Using MARC Records to Populate CONTENTdm

Wendy Robertson

University of Iowa
Using MARC Records to Populate CONTENTdm

Wendy Robertson
The University of Iowa Libraries

Midwest CONTENTdm Users Group
3rd Annual Meeting,
Indianapolis, IN, April 30, 2008
Overview

- Connexion changes
- Connexion setup
- Example workflows:
  - Individual selection of previously cataloged materials
  - Batch selection of currently cataloged materials
  - Individual selection of currently cataloged materials
OCLC Connexion Import for Digital Collections Now Available

The new OCLC Connexion digital import feature allows catalogers using the Connexion client (version 2.0) to add digital items to CONTENTdm collections during the Connexion cataloging process.

This new feature streamlines digital collection creation by integrating it with standard cataloging workflows.

The latest release of the CONTENTdm® Digital Collection Management Software (version 4.3) supports the Connexion digital import and makes it even easier for libraries to integrate digital collection growth into their current cataloging workflows.

"The integration of CONTENTdm and Connexion through the Connexion digital import helps our staff manage digital content in a way that is fully integrated with our existing cataloging workflow. And, more importantly, our users have unprecedented access to our digital collection through our local Montana Shared Catalog and through WorldCat." --Jennie Stapp, Content Specialist Librarian, Montana State Library

Connexion catalogers can attach a digital file (or files, if more than one digital file is associated with the same record) to their WorldCat record by using the Connexion client. The digital file is then added to the designated CONTENTdm collection along with its associated metadata record. Metadata records are mapped to Qualified Dublin Core when added to the CONTENTdm collection, TIFF images are converted to JPEG2000, and OCR is performed for TIFF images comprising compound objects. PDF processing occurs if enabled in the CONTENTdm collection. Additionally, the OCLC number from the WorldCat record is added to the new OCLC number administrative field in CONTENTdm.
Disappointment

OCLC Connexion Import for Digital Collections Now Available

The new OCLC Connexion digital import feature allows catalogers using the Connexion Client to add digital items to CONTENTdm collections during the Connexion cataloging process.

This new feature streamlines digital collection creation by integrating OCLC cataloging workflows.

The latest release of the CONTENTdm® Digital Collection Management software (version 4.3) supports the Connexion digital import and makes it even easier for library staff integrating digital collection growth into their current cataloging workflows.

"The integration of CONTENTdm® 4.3 with Connexion digital import helps our staff manage digital content in a way that is fully integrated with our cataloging workflow. And, more importantly, our users have unprecedented access to our digital collection through our local Montana Shared Catalog and through WorldCat." -- Jennie Stapp, Content Specialist, Montana State University Library.

Connexion digital import allows the association of a digital file (or files, if more than one digital file is associated with the same record) to the record held in Connexion using the Connexion client. The digital file is then added to the designated CONTENTdm collection as a linked metadata record. Metadata records are mapped to Qualified Dublin Core when added to the CONTENTdm collection. TIFF images are converted to JPEG2000, and OCR is performed for TIFF images comprising the record. OCLC processing occurs if enabled in the CONTENTdm collection. Additionally, the OCLC number from the record is added to the new OCLC number administrative field in CONTENTdm.
Workaround

- Repurpose quality MARC records in CDM in a non-hosted environment
- Takes some data manipulation
- Allows us the opportunity to make metadata format consistent with rest of collections
Connexion Updates

• Connexion client 2.00 (Completed: June 2007)
  – Export bibliographic records in Dublin Core Qualified in XML with the UTF-8 character set.
  – Mappings from MARC to DC not complete

• Connexion client 2.10 (Completed: December 2007)
  – Export bibliographic records in additional Dublin Core formats including Dublin Core Simple, OCLC Dublin Core Simple, and OCLC Dublin Core Qualified.
Connection Set-up

- Under Tools—Options—Export tab

Click to change from MARC21 to Dublin Core

Create file of records

OCLC option includes OCLC number in metadata
International Dada Archive

• Fully cataloged materials
• Digital collection moving to CDM
Search

• Institutional record for each title found
Export

- Institutional Record exported to dat file
Qualified DC fields

- dc:contributor
- dc:creator
- dc:description
- dc:language
- dc:publisher
- dc:relation
- dc:subject xsi:type='http://purl.org/dc/terms/DDC'
- dc:subject xsi:type='http://purl.org/dc/terms/LCC'
- dc:subject xsi:type='http://purl.org/dc/terms/LCSH'
- dc:title
- dc:type
- dcterms:alternative
- dcterms:extent
- dcterms:isPartOf
- dcterms:isReferencedBy
- dcterms:issued <dcterms:tableOfContents
OCLC QDC fields

