Metadata Standards outside of DCMI

Basic Tutorial 2.1

Marcia Lei Zeng
DC-2009 "Semantic Interoperability of Linked Data", Seoul
I. The need for different kinds of metadata
II. Available standards
III. Understanding the differences
I. The need for different kinds of metadata

- Metadata are structured, encoded data
  - that describe characteristics of information-bearing entities
  - to aid in the identification, discovery, assessment, and management of the described entities

- For different types of entities, there will be different data structure.
Think: What did you see on food labels?

REDUCED FAT MILK
2% Milkfat

Nutrition Facts
Serving Size 1 cup (236ml)
Servings Per Container 1

<table>
<thead>
<tr>
<th>Amount Per Serving</th>
<th>Calories (120)</th>
<th>Calories from Fat 45%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Fat</td>
<td>5g</td>
<td>8%</td>
</tr>
<tr>
<td>Saturated Fat</td>
<td>3g</td>
<td>15%</td>
</tr>
<tr>
<td>Trans Fat</td>
<td>0g</td>
<td></td>
</tr>
<tr>
<td>Cholesterol</td>
<td>20mg</td>
<td>7%</td>
</tr>
<tr>
<td>Sodium</td>
<td>120mg</td>
<td>5%</td>
</tr>
<tr>
<td>Total Carbohydrate</td>
<td>11g</td>
<td>4%</td>
</tr>
<tr>
<td>Dietary Fiber</td>
<td>0g</td>
<td>0%</td>
</tr>
<tr>
<td>Sugars</td>
<td>11g</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>8g</td>
<td>17%</td>
</tr>
</tbody>
</table>

Vitamin A 10% • Vitamin C 4%
Calcium 30% • Iron 0% • Vitamin D 25%

*Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs.

NONFAT MILK

Nutrition Facts
Serving Size 1 cup (236ml)
Servings Per Container 1

<table>
<thead>
<tr>
<th>Amount Per Serving</th>
<th>Calories (80)</th>
<th>Calories from Fat 0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Fat</td>
<td>0g</td>
<td>0%</td>
</tr>
<tr>
<td>Saturated Fat</td>
<td>0g</td>
<td>0%</td>
</tr>
<tr>
<td>Trans Fat</td>
<td>0g</td>
<td>0%</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>less than 5mg</td>
<td>0%</td>
</tr>
<tr>
<td>Sodium</td>
<td>120mg</td>
<td>5%</td>
</tr>
<tr>
<td>Total Carbohydrate</td>
<td>11g</td>
<td>4%</td>
</tr>
<tr>
<td>Dietary Fiber</td>
<td>0g</td>
<td>0%</td>
</tr>
<tr>
<td>Sugars</td>
<td>11g</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>3g</td>
<td>17%</td>
</tr>
</tbody>
</table>

Vitamin A 10% • Vitamin C 4%
Calcium 30% • Iron 0% • Vitamin D 25%

*Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs.

Metadata are structured, encoded data that describe characteristics of information-bearing entities ...
Think: What information is in a personal record?

Metadata are structured, encoded data that describe characteristics of information-bearing entities ...
Think: How do we know what a map is about?

Metadata are **structured**, encoded data that describe characteristics of information-bearing entities ...
Think: What should be in a business directory?

Metadata are structured, encoded data that describe characteristics of information-bearing entities ...
Structured data ... for photos?

Metadata are structured, encoded data that describe characteristics of information-bearing entities ...
### Structured data ... for objects?

**Names/Titles:**
- Hagia Sophia
- Church of the Holy Wisdom
- Ayasofya
- Agia Sofia
- Agia Sophia

**Sancta Sophia**

**Language:** Latin

**Current Location:** Istanbul (Marmara region, Turkey)

**Display Creator:** architects: Anthemios of Tralles (Byzantine architect and mathematician in Asia Minor, ca. 5th century) and Isidoros of Miletus, the Elder (Byzantine architect and engineer in Asia Minor, active mid-6th century)

**Related People/Corporate Bodies:**
- Anthemios of Tralles
  - Role: architect
- Isidoros of Miletus
  - Role: architect

**Display Creation Date:**
- original structure dated from 4th century CE; present structure built 532-537 CE; rebuilt in 12th century

