Introduction to Metadata Analysis Using the DLF Metadata Framework

Kate Flynn, Chicago Collections; University of Illinois at Chicago

The Basics

- Baseline
- Glossary: http://bit.ly/2gIzY6W
- Types of metadata
 - Descriptive metadata
 - Technical metadata
 - Preservation metadata
- What people might be working with when assessing
 - Databases
 - Character delimited value formats
 - \circ XML
 - JSON
 - Spreadsheets (XLS, XSLX)

The Basics

Metadata Assessment Framework

http://bit.ly/MdFramework

- A high-level framework for assessing descriptive metadata in digital collections
- Many thanks to:
 - Sara Rubinow
 - Maggie Dickson
 - Kathryn Gronsbell
 - Anna Neatrour
 - Bethany Scott

Migrations or Data Sharing

- Migrations
 - Before or after (or both!)
- Data sharing
- Aggregation (exposing data)
 - Shareability
 - Interoperability

- Standards
- Application profile or guidelines
- Schema
- Other pre-defined structures

Validation

Targeted Enhancement

- What change would have the most impact?
- Enhancement for a specific user community
- To address specific problems
- Consider resources available

 Lets users (internal & external) know what is expected for metadata

Profile Generation

Standards & Best Practices

- Staff changeovers may mean internal knowledge loss
- Systems & contexts can change over time
- These can be internally generated or from external sources

- Exhibits
- Text analysis
- Other digital humanities use

Handling Emerging
Object Types

- Usability
- Staff changeovers and internal knowledge
- External access and aggregation

Impact of Metadata Work

Building a scope + Parameters

- Prioritizing assessment
 - Baseline compliance
 - Vocabulary compliance
 - Repository migration
 - Usability

Building a scope + Parameters

- Use practices and policies for efficiency & effectiveness
 - Guidelines
 - Data dictionaries
 - Application profiles

Building a scope + Parameters

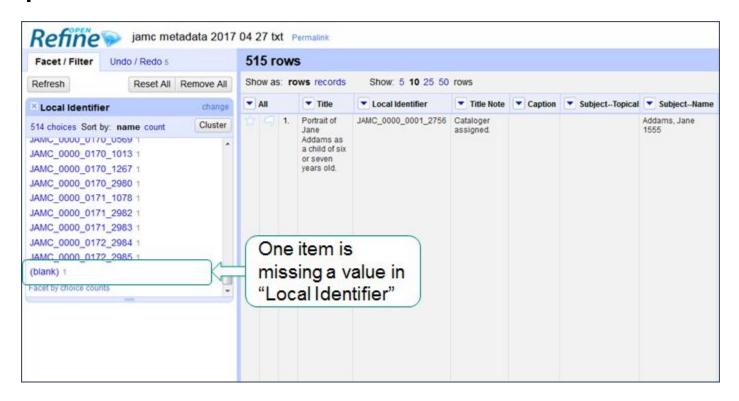
Keep it simple

- Use an approach that allows you to meet your objectives
- Having multiple phases/steps can help

The element/property/attribute is present

Completeness

Completeness



Completeness

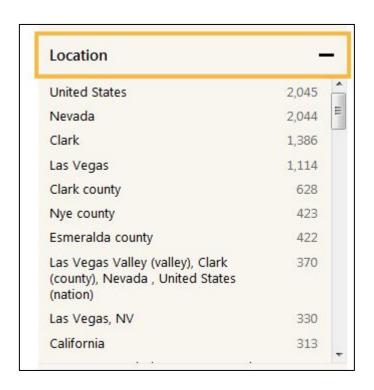
Description: The first paragraph should provide information on when and where the interview took place. The second paragraph should talk about the major topics covered over the course of the interview. The third paragraph is optional and should be a short biography of the interviewee.

Description: This is an audio interview with Jane Smith at her home in Niles, Illinois. It took place on August 7, 2011.

Values are consistent within your domain. Semantic and structural values and elements are represented in a consistent manner across records

Consistency

Consistency



Information is correct and factual both semantically and syntactically

Accuracy

Accuracy



Example

Project Location: Beverly (Chicago, III.)

Project Location: Loop (Chicago, Ill.)

