Dublin Core: The Road from Metadata Formats to Linked Data

Joint NISO/DCMI Webinar
25 August 2010
Makx Dekkers and Thomas Baker

Dublin Core in the Early Web Revolution
Makx Dekkers, Managing Director and CEO, DCMI

First steps
- In October 1994, informal discussion at second Web Conference, Chicago
- Identified a need for a “core” set of descriptors to help discover content on the Web
- 1-3 March 1995, OCLC/NCSA workshop in Dublin, Ohio at OCLC Headquarters

Dublin Core: The Road from Metadata Formats to Linked Data

Joint NISO/DCMI Webinar
25 August 2010
Makx Dekkers and Thomas Baker

Dublin Core in the Early Web Revolution
Makx Dekkers, Managing Director and CEO, DCMI

First steps
- In October 1994, informal discussion at second Web Conference, Chicago
- Identified a need for a “core” set of descriptors to help discover content on the Web
- 1-3 March 1995, OCLC/NCSA workshop in Dublin, Ohio at OCLC Headquarters

Dublin Core: The Road from Metadata Formats to Linked Data

Joint NISO/DCMI Webinar
25 August 2010
Makx Dekkers and Thomas Baker

Dublin Core in the Early Web Revolution
Makx Dekkers, Managing Director and CEO, DCMI

First steps
- In October 1994, informal discussion at second Web Conference, Chicago
- Identified a need for a “core” set of descriptors to help discover content on the Web
- 1-3 March 1995, OCLC/NCSA workshop in Dublin, Ohio at OCLC Headquarters

Dublin Core: The Road from Metadata Formats to Linked Data

Joint NISO/DCMI Webinar
25 August 2010
Makx Dekkers and Thomas Baker

Dublin Core in the Early Web Revolution
Makx Dekkers, Managing Director and CEO, DCMI

First steps
- In October 1994, informal discussion at second Web Conference, Chicago
- Identified a need for a “core” set of descriptors to help discover content on the Web
- 1-3 March 1995, OCLC/NCSA workshop in Dublin, Ohio at OCLC Headquarters

Dublin Core: The Road from Metadata Formats to Linked Data

Joint NISO/DCMI Webinar
25 August 2010
Makx Dekkers and Thomas Baker

Dublin Core in the Early Web Revolution
Makx Dekkers, Managing Director and CEO, DCMI

First steps
- In October 1994, informal discussion at second Web Conference, Chicago
- Identified a need for a “core” set of descriptors to help discover content on the Web
- 1-3 March 1995, OCLC/NCSA workshop in Dublin, Ohio at OCLC Headquarters
Dublin Core: the original idea

- 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
- Making it easier to find information on the Web as it develops (1995!)

1995: The Dublin Core

- “Core” set, simple enough for non-experts to understand and create
- A “library catalog card” for Web objects
- Based on consensus across domains
- Standardized:
  - ISO 15836:2003, revised 2009

1996: Modular metadata

- Not: “One size fits all”
- Different ways to describe one object:
  - MARC records for library catalogs
  - Dublin Core for simpler descriptions
  - Specialized metadata for terms and conditions of use
- Recognized need for a general framework for different types of metadata

1996: Towards metadata “frameworks”

- Warwick Framework: “packages” in “containers”
- Descriptive Dublin Core
- Administrative Dublin Core
- Rights Creative Commons

1997: Qualification to add precision

- Not just any Date, but a Date the resource was Created
- Not just any Subject, but a Library of Congress Subject Heading
- Dumb-down: ignore extra details to see just a “core” description

2000: Application Profiles

- Customized implementations
  - Use “the Dublin Core” with other vocabularies
  - Local rules and guidelines
- Application profile provides documentation so that others can follow
- Not “take it or leave it”, but “take what you want, create what you need”
Dublin Core usage

- The Web exploded (now over a trillion pages!)
- Search engines took care of the open Web
- Dublin Core metadata came to be used widely in "controlled environments"
  - As a basic description mechanism
  - As a basic exchange format
- But never intended to be a “complete” solution

Interoperability in the early Web

- No interoperability of data in the Open Web
  - Mostly pages with links for human navigation
- Controlled environments: “intra-operability” (cooperation between known partners)
  - But “intra-net” can be quite large, e.g., OAI-PMH
- Semantic Web/Linked Data intends to “open up” controlled environments (“silos”)
  - Global interoperability across silos through typed links

Linked Data basic principle

- Object (Resource)
- Subject (Resource)
- Predicate (property)

This presentation Has subject Dublin Core

Dublin Core principles and Linked Data

- Dublin Core principles (1995-2005)
  - One-to-one (describe one and one thing only)
  - Dumb-down
  - Appropriate values
- Corresponding Linked Data design principle
  - A statement is “about” a named resource
  - Sub-property relations
  - Choice between text strings and links to other resources

Dublin Core was one of the inspirations for RDF

Dublin Core development

- Started from a vision for the open Web (HTML)
- Came to be widely deployed in controlled environments (XML)
- Further developed since 2000 in conjunction with Semantic Web and Linked Data (RDF)
- From a “Core Metadata Element Set” for the Web to a “core vocabulary” for Linked Data

Questions?
What Makes the Linked Data Approach Different
Thomas Baker, Chief Information Officer, DCMI

“Open World Mindset”

- “Closed” (“normal”) IT environments
  - Data sources carefully controlled.
  - Data formats “custom-defined” for an application.
- Linked data based on an “open world mindset”
  - Integrating data from the open Web
  - Systems designed to incorporate new information incrementally
  - By design, tolerance of incomplete information

Source A: About a book by Barack Obama

Each bit of information about the book is expressed in the data as a triple – a three-part statement.

