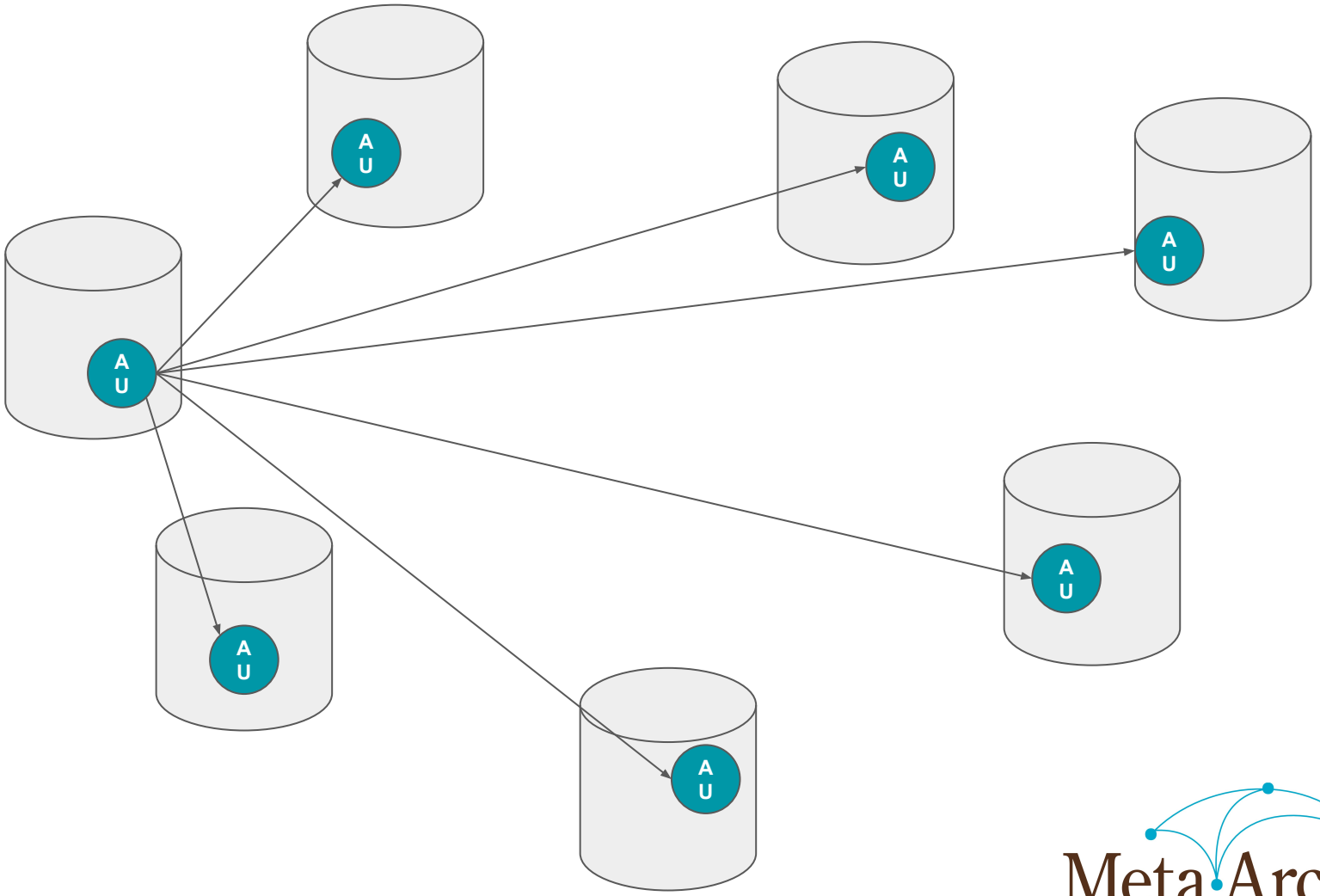
Voting and Polling

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



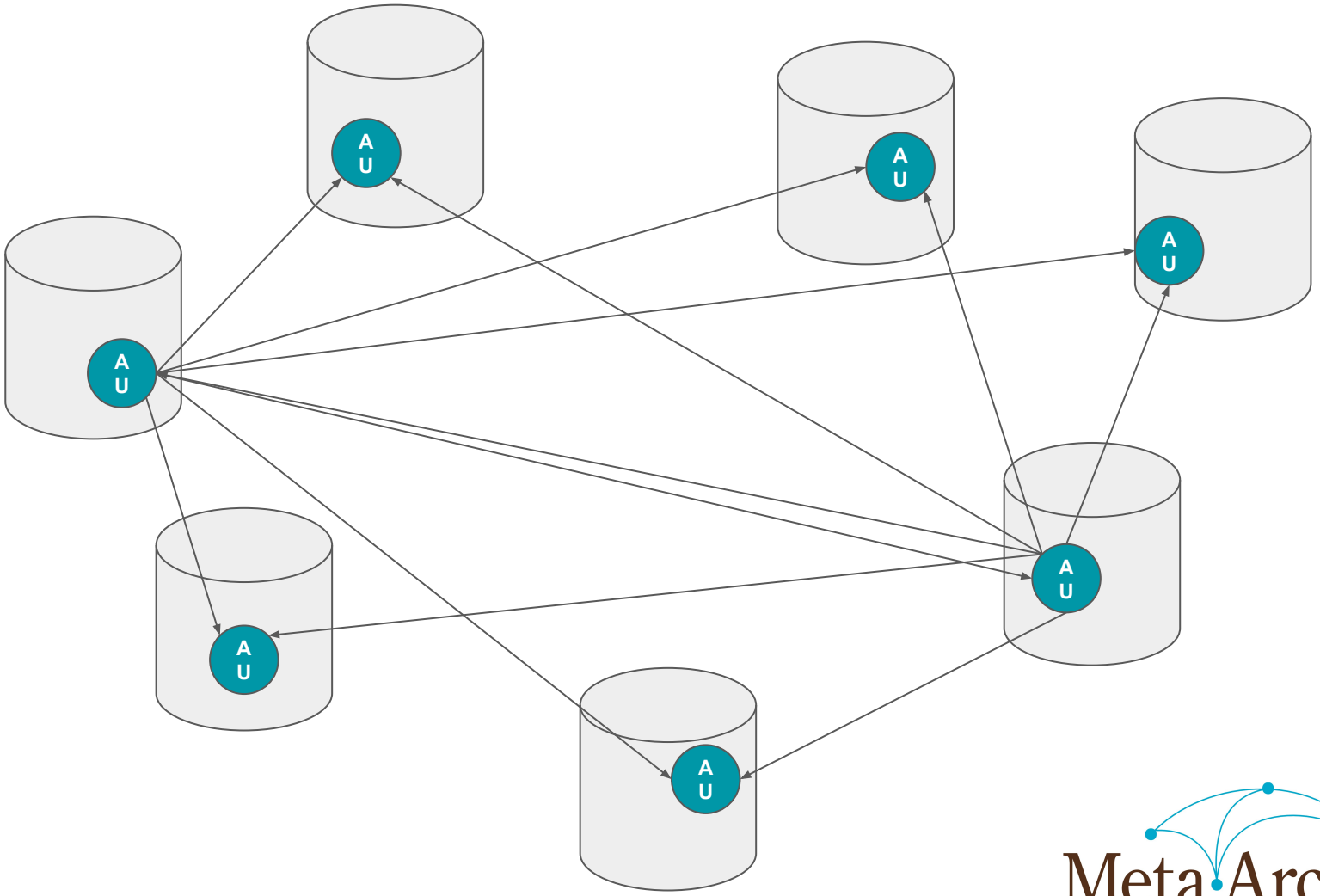
Meta Archive
COOPERATIVE

Voting and Polling



Meta Archive