**Work Type:**
- church
- mosque
- museum

**Technique/Medium display:**
- bearing masonry; interior surfaces are sheathed with polychrome marble, porphyry, and mosaics

**Measurements display:**
- central dome: diameter 31 meters (102 feet); height 56 meters (184 feet)

**Descriptive Note:**
- Related Works:
- Sources:

---

**Metadata are structured, encoded data that describe characteristics of information-bearing entities ...**
**Think: how to encode and share Scientific Data?**

**Carcharodon carcharias**
Name verified: Catalogue of Life; FishBase; Compagno, Leonard J.V.
Organism type: a fish [FishBase](#)

**Data Extent Map (from OBIS Australia/ C Square Mapper)**

<table>
<thead>
<tr>
<th>Metadata Field</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kingdom</td>
<td>Highly Recommended when known</td>
</tr>
<tr>
<td>Phylum</td>
<td>Optional</td>
</tr>
<tr>
<td>Class</td>
<td>Optional</td>
</tr>
<tr>
<td>Order</td>
<td>Optional</td>
</tr>
<tr>
<td>Family</td>
<td>Optional</td>
</tr>
<tr>
<td>Genus</td>
<td>Highly Recommended when known</td>
</tr>
<tr>
<td>Subgenus</td>
<td>Optional</td>
</tr>
<tr>
<td>Species</td>
<td>Highly Recommended when known</td>
</tr>
<tr>
<td>Subspecies</td>
<td>Optional</td>
</tr>
<tr>
<td>Scientific Name</td>
<td>Optional</td>
</tr>
<tr>
<td>Author</td>
<td>Text</td>
</tr>
</tbody>
</table>

The author of a scientific name. An accepted name. Can be more than one. Should be formatted according to the relevant rules of Nomenclature. Parentheses are not used, e.g., if the name of an animal, the authority and year should be mentioned (Hastings, 1986)

*Metadata are structured, encoded data that describe characteristics of information-bearing entities.*
II. Available standards

Developed by many communities for:

- published books
- electronic documents
- archival finding aids
- art objects
- educational and training materials
- scientific datasets
- geospatial information
- e-Government
- business
- multimedia
- etc.
Types of Standards

1. Standards for *data structures and semantics*
   - metadata element sets
     - e.g., Dublin Core Metadata Element Set

2. Standards and Guides for *data content*
   - created to guide the practices of metadata generation
     - e.g., *Using Dublin Core* | AACR2 | DACS (*Describing Archives: A Content Standard*) | CCO (*Cataloging Cultural Objects*)

3. Standards for *data values*
   - value encoding schemes (link to list)
     - e.g., *DCMI Type Vocabulary* | *RFC 4646 Tags for Identifying Languages* | *Art and Architecture Thesaurus* | LCSH
# Metadata Element Sets (a selected list) (1)

<table>
<thead>
<tr>
<th>Purpose of metadata</th>
<th>Metadata Element Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>General resource</td>
<td>DCMES and DCMI Metadata Terms</td>
</tr>
<tr>
<td>Bibliographic resource</td>
<td>MODS (Metadata Object Description Schema)</td>
</tr>
<tr>
<td></td>
<td>MARC21, UNIMARC, MARCXML, TEI (Text Encoding Initiative) Headers</td>
</tr>
<tr>
<td>Cultural objects and visual resources</td>
<td>CDWA (Categories for the Description of Works of Art)</td>
</tr>
<tr>
<td></td>
<td>CDWA Lite</td>
</tr>
<tr>
<td></td>
<td>VRA (Visual Resources Association) Core Categories</td>
</tr>
<tr>
<td>Archives &amp; preservation</td>
<td>EAD (The Encoded Archival Description)</td>
</tr>
<tr>
<td></td>
<td>OAIS (Reference Model for an Open Archival Information System)</td>
</tr>
<tr>
<td></td>
<td>PREMIS (Preservation Metadata: Implementation Strategies)</td>
</tr>
</tbody>
</table>

