Dublin Core: metadata for interoperability

- Objectives of work on Dublin Core metadata
  - To make it easier to find information
  - To offer Dublin Core as one of the core element sets (now: metadata vocabularies) to create wide interoperability across domains and applications
  - To enable mixing and matching of metadata vocabularies through Application Profiles

Rich landscape of metadata specifications

- Dublin Core as a basic set of metadata terms for cross-domain interoperability
- Many domain-specific sets of terms, e.g.
  - MODS (libraries)
  - LOM Learning Object Metadata (education)
  - ISO19115 (geographic information)
  - Etc. etc. etc.

Cross-specification interoperability

- Need to look at issues across specifications
- For example:
  - Different models (flat vs. hierarchical, object-centric vs. event-centric etc.)
  - Different technologies (HTML, XML, RDF etc.)
  - Different objectives (discovery, management etc.)
DCMI: establish co-operation

• With education community
  – Ottawa Communiqué “Harmonisation of Metadata for Education and Training Communities”, August 2001
  – DCMI/IEEE LTSC Taskforce est. 2005 to produce a joint DCMI/IEEE specification on how to use IEEE LOM elements with Dublin Core metadata
• With others:
  – DCMI/RDA Task Group, DCMI/NKOS Task Group, DCMI/FOAF collaboration

This Webinar

• Acknowledging that interoperability across specifications is crucial for shared Web environment
• Looking at challenges and opportunities based on experience with DCMI/IEEE co-operation
• Providing concrete illustrations of harmonization issues and a roadmap for interoperable design

The speakers

• Tom Baker
  – Chief Information Officer at DCMI
  – Co-chair of DCMI Architecture Forum
• Mikael Nilsson
  – Co-chair of DCMI Architecture Forum
  – Leader of DCMI/IEEE LTSC Taskforce
  – Recent dissertation on this very subject (KTH, 2010)
    http://kmi.nada.kth.se/papers/SemanticWeb/FromInteropToHarm-MikaelaThesis.pdf

And don’t forget

Theme: Metadata Harmonization: Bridging Languages of Description

http://purl.org/dcevents/dc-2011
Road to harmonization

NISO Webinar
16 March 2011
Tom Baker

Traditional IT: Integrate across silos using shared formats or “crosswalks”

The metadata ecosystem

2003: DCMI, MARC21, IEEE/LOM... agree to identify “metadata elements” with URIs

Identifying Metadata Elements with URIs

The CORES Resolution

Thomé Bäiker
Bibliothèque de France, France@OzGeOff
<thom@pp@fr>

Mike Quinlan
President, OBiO & OBiO
<mailto@obio.com>

Abstract

On 18 November 2003, a meeting organized by the CORES Project (Information Society Technologies Programme, European Union), several organizations regarded as mainstays of metadata standards achieved consensus on a resolution to assign Uniform Resource Identifiers (URIs) to metadata elements as a second step towards the development of mapping infrastructures and interoperability services. The signatories of the CORES Resolution agreed to promote the consensus in their communities and beyond and to implement an action plan in the following six months. Six months having passed, the maintenance of DCMI, MARC21, IEEE/LOM, and OBiO agreed on their implementations of the resolution and highlighted areas of concern to establish good practice for declaring, identifying, and maintaining metadata elements more generally. In June 2003, the resolution was also endorsed by the maintenance of OBiO/ARIB.
But identifying Apples and Oranges with URIs does not make them comparable!

Overview

- The metadata ecosystem
- Central definitions
- Metadata concepts
- Metadata combinations
- Core harmonization issues
- Harmonization recipe
- Future developments
The metadata ecosystem

Central definitions

➢ Metadata:
   Descriptive data about identifiable things
Central definitions

 Metadata:
 Descriptive data about identifiable things

 Metadata interoperability:
 the ability of two or more systems or components to exchange descriptive data about things, and to interpret the descriptive data that has been exchanged in a way that is consistent with the interpretation of the creator of the data.
Central definitions

Metadata interoperability: the ability of two or more systems or components to exchange descriptive data about things, and to interpret the descriptive data that has been exchanged in a way that is consistent with the interpretation of the creator of the data.

Metadata harmonization: the ability of two or more systems or components to exchange "combined metadata" conforming to two or more metadata specifications, and to interpret the metadata that has been exchanged in a way that is consistent with the intentions of the creators of the metadata.
Central definitions

- **Metadata harmonization:**
  the ability of two or more systems or components to exchange "combined metadata" conforming to two or more metadata specifications, and to interpret the metadata that has been exchanged in a way that is consistent with the intentions of the creators of the metadata.
Central definitions

- Metadata harmonization:
  the ability of two or more systems or components to exchange “combined metadata” conforming to two or more metadata specifications, and to interpret the metadata that has been exchanged in a way that is consistent with the intentions of the creators of the metadata.

