Best Practices for CONTENTdm

CONTENTdm Metadata Working Group

Sheila Bair
Myung-Ja Han
OCLC CONTENTdm Metadata Working Group

- 34 CONTENTdm users
- Represent academic and public libraries, archives, and museums
- Goal: best practices for shareable metadata using the WorldCat Digital Collection Gateway
Need for guidelines

- Loss of richness and meaning when sharing DC records in aggregated environments
  Shreeves, Riley & Milewicz (2006)
  Jackson, Han, Groetsch, Mustafoff & Cole (2008)

- “One-to-one” rule and difficulty of mapping to MARC
  Caplan & Guenther (1996)
  Han, Cho, Cole & Jackson (2009)
  Beisler & Willis (2009)
Best practices draft 1.9

- 12 core elements and 4 recommended “as appropriate”
- Mapping and “shareability”
  - Guidelines for field content standardization and mapping to MARC
- Addenda
  - Treatment of compound objects with respect to OAI harvesting
  - Considerations for consortia using OAI harvesting tools
Best practices draft 1.9

- **Title**: prefer non-use of explanatory or qualifying symbols (e.g., brackets to indicate cataloger-supplied title)

- **Creator**: prefer non-use of ‘junk value’ (e.g., “Unknown”); however, it is appropriate to qualify named entities with “[role]”

- **Publisher**: prefer “digitized by” or other text prefix to qualify value
## Template for photographic collections

<table>
<thead>
<tr>
<th>Field Name</th>
<th>DC Map</th>
<th>Field Name</th>
<th>DC Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title</td>
<td>Type</td>
<td>Type</td>
</tr>
<tr>
<td>Subject-Names</td>
<td>Subject</td>
<td>Format</td>
<td>Format</td>
</tr>
<tr>
<td>Subject-Places</td>
<td>Subject</td>
<td>Identifier</td>
<td>Identifier</td>
</tr>
<tr>
<td>Subject-Topics</td>
<td>Subject</td>
<td>Source</td>
<td>Source</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td>Language</td>
<td>Language</td>
</tr>
<tr>
<td>Publisher</td>
<td>Publisher</td>
<td>Rights &amp; Usage</td>
<td>Rights</td>
</tr>
<tr>
<td>Contributors</td>
<td>Contributors</td>
<td>Date. Original</td>
<td>Date</td>
</tr>
</tbody>
</table>
## Template for archival collections

<table>
<thead>
<tr>
<th>Field Name</th>
<th>DC Map</th>
<th>Field Name</th>
<th>DC Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title</td>
<td>Identifier</td>
<td>Identifier</td>
</tr>
<tr>
<td>Title-Alternative</td>
<td>Title-Alternative</td>
<td>Source</td>
<td>Source</td>
</tr>
<tr>
<td>Creator</td>
<td>Creator</td>
<td>Language</td>
<td>Language</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td>Relation</td>
<td>Relation-Is Part Of</td>
</tr>
<tr>
<td>Abstract</td>
<td>Description-Abstract</td>
<td>Location</td>
<td>Coverage-Spatial</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>Description</td>
<td>Time period</td>
<td>Coverage-Temporal</td>
</tr>
<tr>
<td>Subject</td>
<td>Subject</td>
<td>Rights &amp; Usage</td>
<td>Rights</td>
</tr>
<tr>
<td>Publisher</td>
<td>Publisher</td>
<td>Audience</td>
<td>Audience</td>
</tr>
<tr>
<td>Contributors</td>
<td>Contributors</td>
<td>Relation-References</td>
<td>Relation-References</td>
</tr>
<tr>
<td>Date. Original</td>
<td>Date</td>
<td>Digitization</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specification</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Type</td>
<td>Format</td>
<td>Format</td>
</tr>
</tbody>
</table>
### Digital information seeker


<table>
<thead>
<tr>
<th>Study</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptions of libraries, 2005</td>
<td>• Search engines dominant place to begin</td>
</tr>
<tr>
<td></td>
<td>• Search engine as lifestyle fit</td>
</tr>
<tr>
<td></td>
<td>• Search engines are preferred over libraries</td>
</tr>
<tr>
<td>Sense-making, 2006</td>
<td>• Heavy reliance on Google and other web Information sources</td>
</tr>
<tr>
<td>Researchers and discovery services, 2006</td>
<td>• Common tools include general search engines, specialist search engines</td>
</tr>
<tr>
<td></td>
<td>• Google is often used for relatively simple tasks in conjunction with other sources</td>
</tr>
<tr>
<td>Researcher of the future, 2008</td>
<td>• Majority of BL visits were from search engine</td>
</tr>
<tr>
<td></td>
<td>• Prefer natural-language searching and trust Google to understand them</td>
</tr>
<tr>
<td></td>
<td>• 40% of school-age visits to the BL visits were via an image search</td>
</tr>
</tbody>
</table>
Researchers’ discovery/search strategies

- Discovery
  - Natural language, keyword

- Known searches
  - Names of people, places, events
  - Specific dates, years

“Users want to search names by keyword, search for subjects by browsing, and browse by keyword or name, too. When it comes to using descriptive metadata to discover archival materials and special collections, users want it all.”