- oclcterms:languageOfCataloging
  xsi:type='http://purl.org/oclc/terms/iso639-2b'
- oclcterms:OCLC-MARCProcessingInformation
  xsi:type='http://purl.org/oclc/terms/oclcorg'
- oclcterms:OCLC-MARCProcessingInformation
  xsi:type='http://purl.org/oclc/terms/transactioncode'
- oclcterms:recordContentSource
- oclcterms:recordControlNumberIdentifier
  xsi:type='http://purl.org/oclc/terms/marcorg'
- oclcterms:recordCreationDate
  xsi:type='http://purl.org/oclc/terms/marc008date'
- oclcterms:recordIdentifier
  xsi:type='http://purl.org/oclc/terms/lccn'
- oclcterms:recordIdentifier
  xsi:type='http://purl.org/oclc/terms/oclcrecordnumber'
- oclcterms:recordTranscribingAgency
Conversion to Delimited Output

- Programmer makes changes

<table>
<thead>
<tr>
<th>Source</th>
<th>Control Num</th>
<th>Date</th>
<th>OCLCnum</th>
<th>InstCode</th>
<th>TransCode</th>
<th>TransA</th>
<th>Creator</th>
<th>Contributor</th>
<th>Issued</th>
<th>Description</th>
<th>Extent</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTY-BR</td>
<td>OCoLC</td>
<td>881205</td>
<td>ocn148490109</td>
<td>NUI</td>
<td>CD</td>
<td>NUI</td>
<td>Albert-Briot, Pierre.</td>
<td>[1919]</td>
<td>Ce livre a été tiré de 92 p.; 1 ill.; 19 cm.</td>
<td>fre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTY-BR</td>
<td>OCoLC</td>
<td>900820</td>
<td>ocn149454363</td>
<td>NUI</td>
<td>CD</td>
<td>NUI</td>
<td>Arnauld, Céline.</td>
<td>[1919]</td>
<td>Il a été tiré de 76 p.; 1 ill.; 22 cm.</td>
<td>fre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IaU</td>
<td>OCoLC</td>
<td>850903</td>
<td>ocn148471811</td>
<td>NUI</td>
<td>CD</td>
<td>NUI</td>
<td>Aragon, Louis</td>
<td>[1922]</td>
<td>Library has cop. 95 p. 19 cm.</td>
<td>fre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IaU</td>
<td>OCoLC</td>
<td>810108</td>
<td>ocn148451281</td>
<td>NUI</td>
<td>CD</td>
<td>NUI</td>
<td>Doesburg, Theo van, 1883-1931.</td>
<td>[1921]</td>
<td>31 p., xvii p. of plates; ill.; 23 cm.</td>
<td>fre</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publisher</th>
<th>Subject LCC</th>
<th>SubjectLC</th>
<th>Title</th>
<th>Alternative</th>
<th>IsPartOf</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editions &quot;Ste&quot;</td>
<td></td>
<td></td>
<td>Larountala : polydrame en deux parties</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Editions de &quot;L'esprit nouveau&quot;</td>
<td>PQ2601.R678</td>
<td>1919</td>
<td>Tourvêtre : roman</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verlag der Weissen Bücher</td>
<td>ND548.N351913</td>
<td>Painting. French Painting</td>
<td>Neue französische Malerei</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. Steegemann</td>
<td></td>
<td></td>
<td>Die Wolkpumpe</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nouvelle Revue Française</td>
<td>PQ2601.R2A81919</td>
<td></td>
<td>Anicet, ou, Le panorama</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Editions de la Nouvelle Revue Française</td>
<td>PQ2601.R2A81919</td>
<td>Les aventures de Télémaque</td>
<td>Text</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Au Sans pareil</td>
<td>PQ2601.R2F4</td>
<td></td>
<td>Feu de joie</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erich Reiss</td>
<td>PT260.A37F6</td>
<td>Flametti, oder, Vom Dandysmus der Armen : Roman</td>
<td>Collection de littérature, 7</td>
<td>Text</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Routine Data Manipulation

