Learning Module 1. Metadata Basics: Purpose and Structure

Dr. Oksana L. Zavalina
University of North Texas
Oksana.Zavalina@unt.edu

Week 1 Presentation
Welcome and course overview

1. Welcome!

2. Online course with weekly videorecorded instructor presentations

3. Combination of discussion of readings and practical hands-on exercises to meet student learning outcomes

4. Course Syllabus & Schedule document is the roadmap:
   - Readings schedule
   - Activities schedule
   - Assignment schedule
     • details in Assignments section of the course website
What is Metadata?

- Data about data

- “Structured data about an object that supports functions associated with the […] object” (Greenberg, 2005):

  - **Functions = User tasks**
  - **Find** information objects
  - **Identify** information objects
  - **Select** information objects
  - **Obtain** information objects

(FRBR, 1998, 2008)
Types of information systems:

The ones that use **unstructured** (or less structured) metadata

- search engines
- automatically generated minimum metadata (index terms) by web crawlers

The ones that operate with **structured** metadata

- cultural heritage institutions
- currently largely manually generated metadata
Metadata used in various cultural heritage institutions

- Libraries
  - Public
  - Academic
  - Special
  - National
  - Digital

- Archives
- Museums
- Historical Societies
- ...

Dr. Oksana L. Zavalina, © 2015-2016
Metadata created for variety of user communities

- General audience (anyone)
  - K-12 teachers
  - K-12 students
  - College faculty
  - College students
  - Graduate faculty
  - Graduate students
  - Archivists
  - Art historians
  - Artists
  - Health providers
  - Humanities researchers
  - Genealogists
  - Engineers
  - Musicians
  - Social scientists
  - Scientists
  - Teachers
Metadata created for variety of user communities

- Experts
- Novices
- English speakers
- Global multilingual community
Metadata serves varying user needs

General user tasks (apply to everyone):

- Find
- Identify
- Obtain
- Select

Different needs depending on specific task at hand:

- Deciding where to go on vacation
- Exam in high school class
- Term paper in undergraduate or graduate course
- Doctoral dissertation...

Dr. Oksana L. Zavalina, © 2015-2016
Metadata created to represent a variety of information objects

- Print & Non-print
- Text & Non-text
- Analog & Digital

- Books, ebooks, audiobooks
- Journal runs, journal issues and journal articles
- Newspapers
- Music:
  - folk songs recordings
  - instrumental music recordings
  - notated music...
- Images:
  - Photographs, slides, negatives
  - Posters, drawings, paintings
- ...

Dr. Oksana L. Zavalina, © 2015-2016
Metadata created to represent a variety of information objects

- Print & Non-print
- Text & Non-text
- Analog & Digital

- letters, notebooks, diaries
- lesson plans
- games
- movies
- TV programs/series/episodes
- blogs, websites
- datasets
- 3D objects
- herbaria
- maps...

Dr. Oksana L. Zavalina, © 2015-2016
Important attributes of information objects differ among user communities

Some examples:

<table>
<thead>
<tr>
<th>User community</th>
<th>Important information object attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historians</td>
<td>time period covered by object</td>
</tr>
<tr>
<td>Genealogists</td>
<td>family names</td>
</tr>
<tr>
<td>Musicians</td>
<td>key, instruments, arrangement</td>
</tr>
<tr>
<td>Artists, art historians</td>
<td>style, medium</td>
</tr>
<tr>
<td>Teachers</td>
<td>learning standard, grade level</td>
</tr>
<tr>
<td>Health providers</td>
<td>diagnosis, test, procedure</td>
</tr>
<tr>
<td>Researchers</td>
<td>sampling procedure, research question</td>
</tr>
<tr>
<td>Natural scientists</td>
<td>species, class, date collected</td>
</tr>
<tr>
<td>Photographers</td>
<td>aperture, focal length</td>
</tr>
</tbody>
</table>

Dr. Oksana L. Zavalina, © 2015-2016
Result: multiple metadata schemes, standards, etc.

- To accommodate various information objects
- To meet varying user needs.

http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards.pdf
Of these many metadata schemes, in this course we will learn to create metadata records in...

