

DCMI and the metadata landscape

Tutorial

Dublin Core – Building blocks for interoperability

17 December 2009

Makx Dekkers



Again: the basic idea (1995)

- A basic description mechanism for digital information that:
 - can be used in all domains
 - can be used for any type of resource
 - is simple, yet powerful
 - can be extended and can work with specific solutions
 - Making it easier to find information on the Web as it develops
-



And the Web developed...

➤ Netcraft.com:

- 1995: 18,000 Web sites
- 2000: 10 million Web sites
- 2009: 240M host names, 75M active Web sites

➤ Google.com:

- 1998: 26 million pages
- July 2008: 1 trillion pages



Where did Dublin Core go?

- Manual metadata does not scale to the trillions
- Metadata is unreliable (people lie!)
- Search engines took care of the Open Web
- Dublin Core metadata came to be mostly used in context of controlled collections
 - as a basic description mechanism
 - as an exchange format



What about search engines?

- Effectiveness and scalability of metadata approaches, compared to search engines
- Different environments
 - metadata are applied to controlled collections of all kinds of resources; concentrate on **context**

Sharing information for use

- search engines operate on uncontrolled collections of (mostly) textual resources; concentrate on **content**

Indexing information for discovery



It's about cost and benefit

➤ Metadata:

- expensive in terms of creation and maintenance
- need strong (business) case for delivering high-quality results for specific goals

➤ Search engines:

- rely mostly on relatively cheap background processing and smart algorithms
- deliver ranked results for interpretation by user



Making the best of both worlds

➤ Appropriate solutions:

- metadata systems for professional use with high-quality requirements (e.g. completeness)
- search engines for broad-brush discovery with “best” results first

➤ Crossing over:

- metadata systems and full-text indexing
- search engines with metadata for advanced search



Are metadata luxury?

- Library and information science perspective:
 - too much content to manage centrally
 - distribution of control, metadata creation “at-source”
 - how to manage quality?
 - Business perspective (different words!):
 - product descriptions are part of the process
 - cans without labels don't sell!
 - wrong information may lead to liability
-



The digital environment

➤ Approach is shifting:

- more and more emphasis on integration of metadata in production process, e.g. capturing document properties, time, place
- libraries as aggregators, e.g. institutional repositories

➤ Maintenance of metadata:

- maintenance of quality across collections is crucial
- persistence is an organizational policy!



Inter. Operability.

➤ **Interoperability**

- communication between partners in a network (many-to-many), not a star topology (many-to-one)

➤ **Interoperability**

- processing by machines, not just by human users

➤ Common understanding requires:

Metadata



Interoperability layers

- European Interoperability Framework 2.0:
 - **political (business) context:** common goals
 - **legal interoperability:** rights and obligations
 - **organisational interoperability:** organisational and business processes
 - **semantic interoperability:** meaning of exchanged information
 - **technical interoperability:** tools, protocols

<http://ec.europa.eu/idabc/servlets/Doc?id=31597>



Drivers

- Business/policy requirement:
 - a commitment to work together, to share information
 - interoperability is not a goal in itself, and it may not be the cheapest option!
- Functional requirement:
 - data integration requires common understanding of what the information is and how it is described
 - matching local and common needs



Sharing meaning

➤ Semantic aspects:

- resources (objects, documents) and relationships
- metadata standards, vocabularies, code lists, thesauri

➤ Domain aspects:

- specific domains – specific standards with rich meaning
- across domains – small set of shared meaning:

Dublin Core Metadata



Domain-specific approaches

- Specific domain requirements are the basis of specific solutions and standards, e.g.
 - management of library collections: MARC21
 - management of museum information: CIDOC
 - management of archival resources: EAD
 - management of environmental data: INSPIRE
 - etc. etc.

 - Create “islands of intra-operability”
-



Complete solutions

- Specific requirements are satisfied by specific solutions and standards, e.g.
 - definition of specific types of resources to be covered
 - specific application functionality to be provided
- But: one size does not fit all!
 - emergence of local extensions, local dialects
 - co-operation across domains, national boundaries not easy: requires one-to-one agreements, mappings



Description approaches

➤ The “traditional” approach

- Usually based on a description record, e.g. XML
- Looks at the item at hand and describes all aspects of the item and related resources:

```
<record>
  <name>DCMI Tutorial</name>
  <author>
    <name>Makx Dekkers</name>
    <affiliation>DCMI</affiliation>
  </author>
  <publisher>
    <name>DCMI</name>
    <location>Singapore</location>
  </publisher>
</record>
```