![Diagram]

Core metadata concepts

- Metadata syntaxes
- Metadata vocabularies
- Abstract Models
- Application profiles

Concepts: Metadata syntaxes

- Concrete data formats used to exchange metadata between applications
- Used to implement metadata software and protocols
- Practical tool, but not essence of harmonization issues
- Examples: RDF/XML, DC-HTML, LOM XML

Syntax: LOM XML

```xml
<resourcerepresentation>
  <description>
    <content language="en"/>
    <content language="en">
      This resource represents an airplane flight ...
    </content>
  </description>
  <identifier>
    <content language="en"/>
    <content language="en">
      Flight 1 from Los Angeles to New York ...
    </content>
  </identifier>
  <relation>
    <content language="en"/>
    <content language="en">
      Flight 1 is related to the airplane flight ...
    </content>
  </relation>
</resourcerepresentation>
```
Concepts: Vocabularies

- Sets of descriptive terms for use in metadata descriptions
- Used by metadata "designers" to create metadata records
- Defined according to various models; conflicting models lead to harmonization issues
- Examples: DCMES, MARCREL, LCSH

Example vocabulary: LCSH

Concepts: Abstract models

- Models used to define the meaning and usage of metadata terms
- Used by term designers and syntax creators to create interoperable specifications
- Incompatible models are major barrier to harmonization
- Examples: RDF triple model, IEEE LOM hierarchical model, DCMI abstract model
Concepts: Application profiles

- Defines the structure of metadata records in a particular context (domain, application, etc.)
  - use vocabularies on the basis of an abstract model to define a concrete syntax
- Used by application designers and domain experts to codify domain needs
- Useful for harmonization within the context of a single abstract model
- Examples: ePrints AP, OAI-DC, etc

Syntactic combinations (MODS & LOM)

```xml
<?xml version='1.0' encoding='UTF-8' ?>
<mods ...>
  <subject authority="lcsh">
    <topic>Parachuting</topic>
  </subject>
  <extension>
    <!-- LOM fragment: -->
    <lom:description>
      <lom:string lom:language="en">
        Useful for learning some flight-related French terminology.
      </lom:string>
      <lom:string language="sv">
        Användbar för att lära sig lite flygrelaterad fransk terminologi.
      </lom:string>
    </lom:description>
    <!-- End LOM fragment. -->
  </extension>
</mods>
```
Why syntactic combinations fail

<table>
<thead>
<tr>
<th>Base format</th>
<th>Extended with</th>
<th>Processable by LOM</th>
<th>Processable by MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOM XML</td>
<td>MODS</td>
<td>Only LOM part</td>
<td>No</td>
</tr>
<tr>
<td>MODS</td>
<td>LOM XML</td>
<td>No</td>
<td>Only MODS part</td>
</tr>
</tbody>
</table>

Application A (MODS)

Application B (LOM)

express/encode

LOM abstract syntax

LOM XML binding

MODS (XML)
## Why syntactic combinations fail

<table>
<thead>
<tr>
<th>Base format</th>
<th>Extended with</th>
<th>Processable by LOM</th>
<th>Processable by MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOM XML</td>
<td>MODS</td>
<td>Only LOM part</td>
<td>No</td>
</tr>
<tr>
<td>MODS</td>
<td>LOM XML</td>
<td>No</td>
<td>Only MODS part</td>
</tr>
</tbody>
</table>

**Application A (MODS)**

- express/encode
- interpret

**Application B (MODS)**

- express/encode
- interpret

**Application B (LOM)**

- insert

**LOM abstract syntax**

**LOM XML binding**

---

**Why syntactic combinations fail**

<table>
<thead>
<tr>
<th>Base format</th>
<th>Extended with</th>
<th>Processable by LOM</th>
<th>Processable by MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOM XML</td>
<td>MODS</td>
<td>Only LOM part</td>
<td>No</td>
</tr>
<tr>
<td>MODS</td>
<td>LOM XML</td>
<td>No</td>
<td>Only MODS part</td>
</tr>
</tbody>
</table>

**Application A (MODS)**

- express/encode
- interpret

**Application B (MODS)**

- express/encode
- interpret

**Application B (LOM)**

- insert

**LOM abstract syntax**

**LOM XML binding**

**NO agreement**

---

**Why syntactic combinations fail**

<table>
<thead>
<tr>
<th>Base format</th>
<th>Extended with</th>
<th>Processable by LOM</th>
<th>Processable by MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOM XML</td>
<td>MODS</td>
<td>Only LOM part</td>
<td>No</td>
</tr>
<tr>
<td>MODS</td>
<td>LOM XML</td>
<td>No</td>
<td>Only MODS part</td>
</tr>
</tbody>
</table>

**Application A (MODS)**

- express/encode
- interpret

**Application B (MODS)**

- express/encode
- interpret

**Application B (LOM)**

- insert

**LOM abstract syntax**

**LOM XML binding**

**NO agreement**

---

**Why syntactic combinations fail**

<table>
<thead>
<tr>
<th>Base format</th>
<th>Extended with</th>
<th>Processable by LOM</th>
<th>Processable by MODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOM XML</td>
<td>MODS</td>
<td>Only LOM part</td>
<td>No</td>
</tr>
<tr>
<td>MODS</td>
<td>LOM XML</td>
<td>No</td>
<td>Only MODS part</td>
</tr>
</tbody>
</table>

**Application A (MODS)**

- express/encode
- interpret

**Application B (MODS)**

- express/encode
- interpret

**Application B (LOM)**

- insert

**LOM abstract syntax**

**LOM XML binding**

---
Successful combinations – RDF

This learning object explains parachuting.

