

Introduction to Metadata Analysis Using the DLF Metadata Framework

Kate Flynn, Chicago Collections, University of
Illinois, Chicago

Hannah E. V. Stitzlein, University of Illinois,
Urbana-Champaign

The Basics

- Baseline
- Glossary: <http://bit.ly/2glzY6W>
- Types of metadata
 - Descriptive metadata
 - Technical metadata
 - Preservation metadata
- What people might be working with when assessing
 - Databases
 - Character delimited value formats
 - XML
 - JSON
 - Spreadsheets (XLS, XSLX)

The Basics

Metadata Assessment Framework

<http://dlfmetadataassessment.github.io/Framework>

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

Why do we assess?



Why do we assess?

Migrations or Data Sharing

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

Why do we assess?

Validation

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

Why do we assess?

Targeted Enhancement

- What change would have the most impact?
- Consider resources available

Why do we assess?

Profile Generation

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

Why do we assess?

Standards & Best Practices

- Staff changeovers may mean internal knowledge loss
- Systems & contexts can change over time

Why do we assess?

Handling Emerging Object
Types

- Digital humanity use

Why do we assess?

Impact of Metadata Work

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

Documenting assessment

Why?

- Advocacy
- Demystification
- Sustainability
- Compliance

How / Potential outputs

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

What are you assessing?



What are you assessing?

Building a scope +
Parameters

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

What are you assessing?

Building a scope +
Parameters

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

What are you assessing?

Building a scope +
Parameters

- Keep it simple
 - Use an approach that allows you to meet your objectives

What measures do we use to assess?



What measures do we use to assess?

Completeness

The element/property/attribute is present

Completeness

Headlight: Sights and Scenes Along the Chicago and Northwestern Railroad

[View Description](#)

[Page Flip View](#) [Download](#) [Print](#)



Description

Title	Headlight: Sights and Scenes Along the Chicago and Northwestern Railroad
Subject	Chicago and Northwestern Railroad Company--History; Woodstock, Ill. -- History
Description	The Headlight publications were a promotional tool developed by the Chicago & Northwestern Railroad to highlight the towns along their railway. This publication features Woodstock Illinois.
Creator	Headlight Engraving Company
Publisher	Chicago & Northwestern Railroad
Date	1898
Contributing Institution	
Collection Name	

Missing
rights, type
fields

What measures do we use to assess?

Consistency

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

Example

- + Controlled vocabularies
- + Consistent field use

Consistency

Field Value	Count
text	2070
image	18

Collection uses consistent terms

DPLA OAI Aggregation Tools 1.0 developed by the [North Carolina Digital Heritage Center](#) - Revised and implemented by [CARLI](#)

What measures do we use to assess?

Accuracy

Information is correct and factual both semantically and syntactically

Accuracy



Example

~~Project Location: Beverly (Chicago, Ill.)~~

Project Location: Loop (Chicago, Ill.)

What measures do we use to assess?

Accessibility

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

Accessibility

Type Photograph. Horizontal; Image

Type Image

Type	
image	210521
text	94711
sound	1388
moving image	747
physical object	244

What measures do we use to assess?

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:
 - date
 - description

1

MWDL Dublin Core Application Profile

Version 2.0 (July 20, 2011)

- format
- identifier
- rights
- subject
- title
- type

What measures do we use to assess?

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

TITLE: Evanston's Central Business District

CREATOR: Perkins & Will; Brubaker, C. William

Before Metadata
Guidelines update

Title	Evanston's Central Business District
-------	--------------------------------------

Creator	Brubaker, C. William
---------	----------------------

Contributor	Perkins & Will
-------------	----------------

After Metadata
Guidelines update

What measures do we use to assess?

Provenance

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

Provenance

Pittsburgh



Get full image from
University of Michigan.
Libraries 

Creator	Etienne Turpin
Created Date	2013
Partner	Michigan Service Hub
Contributing Institution	University of Michigan. Libraries
Location	Ann Arbor United States
Format	Image
Type	image
Subject	Labor
Language	Undetermined
Rights	<p>These digitized materials are made available for education and research purposes. Where possible, we indicate what we know about copyright and rights of privacy, publicity, or trademark in the record for each item. Due to the nature of library and archival collections, we do not always know if there is a copyright holder or how to find them in any given case. We are eager to hear from any rights owners to learn more about our collections and obtain accurate information (see more at http://quod.lib.umich.edu/l/lbc2ic?page=rights-permissions).</p> <p>Http://quod.lib.umich.edu/t/text/accesspolicy.html</p>

Resources

DLF Assessment Interest Group Metadata Working Group Toolkit

<http://dlfmetadataassessment.github.io>

Metadata Application Profile Clearinghouse

<https://dlfmetadataassessment.github.io/MetadataSpecsClearinghouse/>

Glossary of Terms (draft)

https://docs.google.com/document/d/1ejPUZs0Qqz_HHx7gWPIUEuP8QdgOEXZRop7bEdjVgq4/edit?usp=sharing

Metadata Assessment Framework

<http://dlfmetadataassessment.github.io/Framework>