Evolving and Adapting: The MetaArchive SuperNode Pilot Project

Kyna Herzinger
Sam Meister

Best Practices Exchange 2019
Columbus, OH
April 29, 2019
Digital preservation network created and hosted by and for memory organizations
Independent membership organization
MetaArchive History

2004 - Founded as part of NDIIPP funding and activities

2006 - Educopia Institute founded to serve as administrative home for MetaArchive

2007 - MetaArchive becomes initial Affiliated Community

2017 - Celebrated 10 years as membership-based preservation network
Hallmarks

Community-based
Affordable
Transparent
Supportive
Diverse
Background / Motivations
Challenges for Small Organizations
Limited local IT resources
Internet bandwidth constraints
Shift to cloud-based services
Challenges for MetaArchive
Supporting complex ingests
Ensuring storage balances
Solution(s)
Simplify ingest process

Lower barriers to participate
Evolution
1. Embrace Bags
2. Offer “Tech Fee” Alternative
3. Explore SuperNode Network
SuperNode Network
SuperNode Network

Staging Server

Staging Server
SuperNode Pilot Project

Purpose

Conduct additional testing and analysis to determine the feasibility and requirements for a SuperNode Network infrastructure.
SuperNode Pilot Project

Objectives

Ingest a lot of content

Collect and analyze cost data

Test and evaluate Transfer Tools

Refine SuperNode resource requirements
SuperNode Pilot Project

Phase 1: Install, Test, Train

SuperNoders and Ingesters install, setup, and configure software and hardware elements
SuperNode Pilot Project

Transfer Tools

DRAG  DROP

FILE SYNC

Staging Server

Member Storage Node
SuperNode Pilot Project

Transfer Tools
SuperNode Pilot Project

Staging Server 1

Boston College Data Center

Virtual Server

ownCloud
SuperNode Pilot Project

Staging Server 2

Virtual Server (EC2 + EBS Volume)

SFTP (Exactly)
SuperNode Pilot Project

Phase 2: Ingest, Ingest, Ingest

Ingesters prepare content and use transfer tools to transfer content to SuperNode staging servers

Track transfer duration and document issues
SuperNode Pilot Project

Phase 2: Ingest, Ingest, Ingest
Ingesters

HBCU Library Alliance
Isabella Stewart Gardner Museum
InDiPres (Butler + Indy Public Library)
University of Louisville
SuperNode Pilot Project

Phase 3: Analyze, Evaluate, Recommend

Project manager and team analyze and summarize data on ingest and storage costs

Get feedback from SuperNoders and Ingesters

Produce Pilot Project report and Implementation Roadmap
SuperNode Pilot Project

Findings

Transfer tools meet basic requirements

Single cloud-based staging server not a viable option

Minimal IT resources required to host and maintain staging server
SuperNode Pilot Project

Next Steps

Refine Recommendations and Implementation Roadmap
institutional profile

limited financial resources
our modestly-resourced institution has been navigating financial issues that, combined with some governance problems, jeopardized our accreditation

finite human resources
we do not have an archivist who is dedicated to handling born-digital collections; these duties are split between four staff members, none of whom have formal training in digital preservation

competing priorities
as little as two years ago we had no policies or procedures in place for our born-digital collections and since then we have developed only basic practices
digital surrogates for materials that are in good condition

born-digital that are held elsewhere; digital surrogates for large, complete projects

born-digital & materials that would be costly to re-digitize
acquire
Bagger/Exactly
quarantine + virus scan
Dark computer station + Endpoint
create accession record
ArchivesSpace
move to network storage
Fixity
extract metadata
DRIOD
process
Normalize files and file names, de-dup, arrange, gather collection data, etc.
copy to preservation storage
Preservation server & MetaArchive
**Contents**

MetaArchive
SuperNode Pilot Project
OwnCloud + Bagger Configuration

**Steps**

**Bagger**

If you haven’t already, download and install Bagger via the link:
https://github.com/LibraryofCongress/bagger/releases/tag/v2.5.1

1. Download “MetaArchive-profile.json” from Google Drive to your local computer.
2. Next you are going to move the “MetaArchive-profile.json” file to a Bag folder.
3. If you are using a Mac, from the Finder navigation menu select
   a. In the search bar enter “~/bagger” (see below)

**Transfer**

Bagger + OwnCloud

1. Open Bagger
2. Create New Bag
3. Select MetaArchive from profile dropdown
4. In the Bag Info section enter values for document information
5. Add data
   a. Select File + Add data or the
   b. Navigate to the folder where the bag
   c. Select Open
6. Save Bag
   a. Select File + Save Bag As or the
   b. In the “Save Bag” dialog box, set
   c. If you want the bag to be saved
   i. Note: The folder location
   j. If you would like to save the
   k. Add the total data size for the bag
   l. Move Bag to local OwnCloud folder
   m. If you have not directly saved
   n. Record the transfer start date

**Metadata**

Note that the file names/identifiers “start” values for the archived units you

**Documentation**

Download your institution-specific Exactly configuration file and SSH private key from
your Google Drive folder to your local computer.
1. The configuration file will look like “Exactly_Configuration_user.xml” with “user” being
   your assigned username.
2. The SSH private key will look like “harvest_madkey_user” with “user” being
   your assigned username.