Metadata can be read by a human and/or a machine

Accessibility

Accessibility

Names: Eacutecole nationale supeacuterieure des beaux-arts (France)

Accessibility

HTML5 Entity Names by Alphabet - E

Character	Entity Name	Hex	Dec	
É	Eacute	000C9	201	Try it
é	eacute	000E9	233	Try it

https://www.w3schools.com/charsets/ref html entities e.asp

Accessibility

Names: École nationale supérieure des beaux-arts (France)

Conformance to Expectations

Values adhere to the expectations of your user communities (both internal and external)

Example

- Consortium requirements
- Aggregator requirements
- Best practices

Conformance to Expectations



Mountain West Digital Library
Dublin Core
Application Profile

Version 2.0 July 20, 2011

I. Best Practices for All Fields

Use of semicolons

Data in one field may be combined with data from another field during the harvesting/aggregating process. To retain integrity of data, put a semicolon after the last word in each field. Likewise, separate multiple entries within a single field by inserting a semicolon and a space between each two entries.

MARC mapping

MARC tags are provided in each table to clarify the type of data contained in a field as well as to facilitate mapping data from Dublin Core records to MARC records.

Requirement status

- Required: Field must be included in record. There are eight required fields:
 - o date
 - o description

1

MWDL Dublin Core Application Profile

Version 2.0 (July 20, 2011)

- o format
- o <u>identifier</u>
- o rights
- subject
- o <u>title</u>
- o type

Timeliness

When the resource changes, the metadata is updated accordingly. When additional metadata becomes available or when metadata standards change, the metadata associated with the resource also changes.

Timeliness

Nature conservation Find in WorldCat

USED FOR:

Conservation of nature

Nature protection

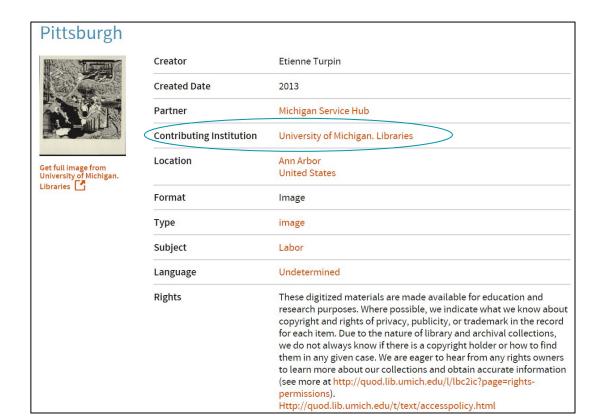
Protection of nature

http://experimental.worldcat.org/fast/1034632/

You can track metadata transformations back to the original form. You have information about the source of the original metadata.

Provenance

Provenance



Documenting assessment

Why?

- Advocacy
- Demystification
- Sustainability
- Compliance

How / Potential outputs

- Best Practices / Guidelines
- Clean-up plans
- Application profile
- Reports
- Other

Example Metadata Guidelines

Empire State Digital Network Best Practices & Guidelines:

http://bit.ly/edsnmap

Date

Field Name	Date		
Description (from DPLA)	Date value as supplied by data provider		
ESDN MODS Element	<pre><origininfo><datecreated [point="start"]="" keydate="yes"></datecreated></origininfo></pre>		
Vocabulary/Encoding Recommendations	W3CDTF EDTF		
Notes and Best Practices	Whenever possible, do not use any punctuation in the date field beyond dashes. For DPLA Timeline, DPLA will use date provided in ESDN MODS with attribute keyDate="yes" Use for the date of the original material, not the date the material was digitized. Dates of digitization should be in discrete fields or consistently indicated so that ESDN can ignore as required as part of transform for DPLA Be consistent in how you are indicating "questionable," "approximate," and "inferred" dates in your dataset		
Example Values	"July 4, 1776" → 1776-07-04 "ca. 1922" → 1922 "November 1956" → 1956-11		

An example clean-up plan

- Edit the following fields
 - Local Identifier should be Identifier
- Delete the following fields
 - □ Genre
- Add the following fields
 - Creator
 - Caption
- Convert values in Subject from LCSH to FAST
- □ Transcribe information written on the image into the new Caption field

Resources

DLF Assessment Interest Group Metadata Working Group Toolkit http://bit.ly/mdassessment

Metadata Application Profile Clearinghouse

https://tinyurl.com/MAPclearinghouse

Glossary of Terms

http://bit.ly/2qIzY6W

Metadata Assessment Framework

http://bit.ly/MdFramework