

An Introduction to Dublin Core

Diane I. Hillmann

National Science Digital Library

DC2004 Tutorial, 11 October 2004

What is Metadata?

- Answers come from three traditions:
 - Database Management Systems (“Schemas of relational databases”)
 - Library Cataloging Traditions (MARC & AACR2)
 - The World Wide Web (since the mid-1990’s)
 - The context for Dublin Core

Types of Metadata

- Administrative
- Descriptive
- Access/Use
- Preservation
- Technical/Structural
- Other?

Comparing Libraries & the Web

- Library tradition:
 - More than 100 years of experience, from catalog cards to MARC
 - MARC Records include:
 - Description (all objects)
 - Subjects and classification (topical context)
 - Holdings information (location)
 - Administrative information required to manage records
 - Distributed metadata creation based on common consensus on standards

The Web: One distributed Library?

- The situation in the mid-1990s:
 - Thousands of information providers, using a variety of metadata schemas (if anything at all)
 - Search engines providing too many hits and very little precision
 - Volatile resources changing addresses, disappearing, etc.
 - Exponential growth in numbers and types of resources available

Fifteen Core Elements (1996)

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

Functions of Metadata

Discover
resources

Manage
documents

Control IP
Rights

Identify
versions

Certify
authenticity

Indicate
status

Mark content
structure

Situate
geospatially

Describe
processes

Characteristics of the Dublin Core

- All elements optional
- All elements repeatable
- Elements may be displayed in any order
- Extensible
- International in scope

Moving Towards a Data Model

- Collective realization that machine-processability requires a coherent data model
- 1996: “Warwick Framework” proposed at DC-2 workshop: DC as one specialized module (“resource discovery”) among many
- 1997: “Qualifiers” proposed for specifying meanings
 - Some early adopters take this to unintended extremes: “DC.Creator.telephone-number”
- 1998: DCMI involvement in emerging Resource Description Framework and clarification of simple data model for Dublin Core
- 2000: First set of qualifiers officially approved

DC Data Model Work-in-Progress

- Provides explicit definitions of resources
- Relates DC principles and practices to the developments outside DCMII
- Makes clear the relationship of DC “packages” of information to other metadata “packages”
- Paves the way for future progress for DCMII

Dublin Core Principles

- Dumb-Down
- One-to-One
- Appropriate Values

Dumb-Down

- The fifteen core elements are usable with or without qualifiers
- Qualifiers make elements more specific:
 - Element Refinements narrow meanings, never extend
 - Encoding Schemes give context to element values
- If your software encounters an unfamiliar qualifier, look it up – or just ignore it!

The One-to-One Principle

- Describe one manifestation of a resource with one record
 - Ex.: a digital image of the Mona Lisa is not described as if it were the same as the original painting
- Separate descriptions of resources from descriptions of the agents responsible for those resources
 - Ex.: email addresses and affiliations of creators are attributes of the creator, not the resource

Appropriate Values

“Best practice for a particular element or qualifier may vary by context, but in general an implementor cannot always predict that the interpreter of the metadata will always be a machine. This may impose certain constraints on how metadata is constructed, but the requirement of usefulness for discovery should be kept in mind.”

-- from *“Using Dublin Core”*

Simple or Qualified Dublin Core?

- Simple Dublin Core is limited to the original 15 elements
- Qualified Dublin Core includes, in addition:
 - New Elements
 - Element Refinements
 - Value Encoding Schemes

Element Refinements

- Make element meanings narrower, more specific:
 - a *Date Created* versus *Date Modified*
 - an *IsReplacedBy* versus *Replaces* Relation
- Depending on syntax chosen, refinements may appear as stand-alone tags instead of with elements:
 - `<dct:created>2002-10-04</dct:created>`, instead of:
 - `<dc:date><dct:created>2002-10-04 </dct:created></dc:date>`
 - Requires a schema to dumb-down *Date Created* to *Date*
 - Dublin Core is simple enough to support both usages

Value Encoding Schemes

- Indicate that the value is:
 - a term from a controlled vocabulary (e.g., Library of Congress Subject Headings)
 - a string formatted in a standard way (e.g., that "05/02" means May 2nd, not February 5th)
- Even if a scheme is not known by software, the value should be "appropriate" and usable for resource discovery.

Dublin Core Grows and Changes

- DCMI Community 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

- Usage Board receives proposals for new elements, refinements, encoding schemes, Type terms
 - Evaluates proposals in light of grammatical principle, usefulness, clarity of definition, overlap with existing terms
- Tiered model of approval status: conforming, recommended, obsolete, registered

DCMI Namespaces and Policies

- All DCMI metadata terms are given unique identity within three namespaces:
 - <http://purl.org/dc/elements/1.1/> - the legacy DC-15
 - <http://purl.org/dc/terms/> - all other elements/qualifiers
 - <http://purl.org/dc/dcmitype/> - a Type vocabulary
 - Example: <http://purl.org/dc/elements/1.1/title>
- Policies promote long-term stability of namespace URIs
 - Changes not substantially “semantic” (i.e., corrections) will not result in change of namespace URIs

Application Profiles & Interoperability

- Implementers want to know how their peers design metadata – to avoid "reinventing the wheel"
- Information providers need to harmonize metadata usage for improved access within domains, e.g.:
 - Between countries (Nordic Metadata Project)
 - Preprint repositories (Open Archives Initiative)
 - Subject gateways (Renardus)
 - Mathematics and physics (MathNet, PhysNet)

