Cornell University

Basic Dublin Core Semantics

DC 2006 Tutorial 1, 3 October 2006

Marty Kurth

Head of Metadata Services

Cornell University Library

Getting started

- Let’s introduce ourselves
- Let’s discuss our expectations for the tutorial

We'll cover

- An introduction to metadata
- Key features of the Dublin Core
- Dublin Core metadata in a broader context
- Important aspects of the DCMI community

Cornell University

An introduction to metadata

What is metadata?

A (hopefully) straightforward place to start:

Metadata consists of statements we make about resources to help us find, identify, use, manage, evaluate, and preserve them.

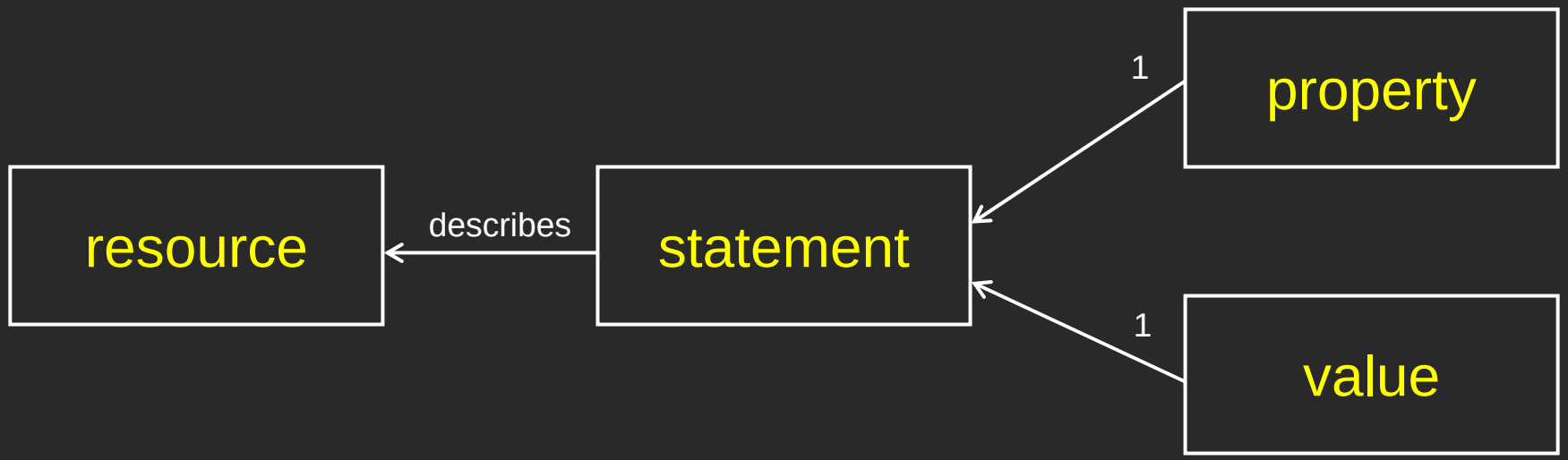
Some typical metadata functions

Discover resources	Manage documents	Control IP rights
Identify versions	Certify authenticity	Indicate status
Mark content structure	Situate geospatially	Describe processes

Metadata building blocks (in words)

1. The basic unit of metadata is a *statement*.
2. A statement consists of a *property* (aka, *element*) and a *value*.
3. Metadata statements describe *resources*.

Metadata building blocks (in pictures)



(An oversimplification of the DCMI abstract model for resources!)

What are the *properties* and *values* in these metadata *statements*?

245 00 $a Amores perros $h [videorecording]

<title>Nueve reinas</title>
<type>MovingImage</type>

Who cares about metadata?

The term “metadata” has meaning in contexts such as:

- Data modeling
- Library cataloging
- Internet/World Wide Web resource discovery
  - Led to a convergence between the first two
  - Formed the context in which Dublin Core arose

Cornell University

Key features of the Dublin Core

How and why did the Dublin Core come to be in 1995?

- Dramatic increase in the number of document-like resources on the net
- Slow improvement in indexing services made resources hard to discover
- Belief that descriptive metadata would improve discovery
- Perceived need for a descriptive standard that was simple to apply (by non-professionals)

Cornell University

Dublin Core Metadata Element Set

Creator	Title	Subject
Contributor	Date	Description
Publisher	Type	Format
Coverage	Rights	Relation
Source	Language	Identifier

Characteristics of the Dublin Core

- A flat element structure, with:
  - All elements optional
  - All elements repeatable
- Elements displayed in any order
- Extensible (elements, qualifiers)
- Syntax independent
- International
- Subject independent

Resources for which DC is often used

DCMI Type Vocabulary

Collection	Dataset	Event
Image	Interactive Resource	Moving Image
Physical Object	Service	Software
Sound	Still Image	Text

Dublin Core principles

- Dumb-down
- The one-to-one principle
- Appropriate values

Dumb-down

- Simple DC does not use *element refinements* or *encoding schemes* and statements only contain *value strings*
- Qualified DC uses features of the DCMI Abstract Model, particularly *element refinements* and *encoding schemes*
- *Dumbing-down* is translating qualified DC to simple DC (*property dumb-down* and *value dumb-down*)
- For more info, see the DCMI Abstract Model!