Jennifer Schaffner (2009)
Catalogers often ‘miss the point,’ failing to apply key saleable descriptors

Catalogers are encouraged to develop their own local controlled vocabularies

Addition of natural language, keywords increase discovery
Adding keywords

- Subject 1: Architecture, domestic
- Subject 2: Dwellings
- Subject 3: Crane, Caroline Bartlett, 1858-1935
- Subject 4: Dormers
- Subject 5: Wooden-frame houses
- Subject-keywords: housing design; 1920s; women; children; family; families; award winning; awards
Results 1-10 of about 55 (.51 seconds)

1. Front exterior view with neighboring houses, Everyman's House
   - Downloadable 2-D image: Picture
   - Publisher: Digitized by Western Michigan University 1924?

2. Copyright Assignment for Everyman's House book
   - by Doubleday, Russell
   - eBook
   - Publisher: Digitized by Western Michigan University 1927-03-29

3. Sowder, Elmer Jesse
   - by Mary Palevsky
   - Internet resource
   - Language: English
   - Publisher: University of Nevada, Las Vegas Libraries
   - Database: OAIster

4. Kitchen interior with sink, window and cupboards, Everyman's House
   - Downloadable 2-D image: Picture
   - Publisher: Digitized by Western Michigan University 1924?
Known names and places

- Subject-LCSH 1: Sly, Eugene R., 1841-1922--Diaries
- Subject-LCSH 2: Andersonville Prison
- Subject-LCSH 3: Danville Prison
- Subject-LCSH 4: Florence Prison (S.C.)
- Subject-LCSH 5: Wirz, Henry, 1823?-1865
Results 1–9 of about 9 (.28 seconds)

Select All  Clear All  Save to: [New List]  Save

1. Plan of Andersonville Prison or "Camp Sumter," April, 1864.
   by Robert Knox Sneden
   Map: Manuscript Archival Material
   Language: English

2. Papers, 1853-1882 [manuscript].
   by Joseph Frederick 1832?-1876 Waring
   Internet resource
   Language: English
   Database: OAIster

3. Eugene R. Sly diary, 1864, United States Civil War.
   by Sly, Eugene R., 1841-1922
   eBook
   Publisher: 1864
Addendum:
Treatment of compound objects

- Compound objects: books, diaries, letters, etc.
- Bring out page-level “gems” in the object-level metadata
- Gateway does not export page-level data to WorldCat
Page-level subjects added

- Subject-LCSH 1: Ide, Alonzo C., 1842-1920--Diaries
- Subject-LCSH 2: Soldiers--Michigan--Diaries
- Subject-LCSH 3: Knoxville (Tenn.)--History--Siege, 1863
- Subject-LCSH 8: Longstreet, James, 1821-1904--Correspondence
- Subject-LCSH 9: McLaws, Lafayette, 1821-1897--Correspondence
Digital Collection Gateway

Map CONTENTdm metadata to MARC
- Collection coordinators can map elements
- Using Qualified Dublin Core is recommended

Ingest metadata into WorldCat

Increase access to collections to a broader user base

However...
CONTENTdm DC to MARC

- *Leader* information is sparse and hard to map
- *Subject* is indexed but not controlled
  - Creates a brief record

- Not exactly a cataloging record,
  - but for **expanded access**
Qualified Dublin Core

- Added refined elements (Qualifiers) to Simple Dublin Core
- Qualifiers can be modifiable by users/repositories
- No specific set of elements
Two broad classes of QDC

- **Element Refinement**
  - make the meaning of an element narrower or more specific

- **Encoding Scheme**
  - identify schemes that aid in the interpretation of an element value
  - schemes include controlled vocabularies and formal notations or parsing rules.

Using Dublin Core - Dublin Core Qualifiers
<http://dublincore.org/documents/usageguide/qualifiers.shtml>
Two broad classes of QDC

- **Element Refinement**
  - make the meaning of an element narrower or more specific

- **Encoding Scheme**
  - identify schemes that aid in the interpretation of an element value
  - schemes include controlled vocabularies and formal notations or parsing rules.
Two broad classes of QDC

- **Element Refinement**
  - make the meaning of an element narrower or more specific
  - Default CONTENTdm OAI metadata format

