Migratory Patterns in IRs: CONTENTdm, Digital Commons and Flying the Coop

Elizabeth Chance
Fort Hays State University

Michele Gibney
University of the Pacific

Kristin Laughtin-Dunker
Chapman University, laughtin@chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/library_presentations

Recommended Citation
Migratory Patterns in IRs: CONTENTdm, Digital Commons and Flying the Coop

Comments
Presented at the 2018 Digital Initiatives Symposium at the University of San Diego.

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

This conference proceeding is available at Chapman University Digital Commons: https://digitalcommons.chapman.edu/library_presentations/20
Migratory Patterns in IRs: CONTENTdm, Digital Commons and Flying the Coop

Elizabeth Chance, Fort Hays State University
Michele Gibney, University of the Pacific
Kristin Laughtin-Dunker, Chapman University
Outline

- Rationales
- Migration Workflows & Usage Stats
  - Fort Hays
  - Chapman
  - Pacific
- Discussion
CHAPMAN & PACIFIC

- Launched an institutional repository using Digital Commons (DC) platform
- Rising dissatisfaction with CONTENTdm, where digital Special Collections and Archives were housed
- DC supports text & image files - the bulk of Special Collections – therefore, combining was cost-effective

FORT HAYS

- Acknowledged that CONTENTdm and Digital Commons do different things
- CONTENTdm cost is shared with other departments so decision was made to maintain both platforms but audit collections to determine which platform was the best choice for the content.
Workflows & Usage Differences

Fort Hays - Targeted collections – on our own

Chapman - All of it – on our own

Pacific - All of it – bepress paid service
Fort Hays State University

Elizabeth Chance, MLIS
Digital Curation Librarian
Fort Hays
Targeted collections

Rationale

CONTENTdm

- Does well with image galleries
- Allows for in-depth description of complex items like scrapbooks
- Has native streaming abilities for audio and video
- Can easily present a variety of file formats in a single collection
- Has a hard time with large .pdf files
- New responsive website doesn’t have a page-flip view for .pdf
- Difficult to embed third-party content

DIGITAL COMMONS

- Great for text-based items
- Requires less metadata to get the discovery job done
- Better integration for OCR and in-text searching
- File formats other than .pdf require extra steps
- Can easily embed third-party content like book readers and YouTube videos
- No mechanism for handling compound objects like front and back of a photograph or items in a scrapbook
Fort Hays
Targeted collections

*Initial Steps*

- Identified qualities that warranted moving a collection
  - Text-based collections
  - .pdf collections
  - Anything that would benefit from better integration of OCR
  - Anything that would be better presented in a “book reader” format
  - Scholarly works that would benefit from inclusion in the Digital Commons network
Fort Hays
Targeted collections
Collections Moved

- Journals
  - Academic Leadership: The Online Journal
  - Journal of Business & Leadership

- Archival Collections
  - Reveille Yearbooks
  - Athletics Programs

- Graduate Student Works
  - Master of Liberal Studies Research Papers
  - Masters Theses Collection (In Progress)
Fort Hays
Targeted collections

General Workflow

- Used Bepress batch upload process
  - pdf’s were reduced in size to get them under 100mb
    - This allowed us to use the first page as the cover image
  - Re-ran OCR on all the files
  - Uploaded files to our library upload server using Filezilla
  - Created new metadata where necessary and cleaned up old metadata where possible
  - When required objects were uploaded to the Internet Archive so we could embed the book reader
- Took the opportunity to inventory all of the master scans so they could be accessioned into our formal preservation system
  - Standardized file names
The Reveille
- Official Yearbook of Fort Hays State University
  - Published from 1914-2003
- Initially Digitized in 2009: Reveille 1.0
  - Presented as its own collection in CONTENTdm
  - Volumes were photographed and .pdf files were created
  - Re-mastered in 2014 to mixed results
- Library leadership expressed a desire to get more out of this collection
  - Wanted to see increased usage
    - Access and discovery were difficult due to metadata issues and loading times
  - Wanted to see improved user experience
Fort Hays
Targeted collections
*The Reveille 3.0*