Element refinements

- Element refinements narrow the meaning of DC elements
  - *hasVersion* and *isVersionof* refine *relation*
  - *bibliographicCitation* refines *identifier*
- Element refinements are *properties*, so we typically render them independently
  - <dcterms:alternative>Nine queens</dcterms:alternative>
  - <dcterms:issued>2000-07-11</dcterms:issued>
- For more information on rendering DC terms, attend the Basic Syntax Tutorial!

Encoding schemes

- *Vocabulary encoding schemes*
  - Indicate that a value comes from a controlled vocabulary (e.g., that "Spanish American literature" is an LCSH term)
- *Syntax encoding schemes*
  - Indicate that a string is formatted in a standard way (e.g., that "1956-11-12" follows ISO 8601)
- DCMI recommends using encoding schemes with *coverage*, *date*, *format*, *language*, *subject*, and *type*
- For more information on rendering encoding schemes, attend the Basic Syntax Tutorial!

The one-to-one principle

- Create one metadata *description* for one and only one resource
  - E.g., do not describe a digital image of the Mona Lisa as if it were the original painting
- Group related *descriptions* into *description sets*
  - I.e., describe an artist and his/her work separately, not in a single description

Appropriate values

- Use elements and qualifiers to meet the needs of your local context, but . . .
- Remember that your metadata may be interpreted by machines and people, so . . .
- Consider whether the values you use will aid discovery outside your local context and . . .
- Make decisions about your local practices accordingly

Origins of the DCMI Abstract Model

- DC community realized early that machine-processing requires a coherent data model
- (1996) “Warwick Framework” proposed at DC-2: Saw DC as one “metadata package” among others
- (1997) Qualifiers proposed for specifying meanings
- (1998) DCMI functional requirements inform W3C Resource Description Framework (RDF) and RDF informs early Dublin Core data model
- (2000) First set of qualifiers officially approved

DCMI Abstract Model (2005)

- Defines *resources* in terms of semantic relationships among *classes*, *properties*, and *values*
- Defines a model for DCMI *descriptions*, *description sets*, and *records*
- Serves as a foundation for future DCMI developments
- Serves as a conceptual model for metadata initiatives outside DCMI

DCMI namespaces and policies

- All DCMI terms have unique identity within three *namespaces*:
  - http://purl.org/dc/elements/1.1/ - DCMES (15 elements)
  - http://purl.org/dc/terms/ - DCMI elements and qualifiers
  - http://purl.org/dc/dcmitype/ - DCMI Type Vocabulary
- DCMI *namespace policies* promote long-term stability of namespace URIs
  - Only substantial semantic changes will result in change of namespace URIs

Cornell University

Dublin Core metadata in a broader context

Metadata creation and distribution models

- *Federation*
  - Extensive specifications, standards, protocols, training
- *Harvesting*
  - Basic agreements, reliance on best practices
- *Gathering*
  - Automated indexing of content, algorithms yield results from search terms, less likely to use descriptive metadata per se

Harvesting model key features

- Integrating metadata from many sources calls for common element sets, record structures, and harvesting protocols
- Open Archives Initiative Protocol for Metadata Harvesting serves as a framework for sharing metadata and mandates ‘simple DC’ as a common metadata format
- Harvesting promotes metadata reuse
- Best practices balance cost and interoperability
- Communities add value to basic infrastructure (more complex metadata, new uses for protocol)

Application profiles & interoperability

Application profiles enable:

- Implementers to use DC metadata in conjunction with non-DC metadata
- Implementers to benefit from the experience of their peers
- Communities to harmonize metadata usage for greater interoperability

For more information on application profiles, attend the Application Profiles Tutorial!

Cornell University

Important aspects of the DCMI community

Dublin Core grows and changes

- DCMI emphasizes *open participation*
  - Conferences, working groups, discussion lists
- DCMI term set evolves as implementers coin new terms and usage patterns emerge
- DCMI Usage Board reviews proposals for new metadata terms

Dublin Core Usage Board

- Considers proposals for new *terms* (elements, refinements, encoding schemes, DCMI Type Vocabulary terms)
  - Evaluates proposals in light of *grammatical principle*, *semantic clarity*, *usefulness*, and *overlap* with existing terms
- Evaluates constructs that use DCMI terms, such as *application profiles*

Find out more about DC

- *DCMI Web Site*
  - http://dublincore.org
- *Using Dublin Core*
  - http://dublincore.org//documents/usageguide/
- *DCMI Abstract Model*
  - http://dublincore.org/documents/abstract-model/
- *DCMI Working Groups*
  - http://dublincore.org/groups
- *AskDCMI*
  - http://askdcmi.askvrd.org/

Cornell University

Questions?

Or please send questions to:

Marty Kurth
Cornell University Library
mk168@cornell.edu

Acknowledgement

Many thanks to Diane Hillmann for sharing the slides from an earlier version of this tutorial with me.

Interested in DC origins and history?

Lagoze, Carl. “The Warwick Framework: A Container Architecture for Diverse Sets of Metadata.” *D-Lib Magazine* July/August 1996. <http://www.dlib.org/dlib/july96/lagoze/07lagoze.htm>

Miller, Eric J. “An Overview of the Dublin Core Data Model.” 6 June 1999. <http://dublincore.org/1999/06/06-overview/>

Weibel, Stuart. “Metadata: The foundations of resource description.” *D-Lib Magazine* July 1995. <http://www.dlib.org/dlib/July95/07weibel.html>

Weibel, Stuart L, and Carl Lagoze. “An element set to support resource discovery: The state of the Dublin Core: January 1997.” *International Journal on Digital Libraries* 1 (1997): 176-86.