Vamp It Up and Make It Better:
Auditing and Upgrading Metadata in the University of Houston Digital Library

Santi Thompson - Metadata Coordinator
Annie Wu - Head, Metadata and Bibliographic Services

2013 TCDL Conference
Austin, TX
UHDL Background

- Founded in 2009
- CONTENTdm
- 56 digital collections to date
- About 46,000+ images
Rationales for Metadata Upgrade

• Improve access of digital collection data in Summon
• Make our metadata interoperable with other systems
• Enhance discoverability of digital objects
• Improve display and retrieval of our digital metadata through our upcoming new UHDL interface

More Robust, Reliable and Retrievable Metadata!
Metadata Upgrade Project Overview

Metadata Upgrade

- Data Analysis Methodology
- New Data Dictionary
- Metadata Upgrade Strategies
- Recommendations
Methodology – Data Collection

- Focus group interviews
- Fields inspection and review
- Bench marking

image source: http://przedsiebiorczykrakow.pl
Recommendations – Focus Groups

Standardize Data
- Controlled vocabularies
- Mapping to DC elements
- Collection names

Integrate Data
- Facets in Summon Discovery tool and in new UHDL interface

image source: http://www.eventbrite.com
Recommendations: Field Inspection and Benchmarking

- Correct inaccurate, incomplete, or missing data
- Add new metadata fields to better describe and preserve items
- Implement templates for titles for consistency
- Standardize mapping to DC elements

image source: [http://portilloa.wordpress.com](http://portilloa.wordpress.com)
Upgrade Strategies – Phase One

- Standardize and update field labels in the digital library
- Standardize and update mapping used to link fields to DC elements
- Add new fields to the digital library
- Link content in the digital library with the library catalog or finding aid, and vice-versa
- Integrate CONTENTdm with Summon to facilitate discoverability
Upgrade Strategies – Phase Two

Collection Level Metadata Upgrade:

- Standardize and update collection names for archival and digital content
- Add and/or edit collection level fields
Upgrade Strategies – Phase Three

Item Level Metadata Revision:

- Apply templates to titles
- Add and/or edit item level fields
  - add ISO date to Date Digital
  - add type terms to Type DCMI field
  - standardize the control vocabulary terms to their source
UHDL Metadata Dictionary

Structure and Purpose
- Provides definitions, examples, and formatting standards for specific metadata fields
- Identifies who is responsible for capturing information
- Defines how fields map to Dublin Core elements
- Will serve as a resource for cross-walking between metadata standards

Types of Metadata:
- Descriptive Metadata
- Technical Metadata
- Preservation Metadata
Descriptive Metadata

- Describes the intellectual content of the object
- Example fields include:
  - Title
  - Creator
  - Description
  - Subjects
- Standardizes data formats, field labels, controlled vocabularies, and mapping to DC elements

image source: [http://ils.indiana.edu](http://ils.indiana.edu)
# UHDL Metadata Dictionary, Descriptive Metadata

<table>
<thead>
<tr>
<th>Field Label</th>
<th>Field Information</th>
<th>Field Content</th>
<th>Field Settings for CONTENTdm</th>
<th>Crosswalk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Definition: &quot;The name given to the resource. Typically, a title will be a name by which the resource is formally known.&quot; (CDD). Note: values of 2 or more title templates should be discussed during project plan meeting. Note: value of resource in other formats (e.g., name of person, corporate body, object, activity, event, geographical area and date)</td>
<td>Format: Letter: &quot;Letter from x to y (if needed: regarding z)?&quot; Newsletter: &quot;title of newsletter and issue date?&quot; Book: Title taken from the catalog record Other formats: Devise a brief descriptive title that combines the nature of the resource (e.g., map, diary, advertisement) and its subject (e.g., names of persons, corporate bodies, objects, activities, events, geographical area and dates)</td>
<td>DC Map Title: Data Type: Text Search: Yes</td>
<td>MARC: 245 a 240.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Required: Yes</td>
<td>Vocals No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Required in UODL Field Template: Yes</td>
<td></td>
</tr>
<tr>
<td>Alternative Title</td>
<td>Definition: An alternative name for the resource. Note: Use this field for primarily for books or other items that have alternative title(s). This information will be found in the catalog record.</td>
<td>Format: Multiple entries should be separated with &quot;,&quot;, (space). Example: &quot;Ameriope para quarta, sine, insignis &amp; admodum habet de reperita primam Occidentales Indiæ à Christophoro Columbo anno M. CCCXCO I scruta ab Hieronymo Becrazion ... qui erit oriëntis XIII. versum dicitur;&quot; non observant: ad una singula feri capita, non corpora cuncta in eadem aequali tamen etiam gentium lobatiae; accession quaedamARBENIUM tarnum tabulis chronographica; omnia decorasitas lingua in sae clausa expressar: Theodore de Bry Leonense, civis Francifortensi.&quot;</td>
<td>DC Map Title:Alternative Data Type: Text Search: Yes</td>
<td>MARC: 246</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Required: No</td>
<td>Vocals No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Required in UODL Field Template: No</td>
<td></td>
</tr>
</tbody>
</table>
Technical Metadata

- “Metadata that documents the creation and the digital characteristics of the files” (DCMI).

