International Conference on Dublin Core and Metadata Applications

Tutorial 1: Dublin Core History and Basics

22 September 2008

Jane Greenberg, Associate Professor
Univ. of North Carolina at Chapel Hill
janeg@email.unc.edu
Overview

• Introduction
• Metadata
  – Metadata defined—pictures, definitions, types
• Dublin Core (history & growth)
  – Dublin Core Metadata Element Set, Version 1.1
    • History workshop → full conference
    • Founding principles (characteristics)
    • Principles guiding Dublin Core metadata creation
  – Guidelines
    – Toward a more Semantic Web… more guidelines
• Dublin Core Metadata Initiative (DCMI)
• Q&A
Introduction

• Who is here?
  √ librarians/archivist/museum professionals
  √ scientists
  √ government specialists
  √ educators
  √ business/corporate workers
  – other?

• Dublin Core
  – Using DC; planning to use DC; want to learn more

• Different disciplines
Metadata defined

- data about data
- information about data
English literature--Early modern, 1500-1700--History and criticism.

O'Connor, John Joseph, 1918 (June 15)–
ix, 308 p. facsims. 25 cm. 9.00
Bibliography: p. 287–293.

1. Amadis de Gaula. 2. English literature—Early modern (to 1700)—History and criticism.

PQ6277.O2
SBN 8135-0622-0
Library of Congress 70 [4]
Family: Pinaceae
Species: Pinus serotina
Date identified: 1958-05-10
County: Pasquotank County
Location collected: Woodland Border, 2.3 miles north east of Nisonton
Collected by: Harry E. Ahles

<Species> Pinus serotina </Species>
<Date.ID><scheme=SPEC.W3C DTF“>1958-05-10</Date.ID>
MS WORD dialog box and output

Title: Functionalities for Automatic-Metadata Generation
Subject: automatic metadata generation
Author: Jane Greenberg, Kristina Spurgin, Abe Crystal
Manager: Jane Greenberg
Company: UNC-CH
Category: AMeGA project
Keywords: metadata, generation, automatic, semi-automatic
Comments: This is a draft circulated for editing

22.09.2008
Dublin Core metadata

<dc:title>Analyzing Metadata for Effective Use and Re-Use</dc:title>
<dc:creator>Naomi Dushay</dc:creator>
<dc:creator>Diane I. Hillmann</dc:creator>
<dc:subject xsi:type="dcterms:lcsh">Metadata</dc:subject>
<dc:subject xsi:type="dcterms:lcsh">Evaluation</dc:subject>
<dc:date xsi:type=dcterms.W3CDTF" content="2003--11" dc:date>
<dcterms:abstract>Using a commercially available visual graphical..</dcterms:abstract>
Metadata

- Structured, descriptive information about a resource (DCMI Glossary; Weibel, 1995)
- Data about the content, quality, condition, and other characteristics of data (FGDC Glossary, 1992)
- Additional information for the data it describes to support “use” (Wikipedia, 2008)
- Structured data about an object that supports functions associated with the designated object (Greenberg, 2003)
Some typical metadata functions

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

schemes and properties (elements)
# Metadata Functions and Classes

<table>
<thead>
<tr>
<th>Typology of 7 types of metadata</th>
<th>Metadata Functions</th>
<th>Element examples*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification / description metadata</td>
<td>RESOURCE DISCOVERY / INFORMATION RETRIEVAL</td>
<td><strong>Creator</strong> (Author), Title, Subject</td>
</tr>
<tr>
<td>Administrative metadata</td>
<td>RESOURCE MANAGEMENT</td>
<td>Price, Condition</td>
</tr>
<tr>
<td>Terms and conditions metadata</td>
<td>RESOURCE USAGE</td>
<td>Rights, Reproduction restrictions</td>
</tr>
<tr>
<td>Content ratings metadata</td>
<td>RESOURCE USE BY APPROPRIATE AUDIENCES</td>
<td>Audience</td>
</tr>
<tr>
<td>Provenance metadata</td>
<td>RESOURCE AUTHENTICATION AND OTHER PROVENANCE-RELATED ACTIVITIES</td>
<td><strong>Creator</strong>, Source</td>
</tr>
<tr>
<td>Linkage / relationship metadata</td>
<td>RESOURCE LINKING WITH RELATED RESOURCES</td>
<td>Relation, Source</td>
</tr>
<tr>
<td>Structural metadata</td>
<td>RESOURCE HARDWARE AND SOFTWARE NEEDS</td>
<td>Compression ratio</td>
</tr>
</tbody>
</table>

