Using “batches” to create workflows in CONTENTdm for shared projects within the library

CONTENTdm Southern Users Group Meeting 2014

Jill Ellern, Systems Librarian
Paromita Biswas, Metadata Librarian
Western Carolina University
Outline:

• History of the use of CONTENTdm at Hunter Library, WCU
• Workflow in a shared project environment
• Using the Client in a shared environment
A collaborative history:

First digitization project: *Craft Revival*

A multi-institutional project

A collection of over 5000 documents, photographs, and artifacts
Craft Revival: Shaping Western North Carolina Past and Present

The Story
The People
The Crafts
The Collection
About
K-12
Resources
Contact Us

The Story

The Crafts

The People

The Collection

Telling the story of an exciting regional movement
Creating handicrafts and preserving traditions from the 1890s to 1940s
History of WCU’s use of CONTENTdm

• Use of Contentdm started with a multi-year grant for the *Craft Revival* project (2005-2009)

• Project partners: Hunter Library and 6 small cultural heritage institutions

• Challenges:
  
  Partners’ inexperience with digital projects and Contentdm; lack of staff time and internet access
Craft Revival project workflow for the partners’:

Partners’ lack of internet access meant:
-- Partners’ contributed their images and metadata outside of Contentdm and sent it to Hunter Library

-- Pictures and metadata were put on flash drives and CDs: Word/Excel files with images in JPEG and TIFF

-- This physical package was called a “batch”

-- Batches were subject specific
A “Batch”
Craft Revival project workflow at Hunter Library:

• Project team:
  --Project Director/Content Specialist (Digital Initiatives Librarian)
  --Cataloging/metadata staff
  --Systems unit staff
  --Project assistants

• Project team members added/enhanced partners’ metadata in the “batches”

• Metadata Librarian added controlled vocabularies, checked the data, uploaded it into Contentdm, and published it online

• After publication the “batches” were returned to Digital Initiatives in Hunter Library
A cumbersome process

A time consuming workflow for *Craft Revival*:

--Entering metadata into “batches”—a linear process of metadata creation in stages

--Team members had to follow detailed instructions for working with metadata in spreadsheets

--Only the Metadata Librarian worked with Contentdm (adding and uploading the metadata from the “batches”/spreadsheets)
Detailed instructions for filling out Excel spreadsheets for creation of single and compound objects:

Craft Revival Batch Instructions for those digitizing items and entering metadata into the spreadsheet
Created February 2011

Where to find the spreadsheet template
- Metadata spreadsheets are available on the H: drive:
  H:\Library\SHARED\Craft Revival\metadata\metadata_template-new.xlsx
- If that file is accidentally overwritten or deleted, there are also extra copies available:
  H:\Library\SHARED\Craft Revival\metadata\templates-extra-copies

Setting up a batch
- Batches should be in folders named by date, as usual. A complete batch will include the tif, files, jpegs, and the metadata spreadsheet for the batch.
- If the batch contains both single items and compound objects, you will need to do one spreadsheet for all of the single items and one spreadsheet for each compound object.

Single items
- Here is an example of how to set up a batch of single items:
  H:\Library\SHARED\Craft Revival\metadata\example-batch-single-items

- Create a “jpegs” folder for the single item jpegs - no other files should go in this folder - no tifs, no metadata spreadsheet. Put the tifs, metadata spreadsheet, and any other files in the batch folder that contains the jpeg folder.

Compound objects
- And here is what a compound object should look like:
  H:\Library\SHARED\Craft Revival\metadata\example-batch-compound-object
Detailed instructions for filling out spreadsheets for metadata creation:

- There are more fields in the spreadsheet than in the Word document. Only fill in the white fields – do not fill in the gray or the black fields. If some of the white fields do not apply to your item, leave them blank (for example, Staff Notes, Contributor, County, and Language are among fields that are not always used on each item).
- Some fields include drop-down boxes to allow you to pick from a controlled vocabulary. These drop-downs will only allow one choice from each list, so if for some reason you need to enter more than one item, you’ll have to clear the cell to delete the drop-down menu.
- For cells with more than one entry, use semi-colons between entries. Example (creator field):

<table>
<thead>
<tr>
<th>Decade</th>
<th>County</th>
<th>Creator</th>
<th>Contributor</th>
<th>Source Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970s</td>
<td>Cherokee Indian Reservation, N.C.</td>
<td>unknown; Wilnoty, John Julius, 1940</td>
<td>Qualia Arts &amp; Crafts Mutual</td>
<td></td>
</tr>
</tbody>
</table>

- Feel free to change the size of the cells in order to make them more readable or easier to use. You can control the size by clicking and dragging a row or column in the alphabetic bar along the top or the numeric bar on the left. To make the text wrap to fit a cell, use the “Wrap Text” button.
- You can copy and paste or copy and drag to fill multiple cells with the same information.
- One new field is the last one, Object Filename. This field is filled in with the corresponding name of the jpeg for each item. The process is slightly different for single items and for compound objects.