Encoding Metadata Records

- Mid-1990s: HTML tags embedded in Web pages
 - Simple, easy to deploy, but inflexible, hard to maintain
 - Bad tags like *DC.Creator.eyecolor* imply a non-existent support for nesting and for entity distinctions
- 2000+: Better XML/RDF alternatives
 - RDF metadata supports complex structures without breaking simple DC grammar
 - Open Archives Initiative promotes mass adoption of an XML schema for simple, unqualified Dublin Core records – along with a protocol to make them available

The World of Metadata Harvesting

- **Service Providers** harvest and integrate metadata from diverse **Content Providers**
 - Presupposes standard element sets, record formats, and harvesting methods
- 1999+: Open Archives Initiative began as a federation of scholarly pre-print providers
 - Today: aggregations using OAI-PMH creating new models of metadata services

Why the Harvesting Model Works

- Exposing metadata facilitates:
 - Reuse of available metadata
 - The creation of value-added services
- Low entry cost is key to deployable digital library infrastructure
 - New Static Repositories version of OAI lowers entry bar even more
- Individual communities have begun to customize the common infrastructure
 - Branching out from Simple DC minimum

Example: National Science Digital Library (USA)

- Principle: “Metadata is too expensive to create much. Use existing metadata when possible.”
- Early strategy:
 - Support standard formats, base Metadata Repository on Qualified Dublin Core
 - Collect DC and native metadata (when possible)
 - Assemble all metadata in a central repository
 - Expose all such records to harvesters
 - Focus limited human effort on metadata for collections
 - Generate metadata automatically when possible

NSDL Search Results for

(Displaying results: 1 - 20 of 60) [Next Results](#) >>>

Title/Description	Resource Format	Found in Collection
<p>Plant nematodes the grower should know [electronic resource] / G. Steiner. [Reprinted from the Fourth proceedings of the Soil Science Society of Florida]. more info [Archived Version]</p>		
<p>The evolutionary position of nematodes Abstract Background The complete genomes of three animals have been sequenced by global research eff ... more info [Archived Version]</p>		
<p>Biodiversity of nematode assemblages from the region of the Clarion-Clipperton Fracture Zone, an area of commercial mining interest Abstract Background The possibility for commercial mining of deep-sea manganese nodules is currently ... more info [Archived Version]</p>		
<p>Nematode Movies This web page presents four videos of nematodes that show their twisting movement. These videos are ... more info [Archived Version]</p>	[text]	
<p>Analysis and functional classification of transcripts from the nematode Meloidogyne incognita Abstract Background Plant parasitic nematodes are major pathogens of most crops. Molecular character ... more info [Archived Version]</p>		
<p>The Effect of Nematode Isolate and Soil Environment on the Tobacco cyst Nematode (Globodera tabacum solanacearum), a Pathogen of Flue-cured Tobacco and Other Solanaceous Crops more info [Archived Version]</p>	[text] [pdf]	

More Information

Title	Nematode Movies
Creator	Michael W Davidson
Subject	Life Science
Subject	Animals
Subject	Microorganisms
Subject	Microscopy
Subject	Worms
Description	<p>This web page presents four videos of nematodes that show their twisting movement. These videos are from a collection of digital movies of microscopic organisms taken at a typical North Florida pond. They were taken under darkfield illumination at magnifications 100X and 200X. Users may access these videos, ranging from 21.1 seconds to 45.4 seconds, by using a playback format at the following connection speeds: 28.8k (modem), 56.6k (modem), or T1/Cable/DSL; or downloading the video clip in .AVI format. (Author/LSR) Copyright 2003 Eisenhower National Clearinghouse (ENC).</p>
Publisher	Florida State University
Contributor	Michael W Davidson
Date	2003
Format	text/html
Identifier	http://micro.magnet.fsu.edu/moviegallery/pondscum/nematode/index.html
Language	en
Rights	Copyright 1995-2002, by Michael W. Davidson and The Florida State University. All Rights Reserved.

NSDL Forges Ahead

- Develops methodology for evaluating metadata quality
 - “Metadata must play well together”
- Builds techniques for normalization that work with large quantities of diverse data
- Moving beyond managing metadata records to managing “descriptive statements about resources”

Steps toward the Semantic Web

- Simple linked data model
 - a Web link means “has something to do with”
 - Formalized since 1997 as [Resource Description Framework \(RDF\)](#)
- XML: universal data exchange format
- URIs: everything has a unique address
 - Both resources and the metadata terms used to describe them
 - XML namespaces: unique identifiers for metadata vocabulary terms
 - URIs as anchors for data merging

Finding Out More about DC

- DCMI Web Site
 - <http://dublincore.org>
- “Using Dublin Core”
 - <http://dublincore.org//documents/usageguide/>
- Participating in a Working Group
 - <http://dublincore.org/groups>
- Ask a question!
 - <http://askdcmi.askvrd.org/>

Questions?

Thank you for your attention!

Diane I. Hillmann

Director of Library Services and Operations, NSDL

Member, DDCMI Usage Board, Advisory Board

Editor, “Using Dublin Core”

Administrator, AskDCMI

DIH1@cornell.edu (fourth character is “one”)