

End to End: Automating Digital Object Workflow in the AT



Jennifer Waxman, Preservation Archivist


Barbara Goldsmith Preservation and Conservation Department, NYU

Nathan Stevens, Programmer/Analyst

Digital Library Technology Services, NYU

End to End:

Automating Digital Object Workflow in the AT

- 
- How do we use our data to support our imaging workflow?
 - How do we automate workflow between archivists and the digital library department to compliment both the digital preservation infrastructure and the publication infrastructure?
 - How do we link digital content to the descriptive record in the Archivists' Toolkit without duplicating time and effort by archivist and staff?

End to End:

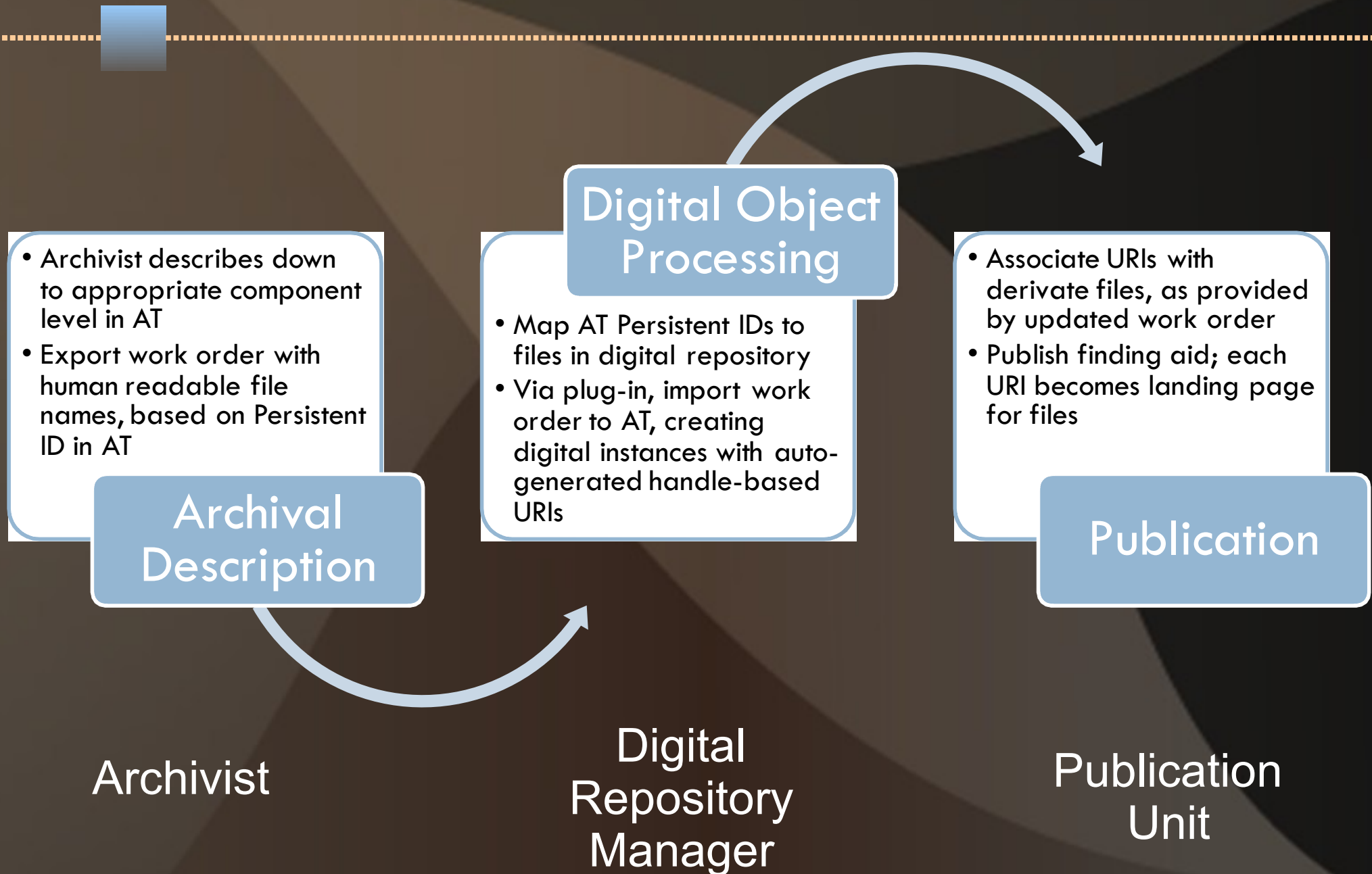
Automating Digital Object Workflow in the AT