- Book Gallery highlights cover art
- Uses the collections tool to create decade sorted sub-galleries
- Embedded book reader
Fort Hays
Targeted collections
Usage Differences

Average Views Per Item

<table>
<thead>
<tr>
<th>Collection</th>
<th>CONTENTdm</th>
<th>Digital Commons</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALJ</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>JBL</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Reveille</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Athletics Programs</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Chapman University

Kristin Laughtin-Dunker, MLIS
Coordinator of Scholarly Communications & Electronic Resources
Step 1: Export metadata spreadsheets from CONTENTdm

Step 2: Recreate collection structures in DC

Step 3: Convert CONTENTdm spreadsheets to DC Batch Upload Excel sheets

Step 4: Missing URLs to the objects in CONTENTdm...

Step 5: Piece together FrankenURLs
The Problems of CONTENTdm URLs

- The metadata exports from CONTENTdm contained the URL for the record for each item.
- However, in order to import an item, Digital Commons needs a direct URL to it, not the record.
- Right-clicking the download button in CONTENTdm and copying the link address also did not work.
- Thus, we had to come up with a system to generate direct URLs to each item ourselves.
For each FrankenURL, the following pieces of information from ContentDM are needed:

- **Instance Identifier**: the unique numeric identifier for your university/instance
- **Collection Identifier**: the unique numeric identifier for the collection
- **Item Identifier**: the unique numeric identifier for each individual item
- **Width and Height** (for images, optional)
- **Filename** (for PDFs or other non-image documents)
The general formula for PDFs is:


- Chapman’s Instance identifier = 15046.

For the item “116.pdf”, which was item 115 in collection 20, the FrankenURL would be:

If you can get all of the separate pieces of information into an Excel sheet, it is easy to use the Concatenate function to create the FrankenURLs.

- Column B: the collection identifier
- Column C: /id/
- Column D: the item identifier
- Column E: /filename/
- Column F: the file name and file extension

You could then concatenate columns A-F together to create the FrankenURLs in column G, then “Paste Values” to add the text of the URLs to your batch upload sheets.
FrankenURL format for image files:


FrankenURL for item 20 in collection 1:

- http://cdm15046.contentdm.oclc.org/utils/ajaxhelper/?CISOROOT=p15046coll1&CISOPT R=20&action=2&DMSCALE=100&DMWIDTH=9999&DMHEIGHT=9999
FrankenURL format for image files:

- Tip! Put 9999 for the width and height, rather than trying to look up the width and height of each image. Digital Commons will import them at their correct size.
  - However, if you put in a number that’s too small, Digital Commons will only import part of that image and cut the rest off.

- You can also use the Concatenate function to build the FrankenURLs in a spreadsheet, then Paste Values into your batch upload sheet:
  - Column A: the URL root:
  - Column B: the collection identifier
  - Column C: &CISOPTR=
  - Column D: the item identifier
  - Column E: &action=2&DMSCALE=100&DMWIDTH=9999&DMHEIGHT=9999
Exceptions

- These tips will work for most instances of ContentDM, especially if the collection identifiers are numeric.

- However, if your instance uses alphabetical or Quick Start collection identifiers, you may have to tweak the FrankenURLs a bit.

- In the base URL, use the following section as the collection identifier in your FrankenURLs:
  - https://[instanceidentifier].contentdm.oclc.org/cdm/ref/collection/[collectionidentifier]/id/[itemididentifier]
  - For alphabetical collection identifiers, it will be a string of letters.
  - For Quick Start collections, it may include the instance identifier, and will end with “qs”.
Exceptions

- Once you have found your **collection identifier**, plug it into the FrankenURL after the /collection/ portion for PDFs or the CISOROOT portion for images:
  - PDFs
    - `http://cdm\[instanceidentifier\].contentdm.oclc.org/utils/getfile/collection/\[collectionidentifier\]/id/\[itemidentifier\]/filename/\[filename\].\[fileextension\]
  - Images
    - `http://cdm\[instanceidentifier\].contentdm.oclc.org/utils/ajaxhelper/?CISOROOT=\[collectionidentifier\]&CISOPTR=\[itemidentifier\]&action=2&DMSCALE=100&DMWIDTH=9999&DMHEIGHT=9999`
- Note that you do not need to put “p\[instanceidentifier\]” before the **collection identifiers** here, as in earlier examples.
Exceptions

- For images: we did experience a few instances where after clicking on or pasting the FrankenURL, we could not view the image on our computers. However, they did import successfully into Digital Commons.