COOPERATIVE

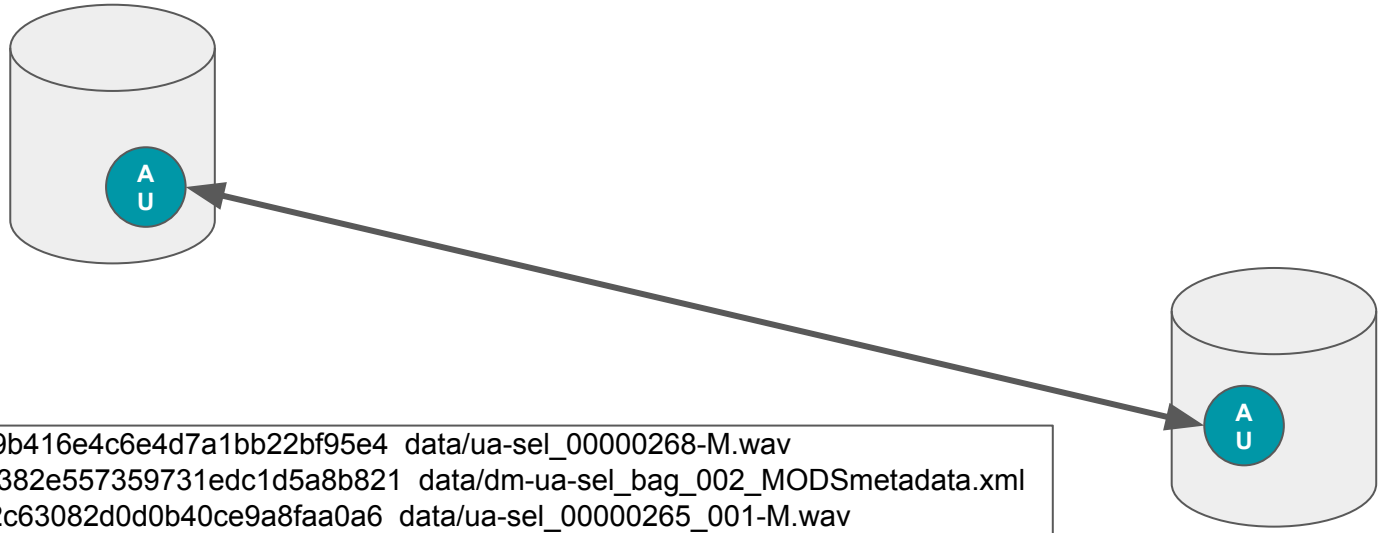
Voting and Polling



Meta Archive
COOPERATIVE

Damage and Repair

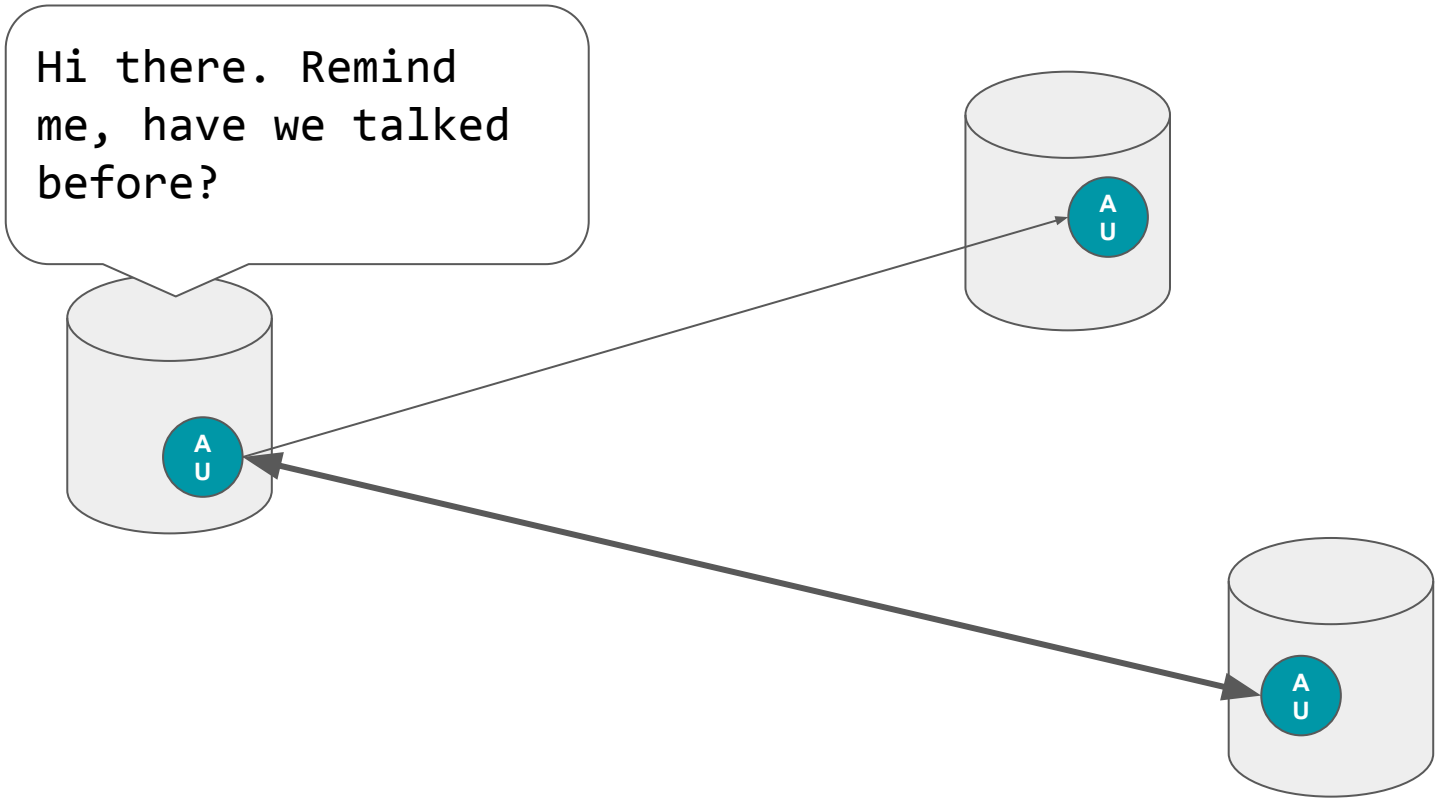
Cf2304e9b416e4c6e4d7a1bb22bf95e4 data/ua-sel_00000268-M.wav
046763d382e557359731edc1d5a8b821 data/dm-ua-sel_bag_002_MODSmetadata.xml
cf9beab2c63082d0d0b40ce9a8faa0a6 data/ua-sel_00000265_001-M.wav
733298738956be7ff4d9ed6b5d021e56 data/ua-sel_00000259-M.wav