Useful for learning some flight-related French terminology.

Harmonization in software

Fedora – digital repository

• Storage of multiple metadata standards

Fedora – digital repository

• Storage of multiple metadata standards
• Exposes two standards for search:
  o Dublin Core (15 properties)
  o Object relations (RDF)
Harmonization in software

Fedora – digital repository

- Storage of multiple metadata standards
- Exposes two standards for search:
  - Dublin Core (15 properties)
  - Object relations (RDF)

Harmonization in software

Fedora – digital repository

- Storage of multiple metadata standards
- Exposes two standards for search:
  - Dublin Core (15 properties)
  - Object relations (RDF)
Harmonization in software

Fedora – digital repository
- Storage of multiple metadata standards
- Exposes two standards for search:
  - Dublin Core (15 properties)
  - Object relations (RDF)

SCAM – RDF-based repository
- Import of multiple metadata standards
- Exposes two standards for search:
  - Dublin Core (15 properties)
  - Object relations (RDF)
Harmonization in software

Fedora – digital repository
• Storage of multiple metadata standards
• Exposes two standards for search:
  o Dublin Core (15 properties)
  o Object relations (RDF)

SCAM – RDF-based repository
• Import of multiple metadata standards
• Exposes full metadata for search

SCAM – RDF-based repository
• Import of multiple metadata standards
• Exposes full metadata for search
Harmonization in software

Fedora – digital repository
- Storage of multiple metadata standards
- Exposes two standards for search:
  - Dublin Core (15 properties)
  - Object relations (RDF)

SCAM – RDF-based repository
- Import of multiple metadata standards
- Exposes full metadata for search

Metadata semantics and harmonization

Semantics: the interpretation of data
- Makes metadata out of data

Three kinds
- Informal: human semantics (language) – “A is a kind of B”
- Machine-processable: semantics encoded as data – `<A> rdfs:subClassOf <B>`.
- Formal: logical semantics
  \[
  \langle x,y \rangle \in \text{EXT}(\{ \text{rdfs:subClassOf} \}) \iff \text{IC} \land \text{ICEXT}(x) \subseteq \text{ICEXT}(y)
  \]

Semantic metadata interoperability

When two systems can exchange machine-processable semantics alongside the metadata and interpret this semantics correctly.
Semantic metadata interoperability

When two systems can exchange machine-processable semantics alongside the metadata and interpret this semantics correctly.

Recipe for harmonization

- **Adopt a core model** with support for machine-processable semantics
- **Construct mappings** of other standards that preserve semantics

Deconstructing the harmonization recipe:
Abstract Model
Semantics
Abstract Syntax
Schema model

Various metadata standards (LOM, ISO MARC, DCAM, etc.)
Metadata formats (XML, N3, XHTML, etc.)
Profile models (DSP etc.)
Application profiles
Element vocabularies
Value vocabularies

Core model
Specifications
Domain definitions

LOM AP
ESWA
UK LOM Core

Profile Models
Example APs
UK LOM Core

Profiles models
Value vocabularies

DC
RDA
LOM
Vocabularies
Vocabulary Models
SKOS etc.
H
LCS
Example element vocabularies
Example value vocabularies
Various metadata standards (LOM, ISO MLR, DCAM, etc.)

Abstract Model
- Semantics
- Schema model

Metadata formats (XML, RDF, JSON, etc.)

Profile models (DSP, etc.)

Vocabulary Models (SKOS, etc.)

Application profiles

Element vocabularies

Value vocabularies

Ontologies

Core model
- Specifications
- Format definitions

Beyond Harmonization: Alignment

NISO Webinar
16 March 2011
Tom Baker

Conclusions & future directions

- Harmonization is not the same as interoperability
- Application profiles are useful for harmonization within specifications, but not between them
- Semantic metadata interoperability basis for harmonization
- Focus on abstract models
- Increased focus on harmonization in standardization activities
- Modularization of standards an important tool
- Harmonization as basis for Linked Data efforts

Beyond harmonization: explicit “alignment”

Problem: proliferation of similar elements
- dc:title – A name given to the resource.
- rdfs:label – human-readable resource name
- gr:name – short descriptive text, like dc:title
- foaf:name – A name for something.

Solution: explicit alignments:
- dct:creator owl:equivalentProperty foaf:maker