- This workaround was only discovered in consultation with two other institutions in January and February 2018, so it has not been tested on ContentDM instances with purely numeric collection identifiers (like Chapman’s were).
On ContentDM, usage had peaked at about 5,000 utilizations (page hits) per month in 2012, and had declined to about 2,300 utilizations per month in 2014.

Digital Commons usage:
- First year (September 2014-August 2015):
  - About 5,100 objects
  - 6,381 page hits (531/month)
  - 4,874 downloads (406/month)
- Second year (September 2015-August 2016):
  - About 5,700 objects
  - 16,623 page hits (1,385/month)
  - 63,892 downloads (5,324/month)
- Third year (September 2016-August 2017):
  - About 7,800 objects
  - 21,276 page hits (1,773/month)
  - 70,764 downloads (5,894/month)
- First five months of fourth year (September 2017-February 2018):
  - About 8,000 objects
  - 11,927 page hits (2,385/month)
  - 41,825 downloads (8,365/month)
Chapman
All of it
Usage
Special Collections objects account for about 40% of our content and 40% of our downloads.

4 or 5 of our Top 10 Downloads are consistently Special Collections objects that had been in ContentDM.

- 2 of these are photos of Miss Universe 1970, who is still a television personality in Puerto Rico. When people Google her, they are finding our photos, often as the top result.

- We have gotten many reproduction requests for our Japanese propaganda poster and United Press International photo collections.
University of the Pacific
Michele Gibney
Digital Repository Coordinator
“Some one else does the work? SIGN ME UP.”
Pacific All of it bepress paid service

35,000+ items
216 structures
bepress requires:

- Structure collection (type of structure, URLs, titles, Grouping, etc.)
- Metadata fields for each structure *(double check all spellings)*
- Administrators
- OAI-PMH URLs from CONTENTdm with *UNIQUE* Dublin Core mapping for each field for each structure
- Migration mapping chart of fields from CONTENTdm to DC *FOR EACH STRUCTURE* *(I’ll say it again: double check all spellings)*

And then you sit back and wait

- And wait
  - And wait
    - And wait....
Pacific All of it OAI-PMH Dublin Core mapping
Problems

- **COMPOUND OBJECTS** (DC doesn’t handle these well)
  - Empty top level rows
  - Copy/paste nightmare
  - Massive # of items = search and revision difficulties

- Certain fields
  - Circa dates/original dates
  - Author standardization

- JPEG2000 image files
  - Download/Access
  - Naming structure

- Embedded audio/video

- Needs of harvesters
Pacific
All of it
Compound Objects
Pacific
All of it
Embedded A/V

CONTENTdm hosted

Digital Commons does not host
Pacific
All of it
Partial Manually

- Multi image records VS PDFs (preservation VS ease of use)
- Multi-format type collections

Led to...

- Manual migrations
  - Metadata export
  - Capturing full text URLs in CDM
  - Conversion to single PDFs
  - Batch uploading
Content added to Digital Commons in May 2017
CONTENTdm subscription ended in October 2017
Pacific
All of it
Usage

Download Map for all collections in Digital Commons
Discussion
Discussion

- Discuss factors affecting migration
  - File preservation
  - Historical metadata preservation
  - Preserving historical usage data

- Determine the pro/cons of migrations
  - Migrate targeted collections
  - Consolidate platforms
  - Consider labor involved: DIY vs. paid service options

- Do your research
  - Look at collections from similar institutions
  - Talk to librarians who have done migrations
  - Document your process to act as a resource in the future
THANK YOU

Questions?

Elizabeth Chance
Digital Curation Librarian
Fort Hays State University
mechance2@fhsu.edu

Michele Gibney
Digital Repository Coordinator
University of the Pacific
mgibney@pacific.edu

Kristin Laughtin-Dunker
Coordinator of Scholarly
Communications & Electronic
Resources
Chapman University
laughtin@chapman.edu