- Example fields include:
  - Capture Tool
  - Digital Creator
  - Access file information

image source: http://melbel.hubpages.com
Preservation Metadata

- Information needed to make resource render-able in the future
- Example fields include:
  - Preservation File name
  - Preservation File Checksum (MD5)
  - Preservation File Notes

image source: http://blog.contractexpress.com
### UHDL Metadata Dictionary, Technical and Preservation Metadata

<table>
<thead>
<tr>
<th>Field Label</th>
<th>Field Information</th>
<th>Field Content</th>
<th>Field Settings for CONTENTdm</th>
<th>Crosswalk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date Digital (ISO 8801)</td>
<td><strong>Definition:</strong> Date that the object was digitized, expressed in ISO 8801 format (YYYY-MM-DD).&lt;br&gt;<strong>Responsible for populating:</strong> DS</td>
<td><strong>Format:</strong> ISO 8801&lt;br&gt;&lt;br&gt;<strong>Examples:</strong> 1078-06-19 2011-06 [10007]</td>
<td>DC Maps: None&lt;br&gt;Data Typem: Text&lt;br&gt;Search: Yes&lt;br&gt;Hides: No&lt;br&gt;Required: No&lt;br&gt;Vocab: No</td>
<td>Required in UHDL Field Template: Yes</td>
</tr>
<tr>
<td>Capture Tool</td>
<td><strong>Definition:</strong> The hardware and/or software used to create a digital file from an analog object.&lt;br&gt;Note: Use defined values from the Technical and Preservation Metadata Template for this field.&lt;br&gt;<strong>Responsible for populating:</strong> DS</td>
<td><strong>Format:</strong> Local standardized language from the Technical and Preservation Metadata Template&lt;br&gt;&lt;br&gt;<strong>Examples:</strong>&lt;br&gt;- Digitization equipment: Epson V310&lt;br&gt;- Camera: 4K CCD camera, 22,000 pixels, Light Source: White LEDs, classified to IC 000225-1, cell 1, no IRV emission.&lt;br&gt;- Digitization equipment: Epson V310&lt;br&gt;- Camera: CCD camera, Light Source: CCD camera, 22,000 pixels, IRV-free cold light lamps, approx. 3,000 Lux.</td>
<td>DC Maps: None&lt;br&gt;Data Typem: Text&lt;br&gt;Search: Yes&lt;br&gt;Hides: No&lt;br&gt;Required: No&lt;br&gt;Vocab: No</td>
<td>Required in UHDL Field Template: Yes</td>
</tr>
<tr>
<td>Digital Creator</td>
<td><strong>Definition:</strong> The individual responsible for creating the digital object.&lt;br&gt;<strong>Responsible for populating:</strong> DS</td>
<td><strong>Examples:</strong> Digital Services Student, Technician&lt;br&gt;- Sandi Thompson&lt;br&gt;- Apex DeVargas</td>
<td>DC Maps: None&lt;br&gt;Data Typem: Text&lt;br&gt;Search: Yes&lt;br&gt;Hides: No&lt;br&gt;Required: No&lt;br&gt;Vocab: No</td>
<td>Required in UHDL Field Template: Yes</td>
</tr>
</tbody>
</table>
Upgrade Tools: Open Refine

- Data Facets and Comparisons
- Data Manipulation
- Data Reconciliation

image source: http://downloadsquad.switched.com
Upgrade Tools: Open Refine

### Google Refine: mtcexportexample.txt

<table>
<thead>
<tr>
<th>#</th>
<th>Title</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Letter from Minnie Fisher</td>
<td>Concerning two ways in which to increase food supplies, forming garden plots and the elimination of waste.</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Letter describing the efforts needed to plant a garden for every home. It encourages women to undertake regular farming on a small scale by renting garden plots. It also encourages women to employ thrift in all measures so as to emphasize the virtues of economy and to suggest methods of economizing.</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Food Supply Efforts Letter</td>
<td>Letter from Minnie Fisher to follow members of the Texas Equal Suffrage Association concerning two ways in which to increase food supplies, forming garden plots and the elimination of waste.</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Letter from the National Woman's Liberty League Committee explaining in detail the results expected to occur.</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
New Workflow: Tab Delimited Files

- Modify collection-level metadata more efficiently
- Store data in a non-proprietary format that can be imported into other metadata programs
- Preserve metadata locally in the department
Upgrade Documentation: Wiki

- Lists tasks in each phase
- Assigns responsible person to complete task
- Tracks tasks for transparency
- Preserves decision-making process for the duration of the project
Barriers and Challenges

- Aligning upgrade workflow and output with requirements and capabilities of CONTENTdm
- Juggling parallel process of metadata upgrade and creation
- Seeking feedback and building consensus
- Balancing need to be transparent and need to do work
- Synthesizing and incorporating multiple (and, sometimes, conflicting) perspectives
Collaboration

- Project driven by shared perspectives
- Collaborative effort to increase DL access points to Summon
- Shared responsibilities of metadata upgrade
- Joint new workflow between MBS and DS

Image resource: http://thevarguy.com
Benefits

- Best Practices
- Data interoperable
- Enhance retrieval of digital data
- New and better standard for future metadata creation

Image source: http://123rf.com
Santi Thompson
sathompson3@uh.edu

Annie Wu
awu@uh.edu