
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

AAA DSD Admin - Remove Item - Metadata Factory

Removing an Item

Collections: top collection selected

Start [Cancel]

Choose the purpose for this removal

Exhibit

Loan

Web/Digital Project

AAA Journal

Research Request

Conservation

Archival/Preservation requirement

Internal request

Public relations request

Other

Next step: enter item information

Done

- 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

Collection: *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".						
<input type="text" value="weinman"/>						<input type="button" value="Choose Collection"/>
Weidner, Roswell T. / Oral history interview with Roswell Weidner, 1989 July 20-27						
Weidner, Roswell T. / Roswell Weidner papers, 1930-1989						
Weihs, Erika / Erika Weihs papers, 1949-1994						
Weil, Stephen E. / Documents relating to the Andy Warhol estate, 1991-1994						
Weill, Erna / Erna Weill papers, [ca. 1934-1983]						
Weimer, Aloysius G. / Aloysius G. Weimer research material, 1933-1942						
Weinberg, Helen Joy / Helen Joy Weinberg papers, 1908-1983						
Weinadel, Carl / Carl Weinadel passport, 1831-1832						
Weingarten, Robert, / Oral history interview with Robert Weingarten, SUPPRESSED RECORD, 20						
Weinman, Adolph A. (Adolph Ale / Adolph A. Weinman papers, 1890-1959						

Step 3: Confirm if item already scanned

Removing an item

Collection: Adolph A. Weisman papers, 1890-1959 [9338]

[Start](#) [Remove item](#) [Print](#) [View](#) [Condition](#) [Digitization](#) [Feedback](#)

Adolph A. Weisman papers, 1890-1959 [9338]
[Change collection](#)

Existing items (total 46 items)

- 18920 To do list
- 18931 Adolph Weisman at work in his studio
- 18921 Model for one of the centaurs in the Fountain of the Centaurs
- 18924 Model for one of the sea urchins in the Fountain of the Centaurs
- 18921 Model for a centaur engaged for the Fountain of the Centaurs, Series 1892, MO, MS504
- 18924 Fountain of the Centaurs with Earl Carter's Louisiana Purchase statue of the Populists
- 18923 Fountain of the Centaurs, Jefferson City, Missouri
- 18921 Fountain of the Centaurs, Jefferson City, Missouri
- 17691 Commission for Seated Lincoln Statue
- 16662 Adolph Weisman and a model

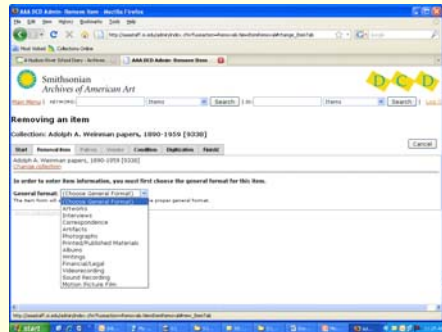
[New item](#)

- If items already exist from the collection, a list is provided

or

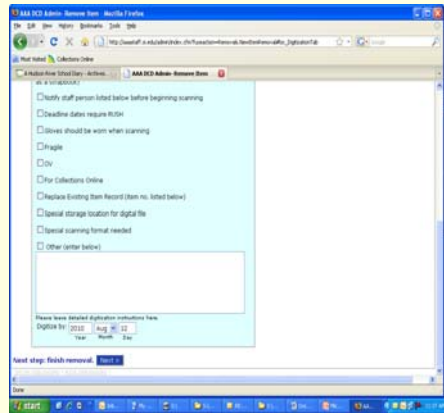
- New item link

Step 4: Descriptive, Rights Metadata creation



- General Format sets the stage for subsequent metadata

Step 5: Digitization instructions



The screenshot shows a web browser window displaying the 'AAMSD Admin Review Form'. The form contains several checkboxes for digitization instructions:

- Notify staff person listed below before beginning scanning
- Deadline dates require RUSH
- Gloves should be worn when scanning
- Fragile
- Dry
- For Collections Online
- Replace existing item record (item no. listed below)
- Special storage location for digital file
- Special scanning format needed
- Other (enter below)

Below the checkboxes is a text input field. At the bottom of the form, there is a section for 'Please enter detailed digitization instructions here' with a 'Digitize by' date field set to '2012 Aug 12'. The date field has sub-fields for 'Year', 'Month', and 'Day'. A 'Next step: Finish removal' button is visible at the bottom of the form.