Voices from the Food Revolution: People Who Changed the Way Americans Eat (MSS 309)

- Born-digital oral histories: 44 total interviewees (with multiple interviews)
- >170 files (.mp3, .doc, .docx)
- Intellectual arrangement: interviewee=series, interview=file
- Grant-funded: tied to deadline (for access)

End to End: Automating Digital Object Workflow in the AT



End to End: Automating Digital Object Workflow in the AT

Archival Description

Describes at
appropriate level

Export work
order

Resources

Title: Voices from the Food Revolution: People Who Changed The Way Americans Eat
An oral history project conducted by Judith Wein

Resource Identifier: MSS 309

Level: file Persistent ID: ref134 Component Unique Identifier:

Other Level:

Title: Interview 1

Wrap in tag...

Date Expression:

Date: Date Begin: Date End:

Bulk Dates: Bulk Date Begin: Bulk Date End:

Language Code:

Repository Processing Note:

Extent: Extent Number: Container Summary:

Instances: Instance Type Instance Label: Digital Object Interview 1

Add Instance Remove Instance

Internal Only Restrictions Apply

Add Child Add Sibling Rapid Data E... Delete Component Transfer Manage Locations Export EAD Export MARC Export Cont. Labels

First Previous Next Last Reports Close Save

Created: Mar 17, 2011 by lisadee | Modified: Aug 8, 2011 by lisadee | Record Number: 281

Add Child Add Sibling

✓ Rapid Data Entry
Default
SMA
DO Workflow
Barcode Cleaner

End to End: Automating Digital Object Workflow in the AT

Digital Object Processing

Map files to Persistent ID in work order

Import work order, via plug-in, to create digital instances



Persistent_Id	Series/Subseries	Title	Use Statement	Filename 1	Handle URI
MSS309_ref134	Dan Barber	Interview 1	Audio-Service		http://hdl.handle.net/10676/TEST-m37pw4x
MSS309_ref139	Lidia Bastianich	Interview 1	Audio-Service		http://hdl.handle.net/10676/TEST-gf1vj25
MSS309_ref25	Ariane Batterberry	Interview 1	Audio-Service		http://hdl.handle.net/10676/TEST-bnzs81p
MSS309_ref27	Michael Batterberry	Interview 1	Audio-Service		http://hdl.handle.net/10676/TEST-6wwq015
MSS309_ref29	Michael Batterberry	Interview 2	Audio-Service		http://hdl.handle.net/10676/TEST-34tmq0p

End to End: Automating Digital Object Workflow in the AT

Publication

Associate URIs
with derivate files

Publish finding
aid

The screenshot displays a digital object management interface with several panels. The top panel, titled 'Resources', shows a tree view of a collection 'MSS 309: Voices from the Food Revolution: People Who Changed The Way Americans Eat'. Underneath, there is a folder 'Interview 1' containing a file 'Interview 1'. The right panel shows metadata for this resource, including 'Level: file', 'Persistent ID: ref134', and 'Component Unique Identifier'. Below this, there are fields for 'Title' (Interview 1) and 'Extent'. The bottom panel, titled 'Digital Object', shows the 'Basic Description' tab with a 'Title' field containing 'Interview 1'. It also includes a 'Dates' section with 'Date Expression', 'Date Begin', and 'Date End' fields. A 'Resource Linked to this Digital Object' table shows a link to 'MSS 309: Voices from the Food R...'. At the bottom, there are 'Add File Version' and 'Remove File Version' buttons, and a 'URI' field containing 'http://hdl.handle.net/2333.1/k3j9kdrd'. There are also 'Export METS', 'Export MARC MARC', 'Export MODS MODS', and 'Export DC Dublin Core' buttons.