- Remove final periods
- Remove c and brackets around dates, pages
- Remove most of the administrative data
- Remove 3 letter code from OCLC number
- Change language abbreviation to full word

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OCLCnum</td>
<td>Creator</td>
<td>C</td>
<td>Issued</td>
<td>Description</td>
<td>Extent</td>
<td>Language</td>
<td>Publisher</td>
</tr>
<tr>
<td>2</td>
<td>148490109</td>
<td>Albert-Birot, Pierre</td>
<td>1919</td>
<td>Ce livre a été 92 p.; 1 ill.; 19 cm.</td>
<td></td>
<td></td>
<td>French</td>
<td>Editions &quot;Stic&quot;</td>
</tr>
<tr>
<td>3</td>
<td>149434363</td>
<td>Arnauld, Céline</td>
<td>1919</td>
<td>Il a été tiré de 76 p.; 1 ill.; 22 cm.</td>
<td></td>
<td></td>
<td>French</td>
<td>Editions de &quot;L'Esprit nouveau&quot;</td>
</tr>
<tr>
<td>4</td>
<td>149945560</td>
<td>Arp, Jean, 1887-1966</td>
<td>1913</td>
<td>15 p., 15 leaves of plates: ill.; 25 cm.</td>
<td></td>
<td></td>
<td>German</td>
<td>Verlag der Weissen Bücher</td>
</tr>
<tr>
<td>5</td>
<td>148636897</td>
<td>Arp, Jean, 1887-1966</td>
<td>1920</td>
<td>Cover title. 23 p. 24 cm.</td>
<td></td>
<td></td>
<td>German</td>
<td>P. Steegemann</td>
</tr>
<tr>
<td>6</td>
<td>149965746</td>
<td>Aragon, 1897-</td>
<td>1921</td>
<td>Ed. originale, 197 p.; 19 cm.</td>
<td></td>
<td></td>
<td>French</td>
<td>Nouvelle Revue Française</td>
</tr>
<tr>
<td>7</td>
<td>148471181</td>
<td>Aragon, Louis</td>
<td>1922</td>
<td>Library has 95 p.; port.; 19 cm.</td>
<td></td>
<td></td>
<td>French</td>
<td>Editions de la Nouvelle Revue Française</td>
</tr>
<tr>
<td>8</td>
<td>148708261</td>
<td>Aragon, 1897-</td>
<td>1920</td>
<td>Title vignette, illus. 19 cm.</td>
<td></td>
<td></td>
<td>French</td>
<td>Au Sans pareil</td>
</tr>
<tr>
<td>9</td>
<td>149424094</td>
<td>Ball, Hugo, 1886-1927</td>
<td>1918</td>
<td>224 p.; 20 cm.</td>
<td></td>
<td></td>
<td>German</td>
<td>Erich Reiss</td>
</tr>
<tr>
<td>10</td>
<td>150559319</td>
<td>Ball, Hugo, 1886-1927</td>
<td>1919</td>
<td>Includes bibli 327 p.; 24 cm.</td>
<td></td>
<td></td>
<td>German</td>
<td>Freie Verlag</td>
</tr>
<tr>
<td>11</td>
<td>148451281</td>
<td>Doesburg, Theo van, 1883-1931</td>
<td>1921</td>
<td>31 p., xvii p. of plates: ill.; 23 cm.</td>
<td></td>
<td></td>
<td>French</td>
<td>De Sikkel; L. Rosenberg</td>
</tr>
</tbody>
</table>
More Data Manipulation

- Add semicolon between multiple elements
- Split extent into different fields
- If “ill.” add “Still image” to the Type field
- Standardize publisher
- Update names with authorized heading
Subject Data Manipulation

• Split subject headings, pulling out geographic and form subdivisions to separate fields
• Reformat geographic subjects hierarchically
• May add standard subject headings to entire collection (e.g. Dadaism; Dadaist literature)
• Add chronological subject based on decade of publication
### Subject Data Manipulation

