MetaArchive Cooperative and the BitCurator Consortium

Long-Term Preservation Strategies & Architecture: Views from Implementers
Sam Meister
Katherine Skinner

iPRES 2015
University of North Carolina Chapel Hill
November 4, 2015
The MetaArchive Cooperative
MetaArchive History

2004

2015
50+ members, 14 US states/districts, 3 continents
Hallmarks

- Distributed digital preservation
- Institutions maintain control over their own content
- Preservation as a process, not a push-button exercise
- Simplicity in ingest, management
Three membership levels
- Collaborative members: $2.5K/year
- Preservation members: $3K/year
- Sustaining members: $5.5K/year

Server cost: <$5K/term

Storage cost: $1K/TB/year*
  - (UNDER REVIEW—will drop in 2015)
Basic processes

- “Producer” (OAIS) determines curation practices; brings SIPs to MetaArchive (low entry barriers)
- Multiple copies of AIPs dispersed across geographical, political, and environmental lines
- Checks and repairs automated across network
- Supports automated versioning
- Deaccession cycle versus data deletion
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
Balance of Research and Action

- Chronicles (NEH, 2012-14; NEH 2014-15)
- Lifecycle Management for ETDs (IMLS, 2012-14)
- DDP Frameworks (2014-?)
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
Collection Details

<table>
<thead>
<tr>
<th>Title</th>
<th>Virginia Tech VA-news</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive</td>
<td>GEN</td>
</tr>
<tr>
<td>Plugin</td>
<td>org.metaarchive.AllFromStart2 by Metaarchive</td>
</tr>
<tr>
<td>Base URL</td>
<td><a href="http://scholar.lib.vt.edu/Chronicles_Ingest">http://scholar.lib.vt.edu/Chronicles_Ingest</a></td>
</tr>
<tr>
<td>Identifier</td>
<td>238</td>
</tr>
<tr>
<td>Description</td>
<td>The Virginia news collection at Virginia Tech. The collection includes the Roanoke Times from January 1992 to February 1997, the Virginian Pilot from 1994 to 1997, and WDBJ7 news transcripts from October 1995 to August 2008</td>
</tr>
<tr>
<td>Repository Type</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

Related Items

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Plugin XML</th>
<th>Archival Units</th>
</tr>
</thead>
</table>

This collection has no metadata.
Adding Collections
Plug-In pathway

New Plugin

Choose one of the plugin creation options below.

Start from a Sample

Plugin Identifier:

eg. AllFromStart

The plugin identifier as seen by LOCKSS. The prefix shown will be prepended automatically, if applicable.

Plugin Template

- All from Start
  
  Crawls everything from a given start path, blah blah blah blah.

Content Provider

The plugin will belong to the selected content provider.

Continue

Upload a Custom Plugin

- or -

Plugin XML Upload

Browse...

No file selected.

Upload your LOCKSS plugin XML file here. The plugin identifier will be extracted from the uploaded file.

Content Provider

The plugin will belong to the selected content provider.

Continue
MetaArchive Bag Creation

Export copies of their digital collections/files, Stage to a server, Ingest into MA
# Auditing Control

## Archival Unit Details

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection</td>
<td>Virginia Tech VA-news</td>
</tr>
<tr>
<td>Plugin</td>
<td>org.metaarchive.AllFromStart2 by Metaarchive</td>
</tr>
<tr>
<td>Status</td>
<td>Test</td>
</tr>
<tr>
<td>Up/Down</td>
<td>Site Up</td>
</tr>
<tr>
<td>Parameter Values</td>
<td></td>
</tr>
<tr>
<td>base_url</td>
<td><a href="http://scholar.lib.vt.edu/Chronicles_ingest">http://scholar.lib.vt.edu/Chronicles_ingest</a></td>
</tr>
<tr>
<td>start</td>
<td>VA-news_bag</td>
</tr>
</tbody>
</table>

## Preservation Status

Coming soon.

## Auditing Tools

Use the form below to request an audit on this archival unit.

When the process completes, a new Report will be created.

**Audit Type**

<table>
<thead>
<tr>
<th>Value</th>
<th></th>
</tr>
</thead>
</table>

The type of audit.

**Cache**

<table>
<thead>
<tr>
<th>Value</th>
<th></th>
</tr>
</thead>
</table>

The cache to use.

![MetaArchive Cooperative Logo](image)
Join Us!

www.metaarchive.org
Members
Penn State University
Duke University
British Library
Stanford University
Yale University
New York Public Library
University of Virginia
The University of Maryland, MITH
Harvard University
University of Colorado Boulder
McMaster University
Massachusetts Institute of Technology
Texas State Libraries and Archives Commission
New York University
Princeton University
The University of Manchester Library
Northwestern University
Gates Archive
Emory University
The University of Maryland, Libraries
University of North Carolina Chapel Hill, SILS
Archives and Forensics

Digital forensics offers archivists new tools and methodologies to:

- survey the extent of a collection
- weed objects that do not fall under collecting policies
- accession the contents of a collection
- preserve the original order of a collection
BitCurator helps archives to...

- Acquire
  - Capture materials at lowest possible level
  - Maintain original order of files on media
  - Identify problems at scale: SEI, PII, viruses
  - Generate lists of contents, file extensions
  - Extract file system level metadata
  - Create copies (access, preservation)
BitCurator helps archives to...

• Appraise
  – Mount content from obsolete media, systems
  – View contents
  – De-duplicate (fuzzy hashing)
BitCurator helps archives to...

• Arrange and Describe
  – Abstract groups of digital files for description
  – Supports MPLP-like arrangement
  – Export contents for more complex arrangements
  – Granular file system metadata for creation, modified dates
BitCurator helps archives to...

- **Access**
  - Mounting disk images in the reading room
  - Document creator’s computer environment
  - Experiment with browsing disk image via browser
  - Facilitates experiments with emulation
What is the Consortium?

The BitCurator Consortium supports digital forensics practices in libraries, archives, and museums in order to help ensure the longevity and reliability of the cultural, scientific, and historical record.
independent

community-led

membership association
Member Benefits

• Use of the members-only BCC mailing list and
• Access Help desk
• Access to the members-only videos and documentation
• Prioritized requests for BitCurator feature development
• Opportunities to serve on BCC committees
• Voting rights for community governance
• Professional development opportunities
• Discounts for events (including the BitCurator User Forum)
BCC Start-up Trajectory

Exploratory conversations between UNC-CH, BitCurator project team, Educopia

**July 2014**
- Kicked off; established subcommittees; began outreach

**Jan-June 2014**
- 14 members; 35+ attend inaugural BC Users Forum; Professional development webinar presented

**Jun 2015**
- Elected first Executive Council; renewed 18 members; added 1 new member; webinar given, 2 meet-ups hosted

**Fall 2015**
- Website launched; help desk designed; workflows project launched; additional meet-ups hosted
Join us!