- **Metadata schemes of particular importance:**
  - Dublin Core
  - MARC XML
  - Metadata Object Description Scheme (MODS)

- **Why are they important?**
  - digitization efforts
  - institutional repositories
  - job descriptions for info professionals
  - Semantic Web and Linked Data
<table>
<thead>
<tr>
<th>Type</th>
<th>Definition/Function</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Descriptive | Metadata used to identify and describe collections and related information resources | • Cataloging records  
• Finding aids  
• Differentiations between versions  
• Specialized indexes  
• Curatorial information  
• Hyperlinked relationships between resources  
• Annotations by creators and users |
| Administrative | Metadata used in managing and administering collections and information resources | • Acquisition information  
• Rights and reproduction tracking  
• Documentation of legal access requirements  
• Location information  
• Selection criteria for digitization |
| Preservation | Metadata related to the preservation management of collections and information resources | • Documentation of physical condition of resources  
• Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration  
• Documentation of any changes occurring during digitization or preservation |
| Technical   | Metadata related to how a system functions or metadata behaves                      | • Hardware and software documentation  
• Technical digitization information, e.g., formats, compression ratios, scaling routines  
• Tracking of system response times  
• Authentication and security data, e.g., encryption keys, passwords |
| Use         | Metadata related to the level and type of use of collections and information resources | • Circulation records  
• Physical and digital exhibition records  
• Use and user tracking  
• Content reuse & multiversioning info  
• Search logs  
• Rights metadata |

Table adapted from Gilliland (2008).
To connect users to information, information organizers (e.g., metadata librarians) create and interact with various representations of:

1. users’ information **needs** (represented through **queries**)

2. information **objects** (represented through **metadata**)

3. **relationships** among information objects (represented through **metadata**)

Dr. Oksana L. Zavalina, © 2015-2016
Focus on meeting user needs:
Cutters’ objectives of the library catalog (1896)

Enable a person to **find** a book when one of the following is known:
1. The author
2. The title
3. The subject.

**Show** what works the library has:
1. By a given author
2. On a given subject
3. In a given kind of literature.

**Assist in the choice** of a book:
1. As to the edition (bibliographically)
2. As to its character (literary or topical).
Focus on meeting user needs: FRBR user tasks (1998; 2009)

- **Find:** Discover if information object exists by searching one or more attributes
- **Identify:** Examine retrieved records to determine the objects that meet user’s search request
- **Select:** Examine retrieved records for those that meet other user needs/requirements
- **Obtain:** Use data in retrieved records to gain physical access to the described object

Dr. Oksana L. Zavalina, © 2015-2016
Steps in representation of information objects

1. Identify **attributes** of information objects
2. Choose those that are **important to users**
   - **Name** of object?
   - **Creator**?
   - **Date of creation**?
   - **Subject**?
   - **Location**?
   - **Color**?
   - **Size**?
   - ... 
3. Think how users will use this information (user tasks)

Dr. Oksana L. Zavalina, © 2015-2016
Representation of information object

represented by

Attributes related to its

Name

Attributes related to its

Creator

Attributes related to its

Content

Dr. Oksana L. Zavalina, © 2015-2016
Representation of relationships among information objects

**Name (Title) attribute**
- *War and peace [electronic resource]*
  - *a novel by Leo Tolstoy. A new and complete translation by Constance Garnett.*

**Creator attribute**
- *Tolstoy, Leo, graf*
- *Russia - History - Alexander I, 1801-1825*
- *Electronic books.*

**Content (Subject) attribute**
- *HathiTrust Digital Library*
- *No Tags, Be the first to tag this record!*

**Additional Information**
- *Dr. Oksana L. Zavalina, © 2015-2016*
Metadata Record Structure: Basics

<table>
<thead>
<tr>
<th>Field</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Bondarchuk</td>
</tr>
<tr>
<td>Title</td>
<td>War and peace</td>
</tr>
<tr>
<td>Subject</td>
<td>Records</td>
</tr>
<tr>
<td>Author</td>
<td>Tolstoy</td>
</tr>
<tr>
<td>Title</td>
<td>War and peace</td>
</tr>
<tr>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Jacobs</td>
</tr>
<tr>
<td>Title</td>
<td>Science dictionary</td>
</tr>
<tr>
<td>Subject</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Wilson</td>
</tr>
<tr>
<td>Title</td>
<td>College Physics</td>
</tr>
<tr>
<td>Subject</td>
<td>Physics</td>
</tr>
</tbody>
</table>