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

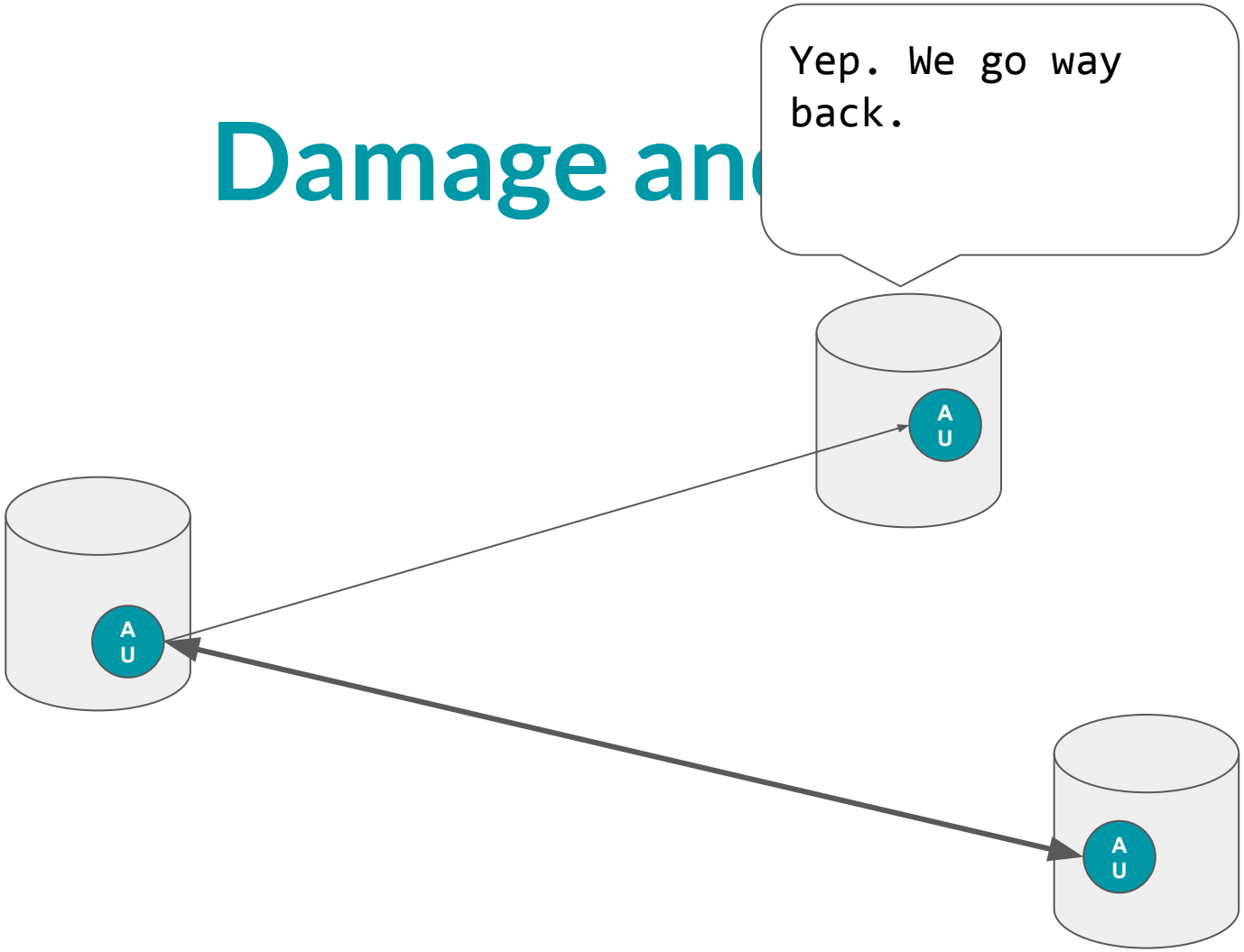
Meta Archive
COOPERATIVE

Damage and Repair



Meta Archive COOPERATIVE

Damage and



Meta Archive COOPERATIVE

Thanks!

metaarchive.org

sam@educopia.org

@samalanmeister

Meta Archive COOPERATIVE