Description approaches

➤ The “semantic” approach

- Based on “statements”, e.g. RDF triples
- Describes the characteristics of the item (***and only of this one item***) – Dublin Core **one-to-one principle!**

```
<Description about=Presentation>  
  <name>DCMI Tutorial</name>  
  <presenter>Makx</presenter>  
  <publisher>DCMI</publisher>  
</Description>
```

```
<Description about=Makx>  
  <ssn>X1234567X</ssn>  
  <affiliation>DCMI</affiliation>  
</Description>
```

```
<Description about=DCMI>  
  <regno>200823602C</regno>  
  <inc_date>2008-12-23</inc_date>  
  <location>Singapore</location>  
</Description>
```



Comparing approaches

- The “traditional” approach
 - single records but duplication of information
 - implicit linkage
 - controlled environments (intra-operability)
 - The “semantic” approach
 - descriptions can be used in other contexts
 - explicit linkage
 - open environments (Linked Data interoperability)
-



Cooperation approaches

- DCMI has always sought to cooperate to come to common understanding
- **“None of us is as smart as all of us”**
- Current joint activities:
 - DCMI/RDA (Resource Description and Access)
 - DCMI/IEEE (Learning Technologies Standards Committee – Learning Object Metadata (LOM))



Example: DCMI/RDA

DCMI/RDA Task Group Wiki

This Task Group is for collaborative work to enable broader use of the Resource Description and Access (RDA), building on agreements made at a [meeting](#) held at the British Library April 30/May 1, 2007. Participants in the meeting came from DCMI and other Semantic Web groups, and the RDA development effort. The Task Group is led by Diane Hillmann (then of Cornell University, now at Syracuse University) and Gordon Dunsire of Strathclyde University.

Charter and Work Plan

Charter: To define components of the draft standard "RDA - Resource Description and Access" as an RDF vocabulary for use in developing a Dublin Core application profile.

Workplan for RDA Vocabularies Project

1. Define RDA modeling entities as an RDF vocabulary (properties and classes).
2. Identify in-line value vocabularies as candidates for publication in [RDFS](#) or [SKOS](#).
3. Develop a DC Application Profile for RDA based on FRBR and FRAD.



Example: DCMI/IEEE

The taskforce will produce a joint DCMI/IEEE specification with the working title

“Recommendation for using IEEE LOM Elements in Dublin Core Metadata”

Scope of the recommendation

This activity will develop a recommended representation of the metadata elements of the IEEE Learning Object Metadata Standard in the Dublin Core Abstract Model. The recommendation will include the specification of a number of terms, including properties, syntax encoding schemes, vocabulary encoding schemes as well as vocabularies, that may be used for expressing metadata conforming to the IEEE LOM Standard in Dublin Core metadata. The recommendation will also include the specification of namespaces to use for the terms, as well as a basic application profile describing how to combine the specified terms in a way that is compatible with the structural constraints of the IEEE LOM Standard.

Purpose of the recommendation

There is an increasing demand for interoperable definition of metadata terms which allows the terms to be used in different kinds of combinations. The IEEE LOM and Dublin Core metadata standards are suffering from a lack of interoperability, making it unnecessarily difficult to use metadata elements from both standards in a single application profile. This activity will approach part of this situation by producing a recommendation for using IEEE LOM and Dublin Core term in combination in Dublin Core applications. This represents a partial and short-term solution to the overall issue, which will still be of great value in the short to medium to implementers that are struggling with these metadata interoperability issues. The recommendation will also be of great value in the longer-term process of trying to align the abstract models of IEEE LOM and Dublin Core, as it will provide an analysis of fundamental incompatibilities between the two models.

<http://dublincore.org/education/wiki/DCMI/IEEE/TSCTaskforce?action=AttachFile&do=get&target=LOM-DCAM-Charter.pdf>



Standards “based on” Dublin Core

- International Press and Telecommunications Council – IPTC Core
- Public Broadcasting Metadata Dictionary Project – PB Core
- European Broadcasting Union – EBU Core
- Publishing Requirements for Industry Standard Metadata – PRISM
- International Digital Publishing Forum eBook Standard – EPUB
- Dutch Government Metadata Standard – OWMS
- Europeana Semantic Elements – ESE



Mappings and crosswalks

- Various flavours of MARC:
UNIMARC, MARC21, MARCXML
- MODS – Library of Congress
Metadata Object Description
Schema
- CIDOC CRM – Cultural Heritage
documentation reference model
- ISAD(G) – Archival descriptions
- CCF – UNESCO Common
Communications Format
- DDI – Data Documentation
Initiative; social science data
- INSPIRE – EU spatial data
infrastructure
- FGDC – US National Spatial Data
Infrastructure
- MPEG-7 – Multimedia content
- JPX – JPEG 2000 extended
image file format syntax