Database File

Records

Fields

Values
Metadata Scheme: 4 Major Components

- **Elements** (represented by fields in database records)
  - e.g., *Title, Author, Subject*, etc. – represent object’s attributes

- **Structure**
  - order of elements in the record
  - which elements are required, repeatable, etc.
  - hierarchy of elements

- **Semantics** = meaning of elements
  - e.g., “*Author* is the person or institution responsible for creation of information object”

- **Syntax**
  - how to input data values in metadata elements (e.g., punctuation, capitalization, order of words)
  - how to encode the elements (e.g., tags, indicators, etc.)
Metadata Record Structure: Example

**Author:** Taylor, Arlene G., 1941- Joudrey, Daniel N.

**Title:** The organization of information.

**Edition:** 3rd ed.

**Imprint:** Westport, Conn. : Libraries Unlimited, 2009.

**Physical Description:** xxvi, 512 p. ; 27 cm.

**Series:** Library and information science text series

**Note:** Includes bibliographical references (p. 479-498) and index.

**ISBN:** 978591587002 (pbk. : alk. paper)

**Subject:** Information organization. Metadata.

Dr. Oksana L. Zavalina, © 2015-2016
• **Define** metadata elements

• Indicate the **kind of data** that should be associated with the element

• Help **distinguish** one element from another
  – Should be unambiguous

• Support **interoperability** (mapping between metadata schemes)
## Semantics: examples

**Term Name:** creator

<table>
<thead>
<tr>
<th>URI</th>
<th><a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Label:</strong> Creator</td>
<td></td>
</tr>
<tr>
<td><strong>Definition:</strong> An entity primarily responsible for making the resource.</td>
<td></td>
</tr>
<tr>
<td><strong>Comment:</strong> Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.</td>
<td></td>
</tr>
</tbody>
</table>

### 4.1. Creator Description Core

**DEFINITION**

The name, brief biographical information, and roles (if necessary) of the named or anonymous individuals or corporate bodies responsible for the design, production, manufacture, or alteration of the work, presented in a syntax suitable for display to the end-user and including any necessary indications of uncertainty, ambiguity, and nuance. If there is no known creator, make a reference to the presumed culture or nationality of the unknown creator.
Syntax

The way to encode and structure the metadata elements for machine processing and exchange purposes

Some (simple) examples:
- Word order
  (e.g. for names)
- Abbreviation
- Capitalization
- Punctuation

Input rules

Author: Taylor, Arlene G., 1941-
Joudrey, Daniel N.

Title: The organization of information.

Edition: 3rd ed.


Physical Description: xxvi, 512 p. ; 27 cm.

Series: Library and information science text series

Note: Includes bibliographical references (p. 479-498) and index.


Subject: Information organization. Metadata.
Syntax: MARC & MARC XML example

100 1_ $a Taylor, Arlene G., $c 1941-
700 1_ $a Joudrey, Daniel N.
245 10 $a The organization of information.
250 __ $a 3rd ed.
300 __ $a xxvi, 512 p. ; $c 27 cm.
490 1_ $a Library and information science text series
504 __ $a Includes bibliographical references (p. 479-498) and index.
020 __ $a 978591587002 (pbk. : alk. paper)
650 _0 $a Information organization.
650 _0 $a Metadata.
Data side of metadata records

• Metadata creator creates or derives data from the information object to associate with individual metadata elements

• Key questions:
  • what is being described?
  • where to find the data?
  • what data go with which element?
  • what is the form of the data to enter into the metadata element?
What is being described?
(a.k.a. entity level)

VS.

Metadata for Digital Collections

VS.

Embedded metadata: friend or foe to our digital collections?