(Lagoze, 1996; mapped in Greenberg, 2005)
Dublin Core history and growth
Internet / WWW

- 1970 ARPANET begin operation
- 1991 Gopher released by Univ. of Minnesota
- 1991 WWW released by CERN
- 1993, Lycos (first engine to achieve commercial success)
- 1994, WebCrawler first full-text Web search engine
- Late 1990s Web directories become popular
- 1998 Google, relevancy ranking

Cataloging/Metadata

- c.1450 printing press
- 18th c. French cataloging code
- 1876 Cutter’s catalog objects
- 1960/61 Lubetsky/Paris principles
- 1994, 2nd WWW Conference in Chicago
  ➢ need for a metadata “core”: a small, common set of metadata elements to describe Web content
- 1995, NCSA/OCLC workshop in Dublin, Ohio
- 1998 FRBR
Dublin Core (DC) history

DC 1: March, 1995, Dublin, Ohio
• Identified 13 core elements (properties) essential for resource discovery on the Internet

• Title
• Author
• Subject
• Publisher
• Other Agent

• Date
• Object Type
• Form
• Identifier

• Relation
• Source
• Coverage
• Rights Management
Dublin Core (DC) history

DC 2: April, 1996, Warwick, England
  • Established the Warwick Framework
  • Establish a syntax for the DC...HTML tags for embedding in web documents (HTML 2.0)

✓ DC 3: September, 1996, Dublin, Ohio
  • DC elements extended to include digital images, 15 properties

✓ DC 4: March 1997, Canberra, Australia
  • Minimalists and structuralists
  • Canberrra qualifiers (meta-meta)
Dublin Core Elements, version 1.1

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

<table>
<thead>
<tr>
<th>Property</th>
<th>Refinement</th>
<th>Outside standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Alternative title&lt;br&gt;Subtitle</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Created&lt;br&gt;Modified</td>
<td>W3CDTF YYYY-MM-DD</td>
</tr>
<tr>
<td>Subject</td>
<td></td>
<td>LCSH, AAT, LCC</td>
</tr>
</tbody>
</table>

Smarter metadata, smarter for machines
Dublin Core (DC) history

DC 5: October 1997, Helsinki, Finland
• Finnish finish, endorse RDF
• More working groups, Include physical objects

• Share implementations

DC 7: October, 1999, Frankfurt, Germany
• Agree to finalize qualification for the DC schema
Dublin Core (DC) history

DC 8: October, 2000, Ottawa Canada
• Business/commerce community

DC 9: October, 2001, Tokyo, Japan
• Conference series, proceedings

DC 10: October, 2002, Florence, Italy

• Open lab
Dublin Core (DC) history

DC 12: October, 2004, Beijing, China

DC 13: Sept. , 2005, Madrid, Spain
• Emphasis on metadata vocabularies

DC 14: August-Sept. , 2006, Mexico
• Spanish speaking population

DC 15: August-Sept. , 2007, Singapore
• Application profiles, DCMI and the NLB
Founding principles

• Guiding the development and growth of DC
DC founding principles (characteristics)

- Simplicity
- Semantic interoperability
- International consensus
- Interdisciplinary
- Extensibility
- Modularity
TITLE: Introduction to cataloging and classification
CREATOR: Wynar, Bohdan S.
CREATOR: Taylor, Arlene G.
DATE: 1992
FORMAT: book
LANGUAGE: en
PUBLISHER: Libraries Unlimited
SUBJECT: Cataloging.
SUBJECT: subject cataloging.
SUBJECT: Classification -- Books
DESCRIPTION: 8th edition
TYPE: text.monograph
IDENTIFIER: (ISBN) 0872879674
RIGHT: Libraries Unlimited
RELATION: Library science text series
Dublin Core principles continued

• Semantic Interoperability
  – Semantic Interoperability Title = title = title
  • Do not conflict meanings of properties
    – Format ≠ method (glazed: sun baked, kiln fired)
  – Crosswalks

• International consensus
  – DC translated into multiple languages
    http://www.dublincore.org/resources/translations/

• Interdisciplinary
  – A wide range of projects using DC
Extensibility


Allows for extensions to metadata schema so as to accommodate the particular needs of any given application.