Link to list: Metadata Standards
<table>
<thead>
<tr>
<th>Category</th>
<th>Metadata Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational resources</td>
<td>LOM (Learning Object Metadata)</td>
</tr>
<tr>
<td></td>
<td>CanCore</td>
</tr>
<tr>
<td></td>
<td>SCORM (Sharable Content Object Reference Mode)</td>
</tr>
<tr>
<td>Publishing</td>
<td>ONIX (ONline Information Exchange)</td>
</tr>
<tr>
<td>Rights management</td>
<td>copyrightMD</td>
</tr>
<tr>
<td></td>
<td>DOI (Digital Object Identifier)</td>
</tr>
<tr>
<td></td>
<td>ODRL (Open Digital Rights Language)</td>
</tr>
<tr>
<td>Scientific resources</td>
<td>CSDGM (Content Standard for Digital Geospatial Metadata)</td>
</tr>
<tr>
<td></td>
<td>Darwin Core</td>
</tr>
<tr>
<td>Multimedia objects</td>
<td>MPEG-7 Multimedia Content Description Interface</td>
</tr>
<tr>
<td></td>
<td>PBCore, The Public Broadcasting Metadata Dictionary</td>
</tr>
<tr>
<td>Agents</td>
<td>vCard</td>
</tr>
<tr>
<td></td>
<td>FOAF (Friend Of A Friend)</td>
</tr>
<tr>
<td>Accessibility</td>
<td>IMS AccessForAll Meta-data</td>
</tr>
</tbody>
</table>
Types of Standards

1. Standards for *data structures and semantics*
   - metadata element sets
     e.g., Dublin Core Metadata Element Set

2. Standards and Guides for *data content*
   - created to guide the practices of metadata generation
     e.g., *Using Dublin Core* | AACR2 | DACS
     *(Describing Archives: A Content Standard)* | CCO
     *(Cataloging Cultural Objects)*

3. Standards for *data values*
   - value encoding schemes (link to list)
     e.g., *DCMI Type Vocabulary* | RFC 4646 Tags for Identifying Languages
     | *Art and Architecture Thesaurus* | LCSH
Standards and Guides for data content

A selected list
DCMI related:
- Using Dublin Core
- NSDL_DC Metadata Guidelines
- Best Practices for Shareable Metadata

Developed by other communities:
- DACS (Describing Archives: A Content Standard)
- CCO (Cataloging Cultural Objects)
- CSDGM (Content Standard for Digital Geospatial Metadata)
- DLESE Metadata Best Practices, Digital Library for Earth System Education
- AACR2 (Anglo-American Cataloguing Rules. 2nd Rev. ed.)
- RDA (Resource Description and Access)
- etc.
4.13. Format

Label: Format

Element Description: The physical or digital manifestation of the resource. Typically, Format may include the media-type or dimensions of the resource. Examples of dimensions include size and duration. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource.

Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [http://www.iana.org/assignments/media-types/] defining computer media formats).

Guidelines for content creation:

In addition to the specific physical or electronic media format, information concerning the size of a resource may be included in the content of the Format element if available. In resource discovery size, extent or medium of the resource might be used as a criterion to select resources of interest, since a user may need to evaluate whether they can make use of the resource within the infrastructure available to them.

When more than one category of format information is included in a single record, they should go in separate iterations of the element.

Examples:

Title="Dublin Core icon"
Type="Image"
Format="image/gif"
Format="4 KB"

Subject="Saturn"
Type="Image"
Format="image/gif 6"
Format="40 x 512 pixels"
Identifier="http://www.not.iac.es/newwww/photos/images/satnot.gif"

Title="The Bronco Buster"
Creator="Frederic Remington"
Type="Physical object"
CCO, a content standard for the cultural heritage community.

- guidance for minimal records
- relationships between work and image records
- describing complex works
- database design and entity relationships
- authority files and controlled vocabularies
- Managing objects/images
- rules for descriptive cataloging
Types of Standards

1. Standards for *data structures and semantics*
   - metadata element sets
     - e.g., Dublin Core Metadata Element Set

2. Standards and Guides for *data content*
   - created to guide the practices of metadata generation
     - e.g., *Using Dublin Core* | AACR2 | DACS (*Describing Archives: A Content Standard*) | CCO (*Cataloging Cultural Objects*)

3. Standards for *data values*
   - value encoding schemes
     - e.g., *DCMI Type Vocabulary Art and Architecture Thesaurus LCSH, DDC, MeSH ...*
Standards for *data values*

value encoding schemes

- Content data for some elements may be selected from a controlled vocabulary:
  - Standardized vocabularies
  - Name authority control
  - Controlled terms