- Instructions for Digital Imaging Technician
- Due Dates

Item Record Presentation

[Research Collections](#) > [Search Images](#) > [Photographs](#) >

Adolph Weinman at the Art School of the Art Students League, New York City, ca. 1888-1897



[View Larger Image](#)

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 ; 18 x 23 cm.

Creator: [Adolph A. \(Adolph Alexander\) Weinman](#), photographer

Subjects: [Adolph A. \(Adolph Alexander\) Weinman](#) / [Art Students League \(New York, N.Y.\)](#) / [Charles Keck](#)

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

[View Full Record](#) | [View Images for this Collection](#)

Citation: [Adolph Weinman at the Art School of the Art Students League, New York City], ca. 1888-1897 / Adolph A. (Adolph Alexander) Weinman, photographer. Photographic print : 1 item ; b&w ; 18 x 23 cm. Adolph A. Weinman papers, 1890-1959. Archives of American Art.

Copyright: [AAA Copyright Policy](#)

Digital ID: 5237

Online Presentation

Research Collections > Search Images >

Search Results

You searched for images matching "Adolph A. Weinman papers"

• There were 45 items matching "Adolph A. Weinman papers"

1/20

Choose View:
Gallery | List



Adolph Weinman at the Art School of the Art Students League, New York City, 1892-1897.

Interior view of the Art School of the Art Students League, New York City, 1892-1897.

Group of people in a hallway, Art School of the Art Students League, New York City, 1892-1897.

Document with handwritten text, Art School of the Art Students League, New York City, 1892-1897.

Framed picture, Art School of the Art Students League, New York City, 1892-1897.

Group of people in a room, Art School of the Art Students League, New York City, 1892-1897.

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

Sample XML

```
<c01 level="series">
<did><unitid>Series 1: </unitid><unittitle>Biographical Material, <unitdate>1928-1954, undated
</unitdate></unittitle>
<physdesc>(Box 1; 8 folders)</physdesc></did>
<scopecontent>
<p>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.</p>
</scopecontent>

<c02><did>
<container type="box">1</container>
<container type="folder">1</container>
<unittitle>Address Lists, <unitdate>undated</unitdate></unittitle>
<physdesc></physdesc>
</did></c02>

<c02><did>
<container type="box">1</container>
<container type="folder">2</container>
<unittitle>Annotated Catalog with Prices, <unitdate>1929</unitdate></unittitle>
<physdesc></physdesc>
</did></c02>
```

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

[View Collection](#)

[More about this Collection](#)

[About Collections Online](#)



Alexander Calder cutting metal, ca. 1955

The papers of **Alexander Calder** were digitized in 2005 by the Archives of American Art. The papers have been scanned in their entirety, and total 1,086 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 studios, with his family at their home in Touraine, France, exhibitions, and artwork. 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, 1928-1954, undated](#)



[Series 2: Correspondence, 1928-1962, undated](#)



[Series 3: Printed Material, 1927-1967, undated](#)



[Series 4: Photographs, 1927-1962, undated](#)



Biographical Series

Alexander Calder Papers, 1926-1967

[View Collection](#)

[More about this Collection](#)

[About Collections Online](#)

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.

Box	Folder	
1	1	Address Lists, undated
1	2	Annotated Catalog with Prices, 1929
1	3	Business Cards, undated
1	4	Ephemera, 1929-1943, undated (Partially scanned)
1	5	Notes, undated
1	6	Tax Document, 1931
1	7	U.S. Passport, 1928
1	8	Writings by Calder and Others, 1954, undated

[Return to top](#)

[PREVIOUS SERIES](#) | [NEXT SERIES](#)



