How to Succeed in Metadata Without Really Trying

SAA Annual Conference
Washington, D.C.
Digital Objects Roundtable
August 11, 2010
Karen Weiss weissk@si.edu
Two metadata creation and digitization workflows

- **Individual items**: Staff removing documents from collections use an online removal tracking system to create basic metadata for the digital files.

- **Large-scale digitization of entire collections**: Processing archivists’ EAD finding aid creates metadata for the digital files.

The Archives of American Art, an archival unit of the Smithsonian Institution, and the world’s largest repository of primary source materials documenting art in America, has developed *Collections Online*, a working model for the large scale digitization and online presentation of entire manuscript collections.

*Collections Online* provides online access to the Archives’ collections that have been digitized in their entirety, primarily the papers of artists, art historians and other art world figures. These are not selected highlights or digital images presented according to themes, but rather entire manuscript collections. The online presence amassed in less than four years of 67 fully digitized collections representing over 450 linear feet and amounting to nearly 600,000 digital files, with more added on a regular basis, underscore that large scale digitization can be achieved, notably by building upon fundamental archival approaches.
Individual Document Scanning

- Began in 2002
- Supports requests from patrons and staff
- Links to collection record
- Finding aid not required
- 10,732 documents scanned
- 27,834 images
Item Level Metadata Creation

- AAA’s registrar maintains “Removal Notice” file to track items removed from collections
- Majority of removals involve digitization
- Paper based system replaced with online form in 2008
- Now serves as both the registrar’s tracking system and digitization metadata creation tool
Online Removal / Metadata Creation Process

- Step 1: Select Purpose for the Removal
  eg. Exhibit, Outgoing Loan, Research Request etc.
Step 2: Select a Collection

Collection Records Import from MARC catalog

Collections: No collection selected

Start | Removal Item | Patron | Vendor | Condition | Digitization | Finish

To select a collection for this removal, begin typing its name in the search box below, choose the matching collection, then click 'Choose Collection'.

- Weidner, Roswell T. / Roswell Weidner papers, 1930-1980
- Wehr, Erika / Erika Wehr papers, 1946-1994
- Welts, Elma / Elma Welts papers, [ca. 1924-1982]
- Wein, Abigail G. / Abigail G. Wein research material, 1933-1942
- Weisberg, Helen J. / Helen J. Weisberg papers, 1908-1983
- Wiercinski, Carl K./Carl Wiercinski passport, 1851-1892
- Wiersma, Adolph A. / Adolph A. Wiersma papers, 1890-1899

Choose Collection
Step 3: Confirm if item already scanned

- If items already exist from the collection, a list is provided
- New item link
Step 4: Descriptive, Rights
Metadata creation

- General Format sets the stage for subsequent metadata
Step 4, cont’d Requester completes item form

- Numerous drop downs
- Online help
- Mandatory fields
- Select name from authority files or add new names
- Restriction and Rights information
Step 5: Digitization instructions

- Instructions for Digital Imaging Technician
- Due Dates
Step 6: Finalizing

- Data about the requester, date, default date to be returned
- When done, paper copy printed, accompanies original document into scanning lab
Adolph Weinman at the Art School of the Art Students League, New York City, ca. 1888-1897

Title: [Adolph Weinman at the Art School of the Art Students League, New York City].

Date: ca. 1888-1897

Physical Details: Photographic print ; 1 item ; b&w ; 19 x 23 cm.


Forms part of: Adolph A. Weinman papers, 1890-1959


Copyright: MMA Copyright Policy

Digital ID: 52397
Online Presentation
Large-Scale Digitization

- Began November 2005
- 97 collections / 695 linear feet digitized
- 958,275 images online; 203,475 under review
- 1,151,750 digital images total
Large Scale - Basic Approach

- Access to the digitized documents is through folder level access instead of item level access.
- All descriptive metadata derived from the EAD finding aid
- Container and folder data forms the file structure for the scanning technician to save the digital files, as well as the descriptive metadata for their discovery
- Ingest of data into database allows for easy way to associate images with metadata, and subsequent “re-packaging” for Digital Asset and other systems

The system integrates the descriptive metadata found in EAD (Encoded Archival Description) finding aids with corresponding digital content at the full collection level – not just links to selected highlights. The structured XML (Extensible Mark-Up Language) data of EAD finding aids provides the only descriptive metadata for collections scanned in their entirety, and serves as a contextual format for online navigation and access. The underlying database incorporates powerful programming and web interfaces that support and integrate the complex internal archival and scanning workflows necessary for sustaining a large scale institutional digitization program. This result is unprecedented access to the Archives’ digital collections and finding aids via the World Wide Web.
So, how exactly do we get this done?

First, the processing archivist creates an EAD formatted xml file. Our particular methodology for this is a modified Cookbook format, using a NoteTab text editor – fairly easy and fairly cheap. However, other EAD creation tools could certainly be used with the same results. The EAD finding aid looks just like any other EAD finding aid, with component levels and container numbers for box and folder.