The screenshot shows a web page titled 'James Beard Interviews'. The page has a navigation bar with 'Home' and 'User login' options. The main content area is titled 'Ariane Batterberry Interview 1' and includes a video player with a play button. Below the video player, there is a 'Playlist' section with the following information: 'Ariane Batterberry Interview 1 Playlist', 'Citation URL: http://hdl.handle.net/2333.1/k3j9kdrd', 'Identifier: Batterberry_Ariane-interview_1', and 'Transcript: Batterberry_Ariane-2009_05_27-transcript.doc'. There is also a 'Log in or register to post comments' link at the bottom right.

End to End: Automating Digital Object Workflow in the AT

Digital Object Workflow 1.0 (07/07/2011)

Component Name **Series A: Miscellaneous, August 1937-May 1938**
Persistent ID **ref14**

Box # barcode item level file level

barcode item level file level file list

Test Only

URI Source Configuration

URI Source

Default URL

Thumbnail URL

User ID Password

End to End:

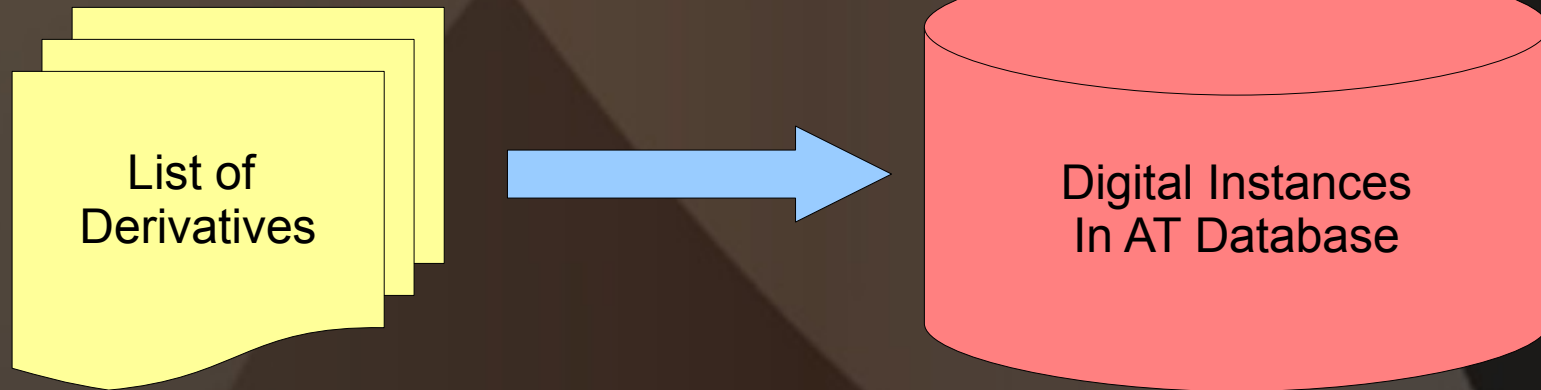
Automating Digital Object Workflow in the AT

Harry Randall: Fifteenth International Brigade Photograph Collection (ALBA PHOTO 011)

- 2 sets of photographs digitized by vendor at different times
- New (>650) and updated (>1100) handle-based digital instances needed to be created
- Filename of derivatives based on container indicator of analog instance
- Resource record contained original set of digital instances with static URIs

End to End: Automating Digital Object Workflow in the AT

Ideal Case: Go from list of derivatives to digital instances link to correct location in AT record



End to End:

Automating Digital Object Workflow in the AT

Input file contains list of derivatives and directives on how to create digital instances

```
# File contains the combined Randall photos master filenames  
# It's used as an import file for the BatchDO2 plug-in for creating and attaching digital  
# objects to the AT record (ALBA PHOTO 011 ).
```

```
instance1,-:_,thumbnail,m.tif:Image-Service
```

```
11_0001m.tif
```

```
11_0002m.tif
```

```
11_0003m.tif
```

```
11_0004m.tif
```

```
11_0005m.tif
```

```
11_0006m.tif
```

```
11_0007m.tif
```

```
....
```

End to End:

Automating Digital Object Workflow in the AT



Total process takes less than 5 minutes:

- Unlink original digital instances containing static URI
- Configure URI source
- Import files list
- Save record with new and updated digital instances

End to End: Automating Digital Object Workflow in the AT

The screenshot displays a digital object management interface. The main window, titled "Resources", shows a tree view on the left with a selected item: "Photo Unit #: A001 American engineers repair road to Quinto which was cut off by Fascists". The right pane shows the "Basic Description" tab with fields for Level (item), Persistent ID (ref15), Title, Date Expression (Aug 1937), and an Instances table.

Resources Record 1 of 2
Title: Harry Randall: Fifteenth International Brigade Photograph Collection Resource Identifier: ALBA PHOTO 011

ALBA PHOTO 011: Harry Randall: Fifteenth International Brigade Photograph Collection
Series A: Miscellaneous, August 1937-March 1938
Photo Unit #: A001 American engineers repair road to Quinto which was cut off by Fascists

Basic Description Names & Subjects Notes Etc.

Level: item Persistent ID: ref15
Other Level:
Title: Photo Unit #: A001 <lb/>American engineers repair road to Quinto which was cut off by Fascists
Wrap in tag...
Date Expression: Aug 1937
Date:
Date Begin: Date End:
Bulk Dates:
Bulk Date Begin: Bulk Date End:

Component Unique Identifier:
Extent: Extent Number:
Container Summary:
Instances:

Instance Type	Instance Label
Digital Object	Photo Unit #: A001...
Digital Object	Photo Unit #: A001...
Mixed materials	item 11-1213

Remove Instance
Restrictions Apply
Export MARC Export Cont. Labels

Digital Object Import "ALBA_RANDALL.txt"

Import Log

Line 1648 Created Digital Object for instance 11_1851, URI : http://dlib.nyu.edu/11_1851m.tif
Line 1648 Created Digital Object for instance 11_1851, URI : http://dlib.nyu.edu/11_1851m.tif?urlappend=/mode/thumb