From Series 1:
Biographical Material,
1928-1954, undated

Collection Title, Series Title, Folder Title compiled and presented for context



50+ Years of Microfilming Paved the Way for Metadata Model

**Microfilming with Zeutschel
microfilm camera 2002**



**Scanning with Zeutschel 1000AI
Scanner 2005**



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

Workflow web interface

AAA Collections Online Workflow - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://aaastaff.si.edu/Workflow/lead/controller/index.cfm?SearchField=Worthington+Y

Getting Started Latest Headlines

Worthington Whittredge papers

[Choose different collection](#)

Collection Progress Checklist Check-In (Upload) Check-Out (Download) Preview Image Processing

Collections Online Workflow Progress

Any tasks that have been ~~eressed-out~~ have been completed. The resource is available to the public once all tasks are completed.

1. ~~Archivist uploads new XML File to Finding Aid Repository~~ (Completed by Erin Corley on 2/4/09 10:44 AM [Redo](#))
2. ~~Supervisor approves XML Finding Aid~~ (Completed by Erin Corley on 2/4/09 10:45 AM [Undo](#))
3. ~~Cataloger provides index terms; archivist adds to XML File~~ (Completed by Erin Corley on 2/4/09 10:45 AM [Undo](#))
4. ~~Second supervisor approves XML Finding Aid~~ (Completed by Erin Corley on 2/4/09 10:45 AM [Undo](#))
5. ~~Create Image Directory / Begin Scanning~~ (Completed by Erin Corley on 2/4/09 10:45 AM [Redo](#) | [Undo](#))
6. ~~Finished Scanning~~ (Completed by Erin Corley on 2/4/09 10:46 AM [Undo](#))
7. ~~Process Images~~ (Completed by Robin Holladay on 2/11/09 10:41 AM [Undo](#))
8. ~~Deploy Images to Staging server~~ (Completed by Robin Holladay on 2/11/09 10:41 AM [Undo](#))
9. ~~Archivist Reviews Collection Online on aaasite.si.edu~~ (Completed by Erin Corley on 2/11/09 11:17 AM [Undo](#))
10. ~~Archivist reports errors to be corrected~~ (Completed by Erin Corley on 2/11/09 11:17 AM [Undo](#))
11. ~~Archivist confirms errors have been corrected and approves Collection Online~~ (Completed by Erin Corley on 2/11/09 11:20 AM [Undo](#))
12. ~~Supervisor deploys XML file, HTML file, and Collection Online to Web~~ (Completed by Karen Weiss on 2/11/09 12:41 PM [Redo](#) | [Undo](#))
13. ~~Supervisor verifies pages on Web site (Collections Online, Col. A-Z, & Finding Aids), notifies staff~~
14. ~~Cataloger Updates Smithsonian Institution Research Information System (SIRIS)~~

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.

XML Output for Digital Asset System

```
-<collection_online_asset>
- <collection>
  <collection_title>Alexander Calder papers, 1926-1967</collection_title>
- <series>
  <series_title>
    Series 1: Biographical Material, 1928-1954, undated
  </series_title>
- <container>
  <container_title>Box 1, Folder 1, Address Lists, undated</container_title>
- <image>
  <imageid>81756</imageid>
  <image_date>undated</image_date>
  <keywords> Address Lists </keywords>
  <keywords> Biographical Material </keywords>
  <group_title>Alexander Calder papers</group_title>
  <collectionid>7294</collectionid>
  <collection_code>caldate</collection_code>
  <rights_summary>Not Researched</rights_summary>
- <structural_path>
  Alexander Calder papers / Biographical Material / Box_0001 / Folder_001
</structural_path>
<order_number>1</order_number>
<file_pathway>%CollectionOnline\caldate\Box_0001\Folder_001\file_pathway>
</image>
- <image>
  <imageid>81757</imageid>
  <image_date>undated</image_date>
  <keywords> Address Lists </keywords>
  <keywords> Biographical Material </keywords>
  <group_title>Alexander Calder papers</group_title>
  <collectionid>7294</collectionid>
  <collection_code>caldate</collection_code>
  <rights_summary>Not Researched</rights_summary>
- <structural_path>
  Alexander Calder papers / Biographical Material / Box_0001 / Folder_001
</structural_path>
<order_number>2</order_number>
```


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.