- Dublin Core
- PREMIS/Preservation Metadata
- ETD-MS (Electronic Theses and Dissertation)
- GEM (Gateway to Educational Materials)
- Other schemes, e.g., Darwin Core
Modularity  (Weibel discusses as Legos)

- Administrative metadata
- Descriptive metadata
- Use metadata
- Technical metadata
- Preservation metadata
Additional principles

• Guiding metadata creation
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 (DCAM), particularly element refinements and encoding schemes
- Dumbing-down is translating qualified DC to simple DC (property dumb-down and value dumb-down)
  - OAI (Open Archives Initiative)
<table>
<thead>
<tr>
<th>Property</th>
<th>Refinement</th>
<th>Outside standard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td>Alternative title, Subtitle</td>
<td></td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>Created, Modified</td>
<td>W3CDTF &lt;YYYY-MM-DD&gt;</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td></td>
<td>LCSH, AAT, LCC</td>
</tr>
</tbody>
</table>

Smarter metadata, smarter for machines
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
    - `<dc:creator>dig. image photographer</dc:creator>`
    - `<dc:date>YYYY-MM-DD</dc:date>`

- Group related descriptions into **description sets**
  - Description set for the Renaissance painting.. *Mona Lisa*
  - Description set for Leonardo Divincci

- Link via “source” property, DCAM
Appropriate values

- Use properties 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
Guidelines

• “a simple scheme in a complex world!”
Standardization

• The set of initial 15 elements (the Dublin Core Metadata Element Set) was proposed as international standard
  – Internet Draft RFC2413, 1998
  – European endorsement as a CEN Workshop Agreement, CWA 13874-2000
  – ISO international standard 15836-2003

• Updated versions:
  – Internet Draft RFC5013, 2007
  – Revision ISO standard 15836-2008 (in process)
Standards (Jane’s view)

- **Data structure standards**
  - (container, label, semantics)

- **Data communication standards**
  - (encoding, markup, data exchange)

- **Data value standards**
  - (authority files, ontologies, taxonomies, etc.)

- **Data syntax standards**
  - element ordering
  - grammar
  - content syntax
Key documents for starting with Dublin Core

1. DC-15 (Legacy scheme): http://purl.org/dc/elements/1.1/
2. DCMI Terms namespace: http://purl.org/dc/terms/ (all properties, refinements, valid encoding schemes, a Type vocabulary, and DCAM classes)
3. Type vocabulary: http://purl.org/dc/dcmitype/
   – Collection, dataset, event, image, service, text, etc.

Encoding guidelines "in use“

Guidelines, recommendations-2

Dublin Core Abstract Model

New encoding guidelines
It's a library, honey--kind of an early version of the World Wide Web.
Toward a Semantic Web

Founding principles
- Simplicity
- Semantic interoperability
- International consensus
- Interdisciplinary
- Extensibility
- Modularity

Key Ideas and approaches
- RDF (very simple)
- Metadata/data reuse
- Consistency
- One size does not fit all
- Global context
- Smarter web, smarter data
  - Machine processable
- Application profiles
DCMI

- DCMI
  - An initiative, a community
  - Committed to open standards, support interoperability
- DCMI Community emphasizes open participation
  - Conferences, communities, discussion lists/wikis
- DC structure
  - Preparing incorporation as non-profit organization in Singapore
  - Governance:
    - Board of Trustees: strategic and financial oversight
    - Directorate: executive, day-to-day management
    - Advisory Board: technical and operational advice
    - Usage Board: maintenance and review of proposals
  - Work by the architecture forum, communities, and task groups
Summary

Dublin Core (history & growth)
  • Dublin Core Metadata Element Set, Version 1.1
    – History workshop → full conference
    – Founding principles (characteristics)
    – Principles guiding Dublin Core metadata creation
  • Guidelines
    • Toward a more Semantic Web… more guidelines

Dublin Core Metadata Initiative (DCMI)

Q&A

Thank you!
janeg@ils.unc.edu

- Makx Dekkers
- Diane Hillmann
- Marty Kurth
- Marcia Zeng