TRADITIONAL PROCESSING MEETS ISLANDORA

Betsy Coles  Caltech Library
Elisa Piccio  Caltech Archives & Special Collections
Mariella Soprano  Caltech Archives & Special Collections
Why Islandora? Advantages

- Open source
- Fedora Commons back-end – “future-oriented”
- Drupal CMS front-end included
- Can be hosted or locally deployed
- Active open source development community; commercial support available
- Highly customizable
- Many “plug-in” modules available to add functionality
- Good support for preservation activities (checksums, preservation metadata, transfer to DPN)
Disadvantages

• Drupal CMS front end included
  • Requires Drupal expertise; new releases of Drupal are not compatible
• Requires significant level of technical support for local deployment
  • Software developer at 50% time for initial migration, 75% time for another year for later local customization activities
• Steep learning curve for both technical staff and archives staff
  • Technology stack (Java, Fedora, Solr, Drupal) requires broad technical expertise
  • Some parts of Islandora staff interface are less-than-intuitive:
    • Metadata entry forms in particular are problematic
    • Drupal interface “requires getting used to”
Initial Islandora Implementation

- Decision to go with Islandora for DAMS was made in late 2012
- Initially we used out-of-the-box Islandora, except for custom theming, custom metadata schema (full MODS), and metadata input forms
- Implementation began in 2013
  - Migration of a legacy database (the ImageArchives)
  - Export and transformation of legacy metadata done locally
  - Islandora implementation and data loading outsourced to discoverygarden.ca
The image archives

- A collection of over 10,000 images representing Caltech’s history, and the people who have made and continue to make it

- Digitization project started in 1993
- Migrated from FileMakePro database to Islandora in 2013
- Collection on OAC linked to Caltech server
Image Archives Demo
Integrating Traditional Archival Processing into Digitization Project

- In this talk we are addressing the digitization of non-digital collections
- Evolution, not revolution
- Attempt to take advantage of efficiencies in established processes
- Tweak them to create the best possible experience for users of digitized content
Paul B. MacCready (1925-2007)

- Caltech MS physics 1948, PhD aeronautical engineering 1952
- A visionary, inventor and entrepreneur, pioneered alternative energy solutions with his company AeroVironment
- Created solar-powered aircraft, solar-powered and electric cars, even a flying pterosaurus
- Designed human-powered aircraft
- First Kremer prize, 1977: Gossamer Condor flew one-mile figure eight, clearing ten-feet pole
- Second Kremer prize, 1979: Gossamer Albatross flew from England to France
Collection overview

- Donated to the Caltech Archives in 2003
- Processing completed in 2014
- Measures 57 linear feet, comprising 112 archival boxes
- Organized in 7 Series
Collection overview - Series

1 AeroVironment
2 Planners and Diaries
3 Notebooks
4 Writing and Talks
5 Biographical and Correspondence
6 Miscellaneous Materials
7 Audio-Visual
Collection overview

- The collection spans 1930 to 2002, documenting most aspects of MacCready's personality and career through a diverse array of documents, media, objects, manuscripts and printed materials.

- Especially prevalent are papers and ephemera from 1977 to 1985, when he was working on human-powered airplanes.

- The papers also document his work in alternative energy solutions.
## Materials and digitization

<table>
<thead>
<tr>
<th>Items</th>
<th>Description</th>
<th>Digitized by</th>
</tr>
</thead>
<tbody>
<tr>
<td>54,000 Papers</td>
<td>300ppi TIFF</td>
<td>In-House digitization by DocuServe – Access and Fulfillment Services at Caltech Library</td>
</tr>
<tr>
<td>2,000 Photos</td>
<td>600ppi TIFF</td>
<td></td>
</tr>
<tr>
<td>130 VHS</td>
<td>mp4</td>
<td>Digitized by USC Shoah Foundation</td>
</tr>
<tr>
<td>10 audiocassettes</td>
<td>wav</td>
<td></td>
</tr>
<tr>
<td>8 16mm reels</td>
<td>mov (uncompressed V210)</td>
<td>Digitized by the California Audio Visual Preservation Project (CAVPP)</td>
</tr>
<tr>
<td>5,600 Slides</td>
<td>600ppi TIFF</td>
<td>Digitized by John Sullivan, Imaging Services, The Huntington.</td>
</tr>
<tr>
<td>14 Oversize drawings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Artifacts</td>
<td></td>
<td>Caltech Graphic Resources Photographer</td>
</tr>
</tbody>
</table>
MacCready → Local Innovation