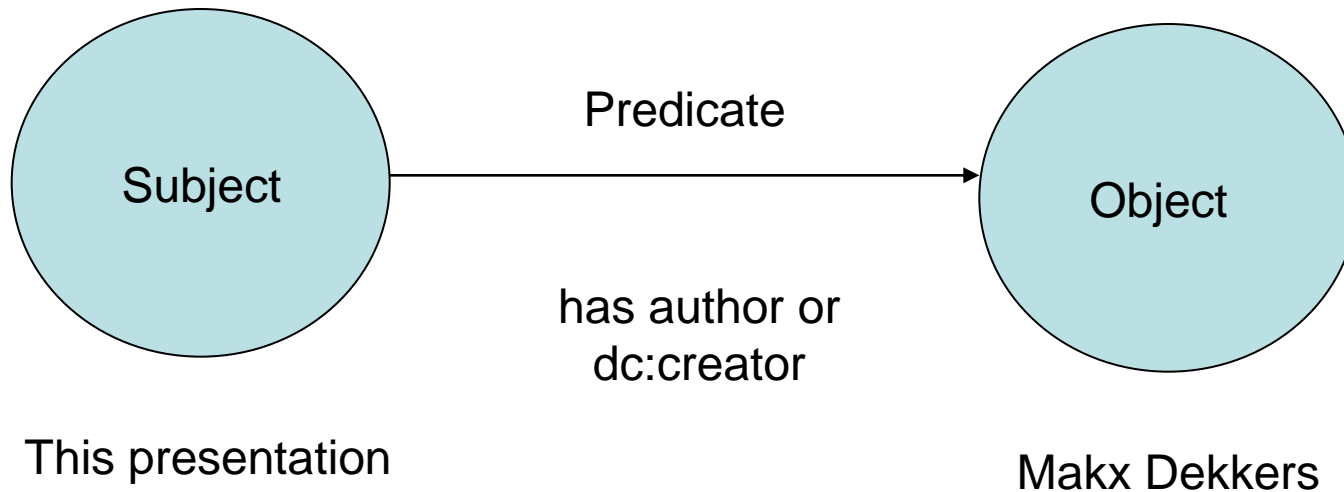
The Semantic Web vision

- Interpretation and processing of metadata by machines
- Rather than building silos of shared meaning, build a common framework for metadata
- Enabling linking of data across products, domains, boundaries
- Distributed publishing and maintenance



Building on RDF

- The main building block:





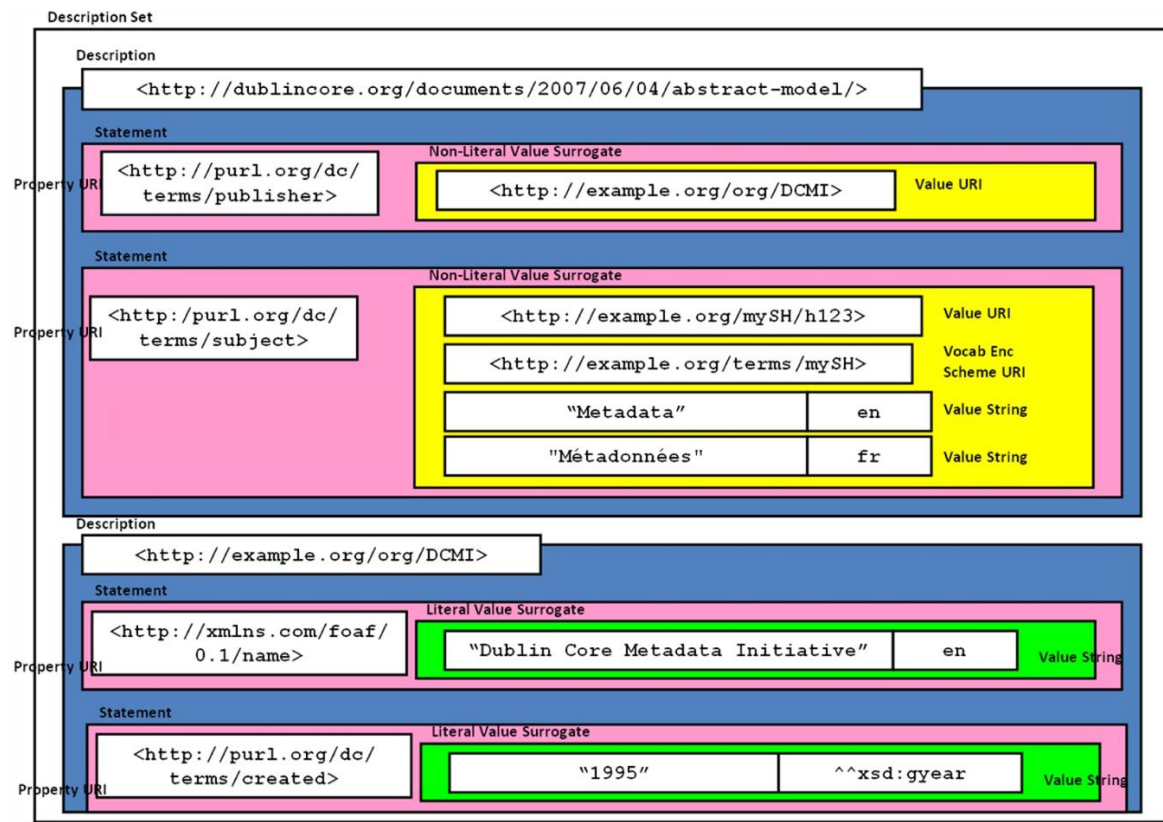
Adding structure

➤ The DCMI Abstract Model

- Formal description of relationships between properties, values, encoding schemes
- Compatible with Semantic Web RDF model
- Independent of syntax
- Specification using UML (Unified Modelling Language) class diagrams
- Recommendation March 2005, revised June 2007



DCMI Abstract Model





The Singapore Framework

- Providing a formalized approach to declaring how an application uses metadata (Dublin Core Application Profiles)
 - Functional Requirements
 - Domain Model
 - Description Set Profile
 - Usage Guidelines
 - Encoding Syntax Guidelines



Linked Data: RDFa

- An RDF-based “microformat” for embedding metadata in HTML documents
- Benefits
 - Publisher independence – choose your own standard
 - Data reuse – no duplication (header/body)
 - Self-containment – separation of HTML and RDF
 - Schema modularity – reuse of attributes
 - Evolvability – add fields as necessary



Use of Dublin Core metadata in Linked Data

- **Mixing and matching metadata vocabularies**
 - Dublin Core terms
 - FOAF
 - vendor-specific vocabularies
- **Examples**
 - Google Rich Snippets
 - Yahoo! SearchMonkey



RDFa at UK government



News | Events | Contact us | Accessibility | Login | Register

Search

Civil Service home About the Civil Service Networks Pensions Cr

- Jobs FAQ's
- Job Search Results
- Jobseeker Profile
- Job-share noticeboard
- Fast Stream
- Undergraduates/Graduates
- Non Graduates/Direct Entry
- Experienced Professionals
- Professional Entry Routes
- Departments and Accredited NDPBs
- Civil Service Commissioners
- Employee profiles
- Office of the Commissioner for Public Appointments
- Opportunities in Europe
- Nationality Requirements
- Rewards and Benefits
- Secondment Opportunities
- Summer Diversity Internships

Vacancy Details
[Back to results](#)

```
<div id="jobAdvert" xmlns:dc="http://purl.org/dc/terms/" xmlns:arg="http://purl.oclc.org/argot/"
xmlns:foaf="http://xmlns.com/foaf/0.1/" typeof="foaf:Document">
<span rel="dc:publisher" href="http://www.civilservice.gov.uk/"></span>
<span rel="dc:type" resource="[arg:Vacancy]"></span>
```

```
<h4>Salary</h4>
<p><span property="arg:salaryFrom" content="22533">£22,533 p.a. pro
rata (includes London Weighting)</span><span property="arg:salaryTo"
content="22533"></span> </p>

<h4>Location</h4>
<p property="dc:coverage">City of London, Greater London,
London<br /></p>

<h4>Closing Date</h4>
<p property="dc:valid" content="2009-05-29">29 May 2009</p>

<h4>Interviews start</h4>
<p property="arg:interviewsStart">10 June 2009</p>
```



Back to basics: open interoperability

- RDFa provides a technology to include metadata in resources
- A potential approach for the open Web
- A tool for Dublin Core's original objective: a common way to describe Web resources
- Dublin Core can be used as a core set of properties RDF/RDFa



Linked (Open/Enterprise) Data

- What can we learn from the last 15 years?
- There are two environments:
 - Open interoperability for resources in the public domain (or with some other requirement to share)
 - Controlled interoperability within the enterprise or partnership or supplier/customer relationships
- Linked Data provides an approach for both – crossing the boundary as and when needed



Conclusions

- Dublin Core was envisaged as “Core” set of descriptors for the Web
 - Main usage evolved in “controlled environments”
 - Cooperation with other initiatives has always been an important aspect of the work of DCMI
 - Linked Data promises to achieve global interoperability through mixing-and-matching of metadata vocabularies
-