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://brn.nada.kth.se/papers/SemanticWeb/FromInteropToForm-MikaelThesis.pdf](http://brn.nada.kth.se/papers/SemanticWeb/FromInteropToForm-MikaelThesis.pdf)

And don’t forget

**DC-2011**

At the National Library of the Netherlands

The Hague, 21–23 September 2011

Theme: Metadata Harmonization: Bridging Languages of Description


Road to harmonization

NISO Webinar

16 March 2011

Tom Baker

The metadata ecosystem

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

- Schema A
- Schema B
- Schema C

2003: DCMI, MARC21, IEEE/LOM... agree to identify “metadata elements” with URIs
But identifying Apples and Oranges with URIs does not make them comparable!

Not just shared URIs, but shared model

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

The metadata ecosystem

Education
Libraries
Government
Multimedia

Useful for learning some flight-related French terminology.
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.

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
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

Example abstract model: DCAM

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 version='1.0' encoding='UTF-8' ?><mods ...

Subject authority="lcsh">
  <topic>Parachuting</topic>
</subject>
Syntactic combinations (MODS & LOM)

```xml
<?xml version='1.0' encoding='UTF-8' ?>
<mods ...>
  <subject authority="lcsh">
    <topic>Parachuting</topic>
  </subject>
  <extension>
  </extension>
</mods>
```

Why syntactic combinations fail

<table>
<thead>
<tr>
<th>Raw 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

Application B (LOM)

- insert
- LOM abstract syntax
- LOM XML testing
Successful combinations – RDF

Harmonization in software

Fedora – digital repository
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 full metadata 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 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
    \[
    x \land y \land \neg \text{IC}(x) \land \text{ICEXT}(y) \Rightarrow \neg \text{IC}(y)
    \]

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**
  - **Core model**

Specifications

- Metadata formats (XML, N3, XHTML, etc.)
- Various metadata standards (LOM, ISO MLR, DCAM, etc.)
Various metadata standards (LOM, ISO MLR, DCAM, etc.)

Abstract Model
Semantics
Element vocabularies
Metadata formats (XML, N3, XHTML, etc.)
Application profiles
UK LOM Core
DC
RDA
LOM

Abstract Syntax
Schema model
Profile Models (DSP etc.)

Ontology Models (OWL etc)
Core model
Specifications
Domain definitions
Ontologies
Value vocabularies
Vocabulary Models (SKOS etc.)
LCS
H
LOM
Vocabs
RDA
Vocabs

Example APs
Example element vocabs
Example value vocabs

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

Questions?

All questions will be posted with presenter answers on the NISO website following the webinar:

THANK YOU

Thank you for joining us today.
Please take a moment to fill out the brief online survey.
We look forward to hearing from you!