These box and folder numbers play a major role in the database tables. The container list automatically creates the file directory for the collection on the scanning technician’s computer, where the digital files will be saved.

Because the processing archivist also physically marks each folder with a box and folder number (for example 1.1; 1.2; etc.), the scanning technician is less likely to confuse folders or lose her place while scanning a large collection.
Digitized Collection Presentation

Alexander Calder Papers, 1926-1967

The papers of Alexander Calder were digitized in 2009 by the Archives of American Art. The papers have been inventoried in their entirety, and total 1,085 images.

The papers of abstract kinetic artist and sculptor Alexander Calder measure 2.5 linear feet and date from 1926 to 1967. Found within the papers are scattered biographical materials, correspondence, printed materials, photographs, scattered prints and sketches by Calder, and a scrapbook. Of particular interest are the numerous photographs of Calder, including many of Calder at work in his studio, with his family at their home in Touraine, France, exhibitions, and artworks. Among the photographs are several taken by photographer and artist Herbert Matter and a photograph of Pierre Matisse at Calder’s home. Read more About this Collection.

Funding for the processing and digitization of this collection was provided by the Terra Foundation for American Art.

Series 1: Biographical Material, 1926-1954, undated

Series 2: Correspondence, 1926-1962, undated

Series 3: Printed Material, 1927-1967, undated

Series 4: Photographs, 1927-1966, undated

Alexander Calder at his studio, ca. 1935

Alexander Calder cutting metal, ca. 1955
Biographical Series

Alexander Calder Papers, 1926-1967

Series 1: Biographical Material, 1928-1954, undated
(Box 1; 8 folders)

Biographical Material includes various address lists and business cards kept by Calder, his passport, notes, a catalog with handwritten prices, and other writings. Also found are a French tax document and other ephemera.

<table>
<thead>
<tr>
<th>Box</th>
<th>Folder</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Address Lists, undated</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Annotated Catalog with Photos, 1928</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Business Cards, undated</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Fisherman, 1928-1941, undated (Partially scanned)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>Notes, undated</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>Tax Document, 1931</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>U.S. Passport, 1928</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>Writings by Calder and Others, 1954, undated</td>
</tr>
</tbody>
</table>

Return to top
The transition was surprisingly smooth, and attests to the compatibility between microfilming and digitization in many areas, including AAA’s organizational structure, staffing, and policies and procedures. This is especially evident with the processing archivists, where much of the digitization workflow is fully integrated into the archival processing workflow even beyond the creation and uploading of the EAD finding aid and includes preparation for scanning and final review in much the same way as microfilm procedures.
Workflows

- Workflow system creates a “checklist” to show the current status of a finding aid
- Users can do, undo, and redo actions
- Records who did what and when
- Checklist items can trigger events, like deployment
- Checklist items can also trigger notifications, e.g. letting a supervisor know that a finding aid is available for review
While the online public interface and access to collections was the primary goal of Collections Online, an important component from the start was the development of an internal web-based interface to simplify the numerous, complex workflows inherent in any large scale processing and digitization project. Little by little, the online workflow and statistical reporting from the database is replacing some of the dependency on spreadsheet management.
Automated Workflows

- Initial uploading of the archivists’ XML EAD finding aid into finding aid repository; subsequent check-in and check-out
- Creation of an internal Collection Online “stub” site for the collection
- Generation of the directory structure (based on the EAD XML container listing) for the digital imaging technician to use when scanning
- Batch processing of the TIFF files to create three jpg derivatives and watermarking
- Linking a representative image to each series description for a visually appealing presentation
- Linking of the digital files to the appropriate folder titles
- Generation of a pdf finding aid accessible from the collection site
- Deployment to the final presentation on the web
- Versioning support

Automated workflows available to staff encompass a range of processes, including initial uploading of the archivists’ XML EAD draft finding aid into the Archives’ finding aid repository, creation of an internal Collection Online “stub” site for the collection; generation of the directory structure (based on the EAD XML container listing) for the digital imaging technician to use when scanning; batch processing of the TIFF files to create three jpg derivatives and watermarking; linking a representative image to each series description for a visually appealing presentation; linking of the digital files to the appropriate folder titles; generation of a pdf finding aid accessible from the collection site; and finally, deployment to the final presentation on the web. Other innovative features include a check-in/check-out feature for finding aid updates and edits when necessary; versioning support; and image count calculations and other statistical reporting features.
Re-purpose existing archival methodologies and cataloging data

- Much easier to re-purpose and enhance existing workflows, rather than invent new ones
- Integrate digitization workflows into archival workflows
- Include web presentation as part of work

We realized that we needed to integrate digitization as much as possible with existing archival processing methodologies and approaches, and to build upon our earlier experiences with reformatting, and more recently, the creation of standardized data in EAD finding aids.