3250 Digital Objects Created ...

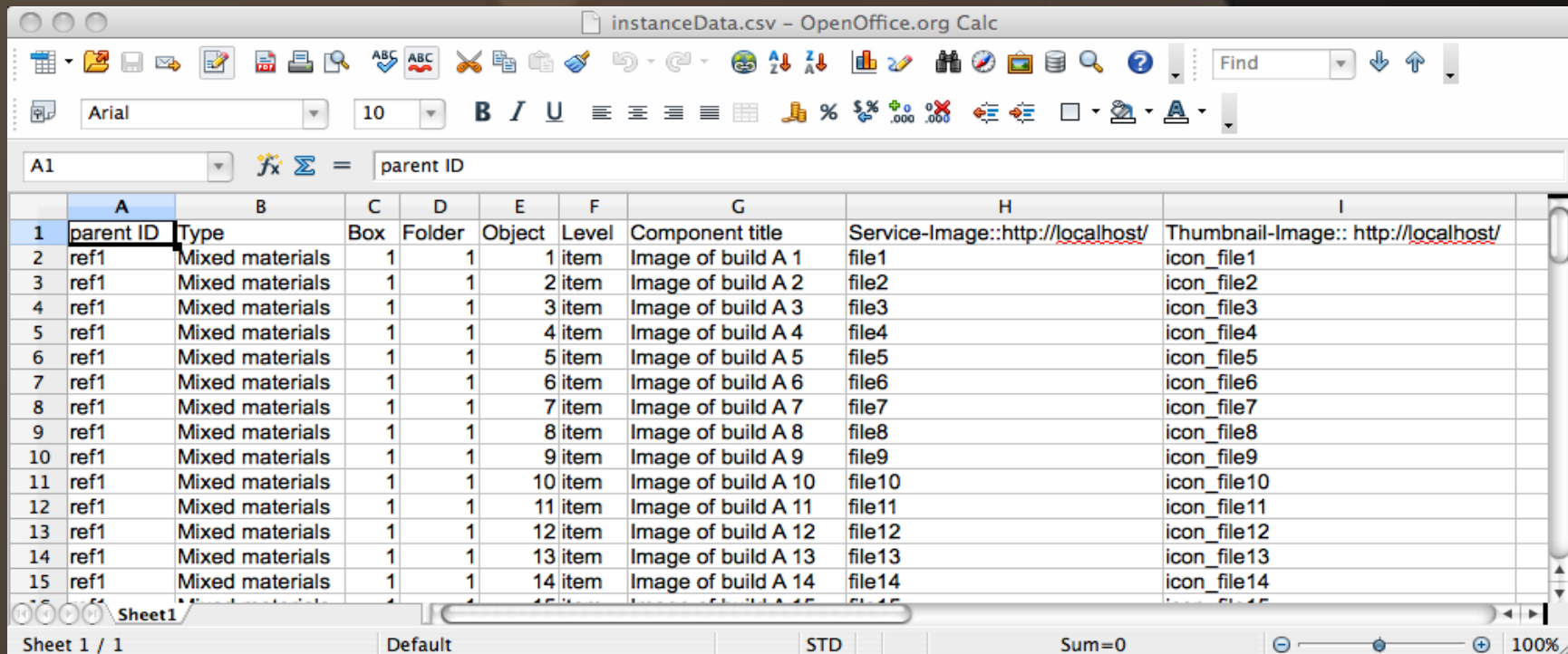
No Component found for "11_1162m.tif" on line # 1174
No Component found for "11_1189m.tif" on line # 1200

Copy Print Save OK

End to End: Automating Digital Object Workflow in the AT

Potential Enhancements?

- More complete automation starting with file name:
file name \Rightarrow component \Rightarrow analog instance \Rightarrow digital instance



The screenshot shows a spreadsheet titled "instanceData.csv" in OpenOffice.org Calc. The spreadsheet contains a table with the following data:

	A	B	C	D	E	F	G	H	I
1	parent ID	Type	Box	Folder	Object	Level	Component title	Service-Image::http://localhost/	Thumbnail-Image:: http://localhost/
2	ref1	Mixed materials	1	1	1 item	Image of build A 1	file1	icon_file1	
3	ref1	Mixed materials	1	1	2 item	Image of build A 2	file2	icon_file2	
4	ref1	Mixed materials	1	1	3 item	Image of build A 3	file3	icon_file3	
5	ref1	Mixed materials	1	1	4 item	Image of build A 4	file4	icon_file4	
6	ref1	Mixed materials	1	1	5 item	Image of build A 5	file5	icon_file5	
7	ref1	Mixed materials	1	1	6 item	Image of build A 6	file6	icon_file6	
8	ref1	Mixed materials	1	1	7 item	Image of build A 7	file7	icon_file7	
9	ref1	Mixed materials	1	1	8 item	Image of build A 8	file8	icon_file8	
10	ref1	Mixed materials	1	1	9 item	Image of build A 9	file9	icon_file9	
11	ref1	Mixed materials	1	1	10 item	Image of build A 10	file10	icon_file10	
12	ref1	Mixed materials	1	1	11 item	Image of build A 11	file11	icon_file11	
13	ref1	Mixed materials	1	1	12 item	Image of build A 12	file12	icon_file12	
14	ref1	Mixed materials	1	1	13 item	Image of build A 13	file13	icon_file13	
15	ref1	Mixed materials	1	1	14 item	Image of build A 14	file14	icon_file14	

End to End: Automating Digital Object Workflow in the AT



BatchDO2 Plug-in can be found here:
<http://archiviststoolkit.org/node/246>

Thank You