3. Open Exactly
4. Select the “Exactly” navigation header on the dropdown menu, select “Import”
5. In the file manager window, navigate to the folder where you saved the Exactly
   configuration file.

6. Select Open
7. Check to confirm that the configuration file was successfully imported.
Error: Verifying Transfer...
(1/1)Transferring files...
Something went wrong while copying files again trying to transfer files...
(1/1)Transferring files...
Something went wrong while copying files again trying to transfer files...
(1/1)Transferring files...
2009 either got corrupted or not copied to the destination
Please try again.

Validating Bag Info
Metadata saved successfully.
Folder already existed in destination with this title. Please change the title.

(1/1)Transferring files...
Something went wrong while copying files again trying to transfer files...

Bag successfully saved; Bag successfully validated
<table>
<thead>
<tr>
<th>date</th>
<th>collection</th>
<th>bag_identifier</th>
<th>total_data_size (GB)</th>
<th>transfer_tool</th>
<th>transfer_start_datetime</th>
<th>transfer_finish_datetime</th>
<th>bagger_transfer_duration</th>
<th>owncloud_transfer_duration</th>
<th>total_transfer_duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>12/12/2018</td>
<td>WAVE Radio transcription discs</td>
<td>1573-443</td>
<td>3.7</td>
<td>Bagger + Owncloud</td>
<td>12/12/2018 12:47:00</td>
<td>12/12/2018 13:16:00</td>
<td>0:29:00</td>
<td>0:02:00</td>
<td>0:31:00</td>
</tr>
<tr>
<td>12/12/2018</td>
<td>August 2009 Flood Collection</td>
<td>2009-106</td>
<td>2.4</td>
<td>Bagger + Owncloud</td>
<td>12/12/2018 14:46:00</td>
<td>12/12/2018 14:58:00</td>
<td>0:13:00</td>
<td>0:02:00</td>
<td>0:15:00</td>
</tr>
<tr>
<td>12/12/2018</td>
<td>Office of the President. James R. Ramsey email</td>
<td>2017-010</td>
<td>19.7</td>
<td>Bagger + Owncloud</td>
<td>12/12/2018 10:06:00</td>
<td>12/12/2018 10:26:00</td>
<td>0:17:00</td>
<td>0:06:00</td>
<td>0:23:00</td>
</tr>
<tr>
<td>12/12/2018</td>
<td>Office of the President. Neville G. Pinto email</td>
<td>2017-040</td>
<td>20.0</td>
<td>Bagger + Owncloud</td>
<td>12/12/2018 10:43:00</td>
<td>12/12/2018 12:15:00</td>
<td>1:32:00</td>
<td>0:09:00</td>
<td>1:41:00</td>
</tr>
<tr>
<td>12/12/2018</td>
<td>Louisville Leader</td>
<td>1973-851</td>
<td>175.5</td>
<td>Bagger + Owncloud</td>
<td>12/12/2018 16:09:00</td>
<td>12/12/2018 17:12:00</td>
<td>1:03:00</td>
<td>0:48:00</td>
<td>1:51:00</td>
</tr>
</tbody>
</table>

*Ingest tracking*
## My Collections

### Active Collections

<table>
<thead>
<tr>
<th>Archive</th>
<th>Title</th>
<th>#AU</th>
<th>Content Provider</th>
<th>Testing AU #</th>
<th>Preserved AU #</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>Newton Owen Postcard Collection (retired)</td>
<td>12</td>
<td>University of Louisville</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SDC</td>
<td>The Herald-Post Collection (retired)</td>
<td>2</td>
<td>University of Louisville</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IH</td>
<td>University of Louisville Yearbooks (retired)</td>
<td>53</td>
<td>University of Louisville</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SDC</td>
<td>Royal Photo Company Collection</td>
<td>0</td>
<td>University of Louisville</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>COMM</td>
<td>Ghost Signs of Louisville</td>
<td>0</td>
<td>University of Louisville</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SDC</td>
<td>Louisville Storefronts</td>
<td>0</td>
<td>University of Louisville</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GEN</td>
<td>Edwards Oral History Collection</td>
<td>1</td>
<td>University of Louisville</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SDC</td>
<td>Arthur Younger Ford (1961-1926) Photograph Albums</td>
<td>1</td>
<td>University of Louisville</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SDC</td>
<td>African American Oral History Collection</td>
<td>20</td>
<td>University of Louisville</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>SDC</td>
<td>Kate Matthews Collection</td>
<td>21</td>
<td>University of Louisville</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>SDC</td>
<td>Bernheim Foundation interviews - 2</td>
<td>53</td>
<td>University of Louisville</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td>Bagger + Owncloud</td>
<td>Exactly</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Need administrative permissions over workstation</td>
<td>• No need for administrative permissions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Metadata profile</td>
<td>• Metadata profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• sha256</td>
<td>• md5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Does not (but can) create a local copy</td>
<td>• Creates a local copy every time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Includes bag-info, manifest, and tag files</td>
<td>• Includes bag-info, manifest, and tag files PLUS file system data and transfer completed summary files</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• No sense of transfer progress</td>
<td>• No sense of progress or even initializing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Metadata disappeared and had to be re-entered</td>
<td>• Transfer error on one bag, reason unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Issues with Java Runtime Environment</td>
<td>• Noticeably slowed my workstation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Thank you!

www.metaarchive.org

@metaarchive

sam@educopia.org

@samalanmeister