[link to list: Value Encoding Schemes and Content Standards](http://www.metadataetc.org/book-website/readings/appendixbencodcschemes.htm)
**Term Name:** type

<table>
<thead>
<tr>
<th>URI:</th>
<th><a href="http://purl.org/dc/elements/1.1/type">http://purl.org/dc/elements/1.1/type</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Label:</td>
<td>Type</td>
</tr>
<tr>
<td>Definition:</td>
<td>The nature or genre of the resource.</td>
</tr>
<tr>
<td>Comment:</td>
<td>Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.</td>
</tr>
</tbody>
</table>
**Term Name:** format

**URI:** http://purl.org/dc/elements/1.1/format

**Label:** Format

**Definition:** The file format, physical medium, or dimensions of the resource.

**Comment:** Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].

**Examples:** values associated with FORMAT element found in the research samples:

- text
- text/html
- text/plain
- plain
- digital TIFF
- image/tiff
- other
- application/msword
- application/Flash (animation)
- ascii
- pdf ps
- 3.6 megabytes
- 1000149 bytes
- language/java
- Application/JAVA applet
- Java
- CLASS
- Model/VRML
- AVI, MOV, QTM
- 1 v. (various pagings)
- p.461-470
- viii, 82 p.
- MPEG-4
Example from LOM (Learning Object metadata)

5.2 Learning Resource Type

Explanation: Specific kind of learning object. The most dominant kind shall be first.

NOTE: --The vocabulary terms are defined as in the OED:1989 and as used by educational communities of practice.

<table>
<thead>
<tr>
<th>Controlled terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value Space:</td>
</tr>
<tr>
<td>ordered exercise</td>
</tr>
<tr>
<td>simulation</td>
</tr>
<tr>
<td>questionnaire</td>
</tr>
<tr>
<td>diagram</td>
</tr>
<tr>
<td>figure</td>
</tr>
<tr>
<td>graph</td>
</tr>
<tr>
<td>index</td>
</tr>
<tr>
<td>slide</td>
</tr>
<tr>
<td>table</td>
</tr>
<tr>
<td>narrative text</td>
</tr>
<tr>
<td>exam</td>
</tr>
<tr>
<td>experiment</td>
</tr>
<tr>
<td>problem statement</td>
</tr>
<tr>
<td>self assessment</td>
</tr>
<tr>
<td>lecture</td>
</tr>
</tbody>
</table>
III. Understanding the differences

1. purpose (see previous list)
2. elements
3. schema’s structure
Not only describing an object, but also describing the *use* of objects.

**Categories and elements**

- **general**: title, language, description, keywords...
- **life cycle**: contribute-role, version...
- **metametadata**: scheme, language, contribute ...
- **technical**: format, size, requirements, installation remarks, platform requirements, duration ...
- **educational**: learning resource type, learning time, interactivity level, intended end user role, typical age range, typical learning time, difficulty, context...
- **rights**: price, copyright...
- **relation**
- **annotation**
- **classification**: purpose
Hierarchical Structure
Elements, Sub-element, & Attributes

[example from VRA Core 4.0]

- **agent**
  - attribution
  - culture
  - **dates** *(type)*
    - **earliestDate** *(circa)*
    - **latestDate** *(circa)*
  - name *(type)*
  - role
Attributes & Attribute Values

[element]
- [sub-element]
  - [sub-sub-element]
  [example from VRA Core 4.0]
Wrapper, <display>, annotation

(using vra:agent element as an example)
For your metadata project

• Set your goal for the metadata to accomplish
• Compare existing and desired metadata
• Work with a metadata standard — whether to adopt, modify, reuse, or start a new one depends on many factors
• Sort our relationships between associated objects
• Decide level of granularity
For your metadata project

• understand **how** and **why**
  ■ implementing existing standards
  ■ developing element sets and application profiles

• identify **where and when** the project is situated
  ■ the stage
  ■ entities to be dealt with
  ■ functional requirements
  ■ collaboration opportunities

• find out **who** is responsible for **what**

• contribute to the creation of good quality, shareable metadata