Getting Started

November 2010
Attended 5-day workshop
"Digital Preservation Management"
University of Michigan

August 2011
Compared Digital Preservation
Repository options

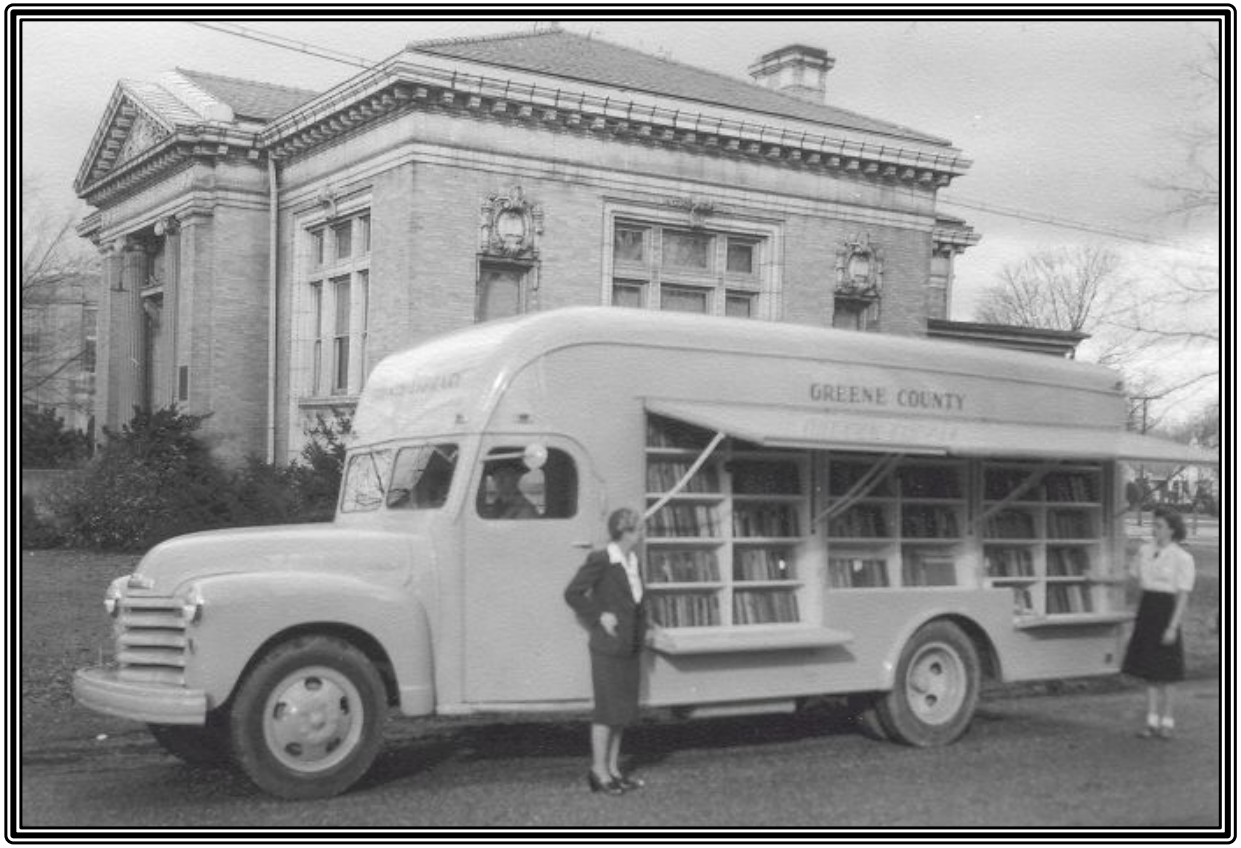
April 2012
Joined MetaArchive as a
Preservation Member