RDF – a grammar for the language of data

1. Describe resources using interrelated “statements” (“triples”).
2. Use URIs – unique, globally managed identifiers – as the “words” of statements.
Each triple in the data corresponds to a link in a conceptual graph.

Each bit of information about the book is expressed in the data as a triple – a three-part "statement".

http://isbn:14000082773 dcterms:ISSN "2004"

http://isbn:14000008273 dcterms:creator "Barack Obama"
Source A: About a book by Barack Obama

Each triple in the data corresponds to a link in a conceptual graph.

Each bit of information about the book is expressed in the data as a triple – a three-part “statement”.

http://isbn/978-14000082773 dct:title "Dreams from My Father"

http://isbn/978-14000082773 x:isTranslationOf http://isbn/978-2258075979
dct:title "Les rêves de mon père"

http://isbn/978-2258075979 x:translator _:placeholder

_:placeholder foaf:name "Danièle Darneau"

Source B: About the French translation

http://isbn/978-2258075979 x:isTranslationOf http://isbn/978-14000082773
dct:title "Les rêves de mon père"

http://isbn/978-14000082773 x:translator _:placeholder

_:placeholder foaf:name "Danièle Darneau"
Merge the two data sources into one set of triples
Merge the two data sources into one set of triples

Software detects matching URIs...

...and merges the data

“Enrich” the data with a DBPedia URI identifying Barack Obama...
“Enrich” the data with a DBPedia URI identifying Barack Obama...

In Wikipedia, a book in German about Obama is linked to the DBPedia URI

New York Times’ URI for Obama, linked to Dbpedia’s, leads to a NYT Topic Page

Example: BBC Wildlife Finder

Humboldt Squid page, pulled together from a diversity of Linked Data sources
Under the hood of the Squid page, triples use URIs to make connections…

Many basic relations are Dublin Core “predicates”

Or mappable equivalents of Dublin Core predicates
Using triples about many animals to pull together a Web page about "nocturnal" wildlife…

Questions?

Dublin Core Metadata Initiative
Designing Interoperable Metadata on Linked Data Principles
Thomas Baker, Chief Information Officer, DCMI

...from Functional Requirements and a Domain Model

What a metadata application must do
> "Support navigation between ‘versions’.”
> "Enable searching on the ‘owner’ of a collection.”
...to a Description Set Profile and Data Format

How things of the Domain Model are described
- Use of Properties and Classes.
- Their cardinality
- Whether mandatory or optional
- Lists of allowable values...

Structure of the actual data format used by an application
- Components of the data structure.
- How the components map to Linked Data.

...on the basis of Community Standards, which in turn are grounded in...

- Community Domain Models
- Using well-known entities, such as FRBR classes
- Metadata Vocabularies
- Using well-known, such as Dublin Core and FOAF

...in Foundation Standards for Linked Data (RDF).

DCMI guidelines, under review, offer one approach¹ to designing Linked-Data-compatible metadata records

* There is more than one way to do it. To be discussed at DC-2010 in Pittsburgh, October 2010, http://dc-2010.org.
“Closed” ("normal") IT: Integrate across silos by mapping ad-hoc data structures

New Linked Data approach: Diverse applications create good triples

Good triples can be merged coherently

Applications come and go…
Applications come and go…

The data remains

Good triples, based on known vocabularies such as Dublin Core, make data “self-descriptive”

Search engine

Queries

Questions?

Bridging the Gap to the Linked Data Cloud

This is not about starting all over

- A lot of metadata already exists
  - Formal collections: libraries, enterprise data
  - Embedded descriptions (e.g., captions, tags)
- A lot of metadata is being produced now, manually and automatically
  - Catalogue data, product descriptions
  - Device-generated data
- Linked Data to leverage what is already there
Metadata is where the meaning is

- Metadata adds meaning to resources
- Specific statements about specific data
- No Linked Data without meaning
- Technology is a tool, not the goal
  - Linked data technology may (will) change
  - The meaning of data is anchored in the real world
  - Any implementation will need to allow evolution

"Open World" assumption does not mean all data must be available without restrictions
- Private data will have to remain private
  - Personal information, commercial secrets
- Linked Open Data vs. Linked Enterprise Data
  - Both types of implementation are useful
  - Open paradigm allows data to be shared more easily if and when necessary

Convert or expose?

- Existing metadata approaches are based on business needs and will therefore live on
- Consider future needs for flexibility and sharing
- There is a choice:
  - Create RDF from existing database format and expose (e.g., for harvesting, get better ranking)
  - Migrate database to triple store with emerging tools (maximum flexibility in return for investment)

Data

Manage

Expose

Convert

Metata

Data

Triple store

Database

Manage

Expose

Convert

Metadata

Data

Triple store

Database
Using Linked Data vocabularies

- Linked Data provides enormous flexibility
- A choice to be made:
  - Define your own vocabulary
  - Use existing ("standard") vocabularies
- Using trusted vocabularies reduces complexity for interoperability
- Well-known vocabularies:
  - Dublin Core, Good Relations, FOAF, SKOS

DCMI trusted operational principles

- Development and maintenance of the Dublin Core vocabulary, "DCMI Terms"
  - Open consensus building
  - International scope and participation
  - Neutrality of purposes and business models
  - Neutrality of technology
  - Cross-disciplinary focus
- Unrestricted use of specifications and guidelines

DC-2010

- Tenth International Conference on Dublin Core and Metadata Applications
  - Pittsburgh Hilton, 20-22 October 2010
- Sessions and presentations on:
  - Modularity, migration, models, communities
  - Practical examples, tools, knowledge management
  - … and a lot of Linked Data
- http://dc-2010.org/

Dublin Core: The Road from Metadata to Linked Data

Questions

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

www.niso.org/news/events/2010/dublincore