• Naming scheme for digitized files reflecting container list structure at folder and page level
• Navigation via finding aid: automated links from container list to digital objects in Islandora
• Implementation of a paging display that preserves context within folder objects
Innovation 1: From arrangement to filenames

PBM_7_23_5_0001.tif

Collection_Series_Box_Folder_File
Innovation 1: From arrangement to filenames

- Only Series, Box and Folder numbers are used, not Subseries
- Box numbering restarts from 1 in each Series, allowing digitization to begin before processing of Series was completed
- Files get a 4-digit suffix: PBM_4_2_1_0023.tif
- Descriptive metadata is drawn from finding aid at folder level, and metadata files are numbered the same way as digital object files.
Automated metadata generation from Finding Aid

- Folder level information created as part of traditional processing
- We can use this information to automatically generate MODS metadata for Islandora, at the folder level.
- Start with container list in EAD form of finding aid
- Transform with various tools (OpenRefine, XSLT, perl scripts) to produce DLF/Aquifer compliant MODS/XML files, one per folder
- Key for later ingest: MODS files are named using Series/Box/Folder convention, e.g. PBM_7_23_5.xml
MODS/XML example

```xml
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/v3" xmlns:mods="http://www.loc.gov/mods/v3"
xmlns:xlink="http://www.w3.org/1999/xlink">
  <titleInfo>
  </titleInfo>
  <typeOfResource>moving image</typeOfResource>
  <originInfo>
    <dateIssued keyDate="yes">1991 June</dateIssued>
  </originInfo>
  <language>
    <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
  </language>
  <identifier type="local">PBM_7_23_5</identifier>
  <physicalDescription>
    <form authority="marcform">videorecording</form>
    <extent>VHS. 8 min. 32 sec.</extent>
    <digitalOrigin>digitized other analog</digitalOrigin>
  </physicalDescription>
  etc. ....
</mods>
```
Automated ingest of metadata and digital objects into Islandora

- Islandora has batch ingest capabilities
- Congruity of file names for digital objects and metadata files allows creation of scripts that match them up and feed them to Islandora together.
Innovation 2: Automated Linking From Finding Aid

- We started with UCLA’s work on the Islandora Manuscript Solution Pack
- EAD Finding Aid is loaded into Islandora to provide Collection Guide navigation
- We create links on-the-fly from the EAD container list to objects in the collection
MacCready Collection Demo
Innovation 3: IIIF and the UniversalViewer

- A community driven image framework with well defined APIs for making the world's image repositories interoperable and accessible
- UniversalViewer: Open source project, backed by British Library, implementing IIIF
UniversalViewer Demo
What Have We Accomplished?

- Retained advantages of traditional processing workflow
- Gained efficiencies in digitization and ingestion workflow
- Improved user experience
  - Navigation via finding aid
  - Display (once UniversalViewer is implemented)
Future Directions

Donald A Glaser Collection - Nobel Prize winner in Physics (underway)

Materials from various already-processed collections, as an ongoing effort
Acknowledgements

• MacCready family
• Caltech Development & Institute Relations
• Caltech Library DocuServe
• USC Shoah Foundation
• John Sullivan, Imaging Services, The Huntington
• California Audiovisual Preservation Project (CAVPP)
• Jim Staub, Caltech Graphic Resources
• Kristen Abraham and Bianca Rios
Contacts

• Betsy Coles, Library Services
  • bcoles@caltech.edu

• Elisa Piccio, Archives & Special Collections
  • epiccio@caltech.edu

• Maria Soprano, Archives & Special Collections
  • mariella@caltech.edu