- **Encoding Scheme**
  - identify schemes that aid in the interpretation of an element value
  - schemes include controlled vocabularies and formal notations or parsing rules.
<table>
<thead>
<tr>
<th>ID Number</th>
<th>A149-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Frances Abington (nee Barton) as Aurelia in &quot;The Twin Rivals&quot;</td>
</tr>
<tr>
<td>Role</td>
<td>Aurelia</td>
</tr>
<tr>
<td>Play</td>
<td>The Twin Rivals</td>
</tr>
<tr>
<td>Subject</td>
<td>Abington, Mrs., (Frances Barton), 1737-1815</td>
</tr>
<tr>
<td>Actors</td>
<td></td>
</tr>
<tr>
<td>Costume</td>
<td></td>
</tr>
<tr>
<td>Theater--History--18th century--Pictorial works</td>
<td></td>
</tr>
<tr>
<td>Women in the theater</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Engravings</td>
</tr>
<tr>
<td>Dimensions</td>
<td>5 x 3 3/4 inches</td>
</tr>
<tr>
<td>Technique</td>
<td>Engraving</td>
</tr>
<tr>
<td>Description</td>
<td>Three-quarter-length. Wearing dress with ornaments and headdress with plumes, holding fan. With text from Diary of Henry Crabb Robinson: &quot;1811_June 16th_met the once celebrated Mrs. Abington_She herself laughed at her snub nose&quot;</td>
</tr>
<tr>
<td>Rights</td>
<td>These images are made available for educational use only and may not be used for any non-educational or commercial purpose. Approved educational uses include faculty research, teaching, and student projects. For additional information please contact <a href="mailto:dsd@library.uiuc.edu">dsd@library.uiuc.edu</a>. Please see <a href="http://images.library.uiuc.edu">http://images.library.uiuc.edu</a> if interested in obtaining higher resolution images for your research or a publication. Fees may apply.</td>
</tr>
<tr>
<td>Digital Collection</td>
<td>Portraits of Actors, 1720-1920, University of Illinois Library</td>
</tr>
<tr>
<td>Repository Collection</td>
<td>University of Illinois Theatrical Print Collection</td>
</tr>
</tbody>
</table>
Frances Abington (nee Barton) as Aurelia in "The Twin Rivals"

Abington, Mrs. (Frances Barton), 1737-1815; Actors; Costume; Theater--History--18th century--Pictorial works; Women in the theater

Engravings

5 x 3 3/4 inches

Engraving; Etching

Three-quarter-length. Wearing dress with ornaments and headdress with plumes, holding fan. With text from Diary of Henry Crabb Robinson: "1811_ June 16th ___met the once celebrated Mrs. Abington___She herself laughed at her snub nose"

These images are made available for educational use only and may not be used for any non-educational or commercial purpose. Approved educational uses include faculty research, teaching, and student projects. For additional information please contact dsd@library.unc.edu. Please see http://images.library.unc.edu if interested in obtaining higher resolution images for your research or a publication. Fees may apply.

Portraits of Actors, 1720-1920, University of Illinois Library

University of Illinois Theatrical Print Collection

http://images.library.unc.edu:8081/u?/actors,4
<table>
<thead>
<tr>
<th>Field name</th>
<th>DC map</th>
<th>Data type</th>
<th>Large</th>
<th>Search</th>
<th>Hide</th>
<th>Vocab</th>
<th>add field</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID Number</td>
<td>None</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Title</td>
<td>Title</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Date</td>
<td>Date-Created</td>
<td>Date</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Role</td>
<td>None</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>edit</td>
</tr>
<tr>
<td>Play</td>
<td>None</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>edit</td>
</tr>
<tr>
<td>Subject</td>
<td>Subject</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>edit</td>
</tr>
<tr>
<td>Type</td>
<td>Type</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>edit</td>
</tr>
<tr>
<td>Dimensions</td>
<td>Format-Extent</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Technique</td>
<td>Subject</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>edit</td>
</tr>
<tr>
<td>Creator</td>
<td>Creator</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>edit</td>
</tr>
<tr>
<td>Publisher</td>
<td>Publisher</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td>Text</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Rights</td>
<td>Rights</td>
<td>Text</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Digital Collection</td>
<td>Relation-Is Part Of</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
<tr>
<td>Repository Collection</td>
<td>Source</td>
<td>Text</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>edit</td>
</tr>
</tbody>
</table>
650 _0 Abington, Mrs., (Frances Barton), 1737-1815.
650 _0 Actors.
650 _0 Costume.
650 _0 Women in the theater.
Then what?

- Improve crosswalk from CONTENTdm OAI QDC to MARC?
- Create MARC records from the beginning?
- Defeat the original purpose of using Dublin Core
- It is too expensive!

*WorldCat increases access and discovery of resources!*
Using Dublin Core!

- Preserve the integrity of metadata
- Original metadata quality can be kept as is
- Simple and easy to use
Acknowledgements

- Members of CONTENTdm Metadata Working Group
- OCLC
  - Geri Ingram
  - Jason Lee
  - Taylor Surface
Thank you!

- Sheila Bair
  Western Michigan University
  sheila.bair@wmich.edu

- Myung-Ja Han
  University of Illinois
  mhan3@illinois.edu
References


## Best practices draft 1.9
### core elements

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Field Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Format</td>
</tr>
<tr>
<td>Creator</td>
<td>Type</td>
</tr>
<tr>
<td>Description</td>
<td>Contributor</td>
</tr>
<tr>
<td>Subject</td>
<td>Date</td>
</tr>
<tr>
<td>Identifier</td>
<td>Rights</td>
</tr>
<tr>
<td>Language</td>
<td>Publisher</td>
</tr>
<tr>
<td>Source</td>
<td>Relation-Is Part Of</td>
</tr>
<tr>
<td>Coverage-Spatial</td>
<td>Coverage-Temporal</td>
</tr>
</tbody>
</table>