Abigail L. Dariger
San José State University, School of Library & Information Science
San José, California, United States

Library Student Journal, January 2011
Abstract

This paper examines the emerging trend of embedded metadata in regards to the successful preservation of digital collections, particularly those comprised of non-textual works. Since the long-term effects of using embedded metadata are not fully understood yet, it is either considered to be an important tool for digital preservation and the interoperability of resources or an unstable process that will only lead to data corruption. As a result, the use of embedded metadata is far from widely accepted or standardized as a best practice in the United States.

Dr. Oksana L. Zavalina, © 2015-2016
QUOTE: “cataloging rules and codes, [ ...] guidelines for the format and syntax of the data values that are used to populate metadata elements”

A. Gilliland (2008)

Sy tax of a data value

<dc:creator>Shevchenko, Taras</dc:creator>
<dc:title>The Gypsy Fortune Teller</dc:title>
<dc:type>Painting</dc:type>
<dc:format>Oil on canvas</dc:format>
<dc:date>1841</dc:date>
<dc:coverage>Ukraine</dc:coverage>
<dc:subject>Fortune telling</dc:subject>
<dc:subject>Women</dc:subject>
<dc:subject>Children</dc:subject>
<dc:subject>Dogs</dc:subject>

Syntax of a metadata scheme
Data Content Standards in Libraries

• International Standard Bibliographic Description (ISBD)

• Anglo-American Cataloguing Rules (AACR) + Resource Description and Access (RDA)

• Library of Congress Subject Cataloging Manual

• Library of Congress Classification + Dewey Decimal Classification + Universal Decimal Classification

Identifies components in a metadata record

Content rules for descriptive metadata

Content rules for subject metadata

Rules for assigning classification numbers

Dr. Oksana L. Zavalina, © 2015-2016
Data Content Standards beyond Libraries

- International Guidelines for Museum Object Information: The CIDOC Information Categories
  - General International Standard Archival Description (ISAD(G))
- Cataloging Cultural Objects (CCO)
- Describing Archives: A Content Standard (DACS)

Identifies components in a museum metadata record

Identifies components in archival metadata record

Content rules for museum metadata

Content rules for archival metadata

Dr. Oksana L. Zavalina, © 2015-2016
Data Content Standards Tell Us

• **WHERE** to find the data for particular metadata element
  – e.g., take *Title* from the title page NOT from cover or spine for a book

• **WHAT** information to record, e.g.:
  – main title (aka title proper)
  – parallel title
  – remainder of the title (e.g., subtitle)

• **HOW** to record information in the record:
  – Input rules (punctuation, capitalization, word order, etc.)
  – Format of data

---

**Example:**

- **Creator:** Shevchenko, Taras
- **Title:** The Gypsy Fortune Teller
- **Type:** Painting
- **Format:** Watercolor on Bristol paper
- **Date:** 1841
- **Coverage:** Ukraine
- **Subject:** Fortune telling
- **Subject:** Women
- **Subject:** Children
- **Subject:** Dogs
Data Value Standards

Use of data value standards affects information representation and retrieval:

- increases consistency of representation
- improves information retrieval results

QUOTE: “controlled vocabularies, thesauri, controlled lists, [...] terms, names, and other values that are used to populate data”

Gilliland (2008)
Enable a person to **find** a book when one of the following is known:
1. The **author**
2. The title
3. The **subject**.

**Show** what works the library has:
1. By a given **author**
2. On a given **subject**
3. In a given **kind of literature**.

Assist in the **choice** of a book:
1. As to the edition (bibliographically)
2. As to its character (literary or topical).
Data Value Standards in Libraries

- Library of Congress Name Authority File (LCNAF)
- Library of Congress Subject Headings + Library of Congress Subject Authority File (LCSAF) + Medical Subject Headings (MESH)
- Library of Congress Classification + Dewey Decimal Classification + Universal Decimal Classification

Controlled vocabulary: data values for names
Controlled vocabulary: data values for subject metadata
Data Value Standards beyond Libraries

- Union List of Artist Names (ULAN)
- Thesaurus for Geographic Names (TGN)
- Virtual International Authority File
- Art and Architecture Thesaurus (AAT)
- Book Industry Study Group’s Subject Headings
- Book Industry Standards and Communications system of classification (BISAC)

Controlled vocabulary: data values for names

Controlled vocabulary: data values for subject metadata
Welcome to the Fun!
References