<table>
<thead>
<tr>
<th>Place of Publication</th>
<th>Subject LCSH</th>
<th>Chronological Subject</th>
<th>Geographical Subject</th>
<th>Title</th>
<th>Alternative</th>
<th>isPartOf</th>
<th>Q</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>France -- Paris Dadaism; Dadaist literature</td>
<td>1910-1920</td>
<td></td>
<td>Larountala : polydrome en deux parties</td>
<td></td>
<td></td>
<td></td>
<td>Text: Still image</td>
</tr>
<tr>
<td>3</td>
<td>France -- Paris Dadaism; Dadaist literature</td>
<td>1910-1920</td>
<td></td>
<td>Tou碨vre : roman</td>
<td></td>
<td></td>
<td></td>
<td>Text: Still image</td>
</tr>
<tr>
<td>4</td>
<td>Germany -- Leipzig Dadaism; Dadaist literature; Painting, French</td>
<td>1910-1920</td>
<td></td>
<td>Noue franzosische Malerei</td>
<td></td>
<td></td>
<td></td>
<td>Text: Still image</td>
</tr>
<tr>
<td>5</td>
<td>Germany -- Hanover Dadaism; Dadaist literature</td>
<td>1920-1930</td>
<td></td>
<td>Die Wolkenpumpe.</td>
<td></td>
<td></td>
<td></td>
<td>Text</td>
</tr>
<tr>
<td>6</td>
<td>France -- Paris Dadaism; Dadaist literature</td>
<td>1920-1930</td>
<td></td>
<td>Anicet, ou, Le panorama</td>
<td></td>
<td></td>
<td></td>
<td>Text</td>
</tr>
<tr>
<td>7</td>
<td>France -- Paris Dadaism; Dadaist literature</td>
<td>1920-1930</td>
<td></td>
<td>Les aventures de T纪委maque</td>
<td></td>
<td></td>
<td></td>
<td>Text: Still image</td>
</tr>
<tr>
<td>8</td>
<td>France -- Paris Dadaism; Dadaist literature</td>
<td>1920-1930</td>
<td></td>
<td>Feu de joie</td>
<td></td>
<td></td>
<td></td>
<td>Text: Still image</td>
</tr>
<tr>
<td>9</td>
<td>Germany -- Berlin Dadaism; Dadaist literature</td>
<td>1910-1920</td>
<td></td>
<td>Flametti, oder, Vom Dandymsus der Armen : Roman</td>
<td></td>
<td></td>
<td></td>
<td>Text</td>
</tr>
<tr>
<td>10</td>
<td>Switzerland -- Bern Dadaism; Dadaist literature; National characteristics, German</td>
<td>1910-1920</td>
<td>Germany</td>
<td>Zur Kritik der deutschen Intelligenz</td>
<td></td>
<td></td>
<td></td>
<td>Text</td>
</tr>
<tr>
<td>11</td>
<td>Belgium -- Antwerp; France -- Paris Dadaism; Dadaist literature; Art, Art, Modern, Aesthetics</td>
<td>1920-1930</td>
<td></td>
<td>Classique, baroque, moderne</td>
<td></td>
<td></td>
<td></td>
<td>Text: Still image</td>
</tr>
</tbody>
</table>
The Problem of Using Old Metadata

• Out of date coding won’t convert well
  – Prior to 1993 an alternative title would be in a 740
    • 740 0^ |a Anicet.
    • <dc:relation>Anicet.</dc:relation>
  – Authorized entries may have changed
    • Aragon, 1897-
    • Aragon, 1897-1982
Missing Information

• Records lack place of publication.
  – Manually added this information (in hierarchical form)
  – Would be great if this could be included in export

• Add standard information for compound object (collection, contributing institution etc.)
Upload Records

• Save file as Tab-delimited text file
• Load collection to CDM like usual
Edit item

Edit metadata then click **save changes**. To remove this item from the collection, click **delete**. After making changes, you must rebuild the text index.