**Single Items**
- The first row always includes the name of each field. Then, each item has one row below the field names, and the last column of that row has the jpeg name. Example:

<table>
<thead>
<tr>
<th>AE</th>
<th>AF</th>
<th>AG</th>
<th>AH</th>
<th>AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Format</td>
<td>Hardware</td>
<td>Software</td>
<td>Larger Image</td>
<td>Object filename (ex: SHCG_Photo_4_5_2.jpg)</td>
</tr>
<tr>
<td>JPEG 300dpi</td>
<td>HP Scanjet 8300</td>
<td>Adobe Photoshop</td>
<td>QACM_Wilnoty_JohnJulius_02.jpg</td>
<td></td>
</tr>
<tr>
<td>JPEG</td>
<td>HP Scanjet</td>
<td>Adobe</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table Data**

- JPEG
- 300dpi
- HP Scanjet
- Adobe
- Photoshop
- QACM_Wilnoty_JohnJulius_02.jpg
Newer & more streamlined workflow for digital projects at Hunter Library:

• No longer entering metadata in spreadsheets

• Converted the flash drives/CDs/metadata ("batches") to a shared server space

• Continued to allow multiple people to enter metadata at different levels of record creation: but everyone working with digital collections now had access to the Contentdm Client and put in data directly into the Client during record creation

• Continued using our old term “batches” for shared “projects” in the Client
Model we opted for: shared “batches” not individual “batches”

• Shared batches/projects:
  – Files are located on a shared drive
  – Multiple people do different parts of each record
  – Cataloging/Metadata staff review each record, upload, approve, and publish it online

• Individual batches/projects:
  – One person does all the data entry
  – Actual client files can be anywhere
  – Uploads for approval by Web Administration
CONTENTdm/Digital Projects workflow at Hunter Library

Scanning and visible data entry
- Digital Tech
- Image scanning and watermarking
- Basic data identified and described

Description and initial subject
- Digital Librarian
- Augmenting of visible data
- Description and selection of basic subject data

Subject and Authority Work
- Metadata Tech/Librarians
- In depth checking of all fields
- Adding of new subjects

Published to CONTENTdm
Shared metadata work for a “batch”:

Examples of metadata entry work for different fields in Contentdm Client:

- **Digital Technician**: Type; Medium of original; Dimensions; Location; Source institution

- **Digital Initiatives Librarian**: Date of original; Description; Local Subject

- **Cataloging staff/ Metadata Librarian**: Controlled vocabularies; overall quality control of the record before publishing it online
Client software in a shared environment

- Basic architecture of the CONTENTdm
- Benefits of shared environment
- Challenges with the current software
- Limitations of the client
Basic CONTENTdm Shared Environment
What Client Project files look like

* Subdirectory for each project
  - Hosted are named by server and collection
    - p16232coll1
  - Subdirectories under project for compound items
  - Localdataset.xml
    - What files are currently active in the subdirectory
  - Client is not good at deleting files once done with them
* Types of files
  - Jpg, Icon’s jpg, desc, history, profileInformation
    - In xml format
    - Filename are often include the IP of the creator, full date, number
Benefits of shared environment

• Many people look at each record
  – Standardization and consistency
  – Quality of cataloging
• Use of shared fields across collections
CONTENTdm IT Challenges

• Naming problems
• Locations of project files
• Username issues
• Not flexible for project needs that change over time
Naming problems

• Limited space in the window for open and existing project
  – Doesn’t have left and right slider
  – Limits the names of projects and collections because the names need to be short so you can fit in screen
• Finding the name of a project under the “pnumber” name of the subdirectory can be challenging
Which one is the Craft Revival project in?
File location issues

• Being very careful when creating shared files
  – To find it again and place them together for later

• Frustrating when importing a shared project
  – Doesn’t play nice in a server environment
  – Have to click through all the paths for every project
  – Ctrl C and Ctrl V are your friend
Shared area on the server where all the project files are kept.
Location you pick will dictate the place the directory is placed.
Username issues

• Information about the client is kept in both the registry and the data files
  – Not updated when the client is opened
  – Difficult to modify later
• They have to match or there are problems
  – File location for “all users” not available
  – Projects are installed for each username on machine
• Environment (file location) can’t change
  – Can’t copy the directory and take home
Setup in the CURRENT_USER location

Local dataset file expects data as well

```xml
<cxmldocument>
  <title/>
  <cisoptr/>
  <compoundobjectdirectory>L:\\shared\\depts\\dat\digital projects\\p16232coll10\\great smoky - batch 16\\152_30_162_19_2013-10-10_113810</compoundobjectdirectory>
  <objectfilename>index.cpd</objectfilename>
  <metadatafilename>index.desc</metadatafilename>
  <iconfilename>iconindex.jpg</iconfilename>
</cxmldocument>
```
Not flexible for project needs that change over time

- Server changes
- Renaming files/directories
- Changing your mind about who has access to a project
  - Adding and removing projects assigned to an individual
- Removing fields
Limitations of the client

• Moving images and metadata between projects
  – Excel can be used if it’s worth it for the number of images
• Locked records
• Leaving old files in the folders after uploading
• Data corruption worries
  – Users are warned about too many records in one project
  – Sorting large files in client is slow
Questions?

Read more about the Craft Revival project at:

- See our digital collections at http://www.wcu.edu/hunter-library/collections/digital-collections.asp

elloren@email.wcu.edu
pbiswas@email.wcu.edu
Using “batches” to create workflows in CONTENTdm for shared projects within the library

CONTENTdm Southern Users Group Meeting 2014

Jill Ellern, Systems Librarian
Paromita Biswas, Metadata Librarian
Western Carolina University