January 2013
Started ingesting collections

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

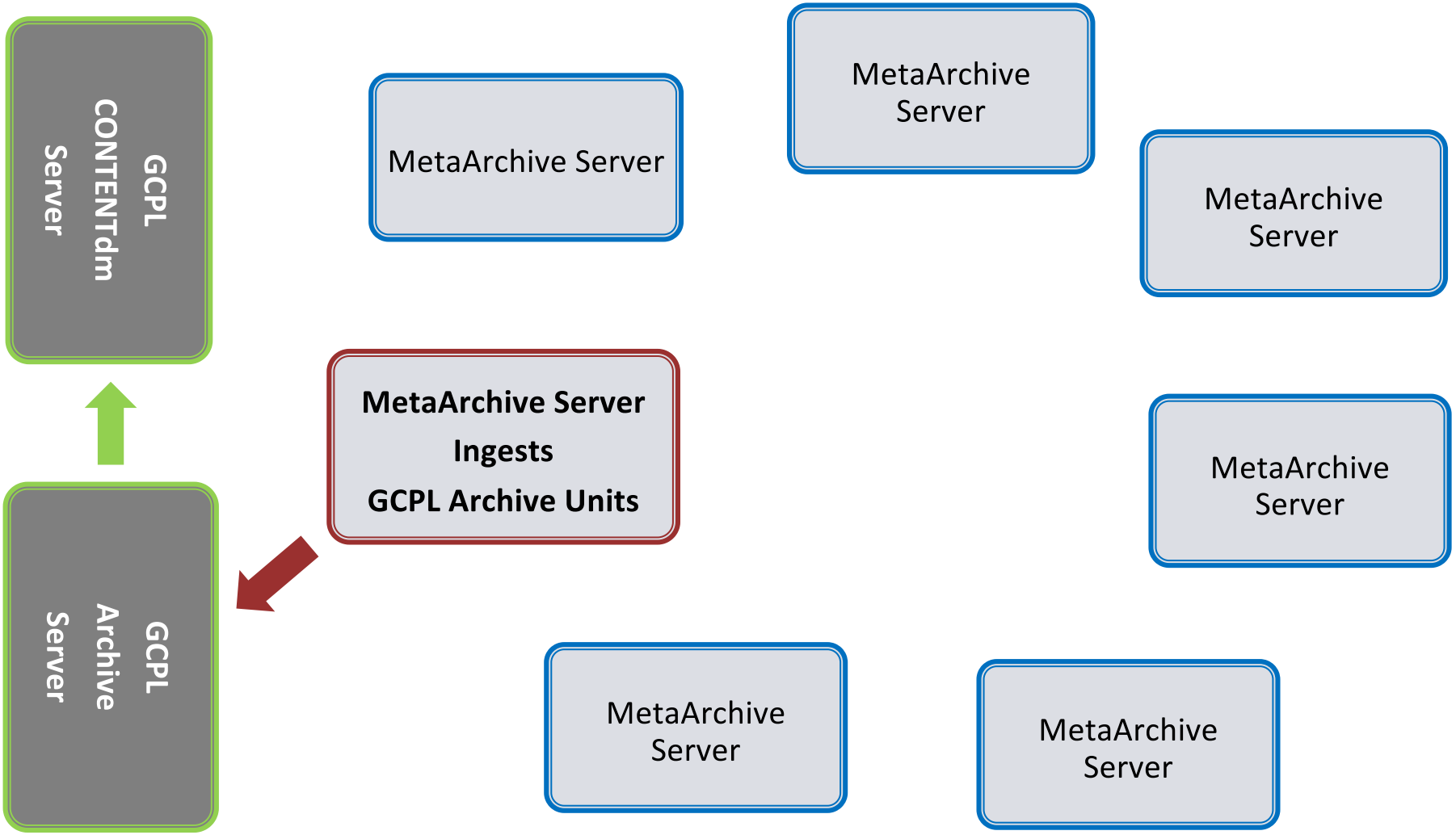
Why MetaArchive

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

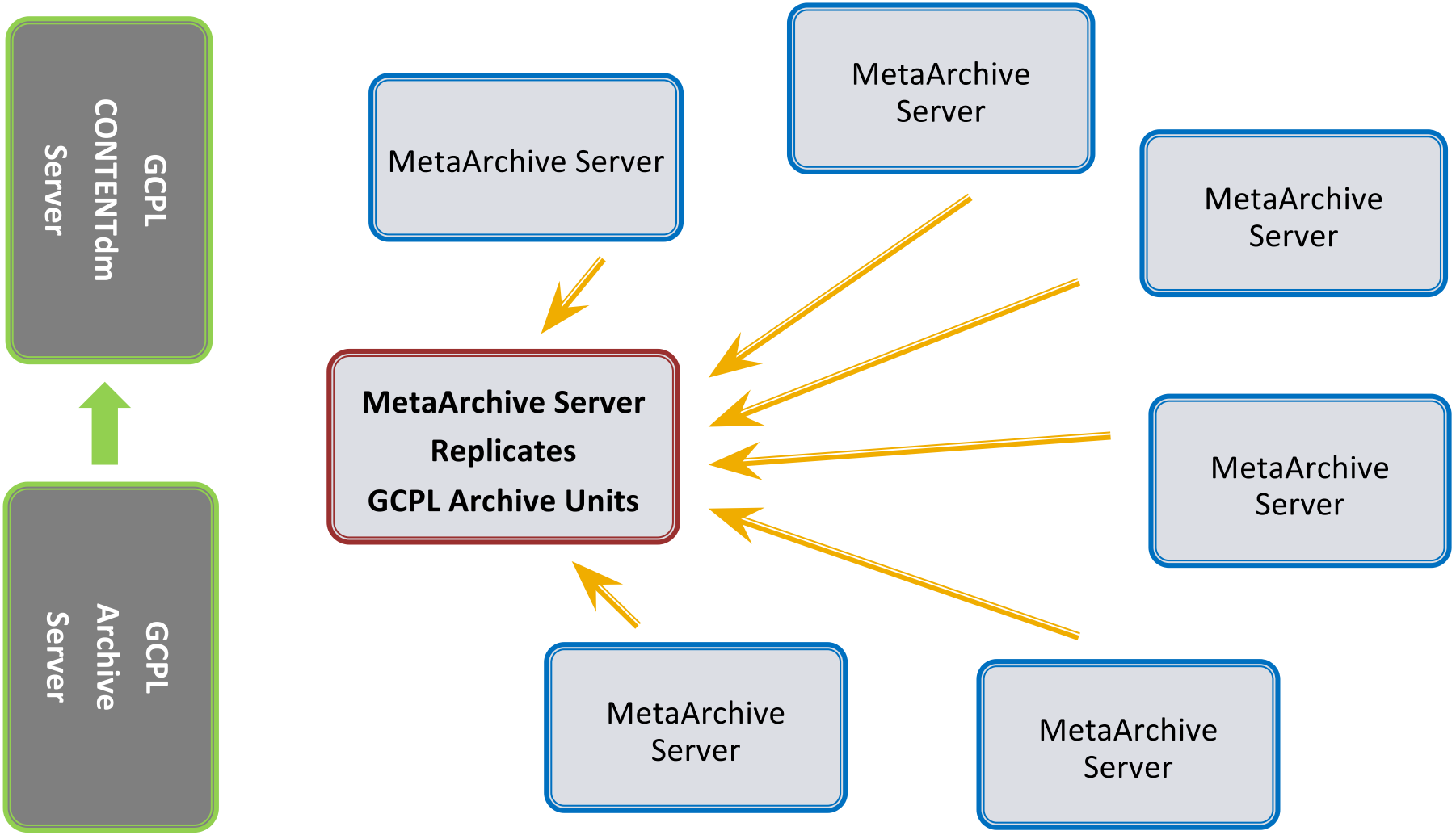


First bookmobile used by the Greene County Public Library from 1948 – 1958. Xenia, Ohio.

Modified Ingest



Modified Ingest



Cost of a Digital Time Capsule….

Library Paid in 2015

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

Preservation Membership	**$3,000**	
Technology Fee	**1,000**	
Storage .50¢ per GB x 3,600 GB		**1,800**
Total MetaArchive Fees 2015	**$4,800**	