<table>
<thead>
<tr>
<th>Title</th>
<th>Lerouteau : polydrame en deux parties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Title</td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Albert-Briot, Pierre</td>
</tr>
<tr>
<td>Publisher</td>
<td>Editions &quot;Sic&quot;</td>
</tr>
<tr>
<td>Place of Publication</td>
<td>France -- Paris</td>
</tr>
<tr>
<td>Date Original</td>
<td>1919</td>
</tr>
<tr>
<td>Description</td>
<td>Ce livre a été tiré à 124 exemplaires, cent vingt sur papier d'arches à la forme numérotés de 4 à 123 et quatre sur chine dont trois numérotés 1, 2, 3, et un imprimé au nom de l'auteur--9. [3]. Library's copy is unnumbered and has author's holograph inscription: &quot;Le Lois de...&quot;</td>
</tr>
<tr>
<td>Table of Contents</td>
<td></td>
</tr>
</tbody>
</table>
Title: Larouncinga: polystyrene en deux parties
Creator: Albert-Birot, Pierre
Publisher: Editions "SIC"
Place of Publication: France -- Paris
Date Original: 1919
Description: Ce livre a été tiré à 124 exemplaires, contenant sur papier d'arches à la forme numérotée de 4 à 123 et quatre sur chine dont trois numérotées 1, 2, 3, et un imprimé au nom de l'auteur - P. "SIC". Library's copy is unnumbered, and has author's holograph inscription: "à Louis de Gonzaque-Frick [2 words illegible] personnages sous la direction de l'auteur Pierre Albert-Birot 30 Mai 1919."
Topical Subject (LCSH): Dadaism
Dadaist literature
Chronological Subject: 1910-1920
Type (DCMI Type): Text
Type (AAT): Books
Type (IMT): pdf
Digital Collection: Dada Digital Collection
Contributing Institution: University of Iowa Libraries, Special Collections Dept.
Rights Management: Educational use only; no other permissions given. U.S. and international copyright laws may protect this digital image. Commercial use or distribution of the image is not permitted without prior permission of the copyright holder.
Contact Information: For further information about the Dada Digital Archive, please contact the curator, Timothy Shipe: timothy-shipe@uiowa.edu
Number of Pages: 92: 148490109
Height (cm): 19
Digitization Specifications: Scanned at 300 dpi, 24-bit full color. Archival tiff images available.
Date Digital: 4/28/2008
Language: French
Aerial Photo Indices

• Materials currently cataloged in our local catalog by paraprofessional
• Records batch exported to OCLC
• Paraprofessional adds OCLC numbers to a spreadsheet
• Digital libraries staff notified when a decade is completed
• Txt file of OCLC numbers created
Import File Into Connexion

Select OCLC number

Click import, navigate to the file you want and click Open

Click Save and then Close
Process Batch in Connexion

If you have a large number of IDs, this step will take a while.

Check these boxes

Click OK
View Batch Reports

Look for any errors (invalid OCLC numbers)

Look for any blank search keys (duplicated OCLC numbers)

Note: you can correct problems in txt file and do this step again

Note save file numbers
Remember Connection Set-up!

- Under Tools—Options—Export tab

Click to change from MARC21 to Dublin Core

Create file of records

OCLC option includes OCLC number in metadata
Retrieve Records in Local Save File

Click OK to retrieve all records in save file (or limit search)
Select Records

Knowing save file numbers can help identify the records.
Export Records

This step can take some time

XML created as in previous example and requires the same type of manipulation
Delete Records in Save File

Identifies you are in your local save file

Delete unneeded records from local save file to make it easier to find records in the future
Manipulate and Load

• Convert QDC XML files to delimited
• Modify data as needed/desired
• Load records to QDM
Ignaz Pleyel Sheet Music

• Digitizing fully cataloged materials
• Records updated by music cataloger in Connexion to meet current standards
  – Master record corrected when possible
  – Otherwise changes made to local record
Awkward Title by Title Workflow

• Set export for MARC

Change destination and record characteristics for each title
Awkward workflow (cont.)

• And change settings back to QDC

Change destination and record characteristics for each title
Better Workflow

- Export records as MARC like usual
- Save records in local save file
- Change settings and export group of records
- Still requires individual catalogers make all the setting changes
- If save file is on server, Digital Library staff could export when needed
Conclusion

• Pros:
  – MARC records are improvement over recreating all the metadata
  – Catalogers are comfortable with MARC

• Cons:
  – Files require quite a bit of manipulation
  – Workflows not always smooth and depend on how Technical Services staff want to do work

• Looking forward to improved functionality for non-hosted sited
Thanks!

Questions?

Wendy Robertson
Electronic Resources Systems Librarian
Digital Library Services
The University of Iowa Libraries
wendy-robertson@uiowa.edu