International Conference on Dublin Core and Metadata Applications

Tutorial 1: Dublin Core History and Basics

22 September 2008

DC-2008
BERLIN

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
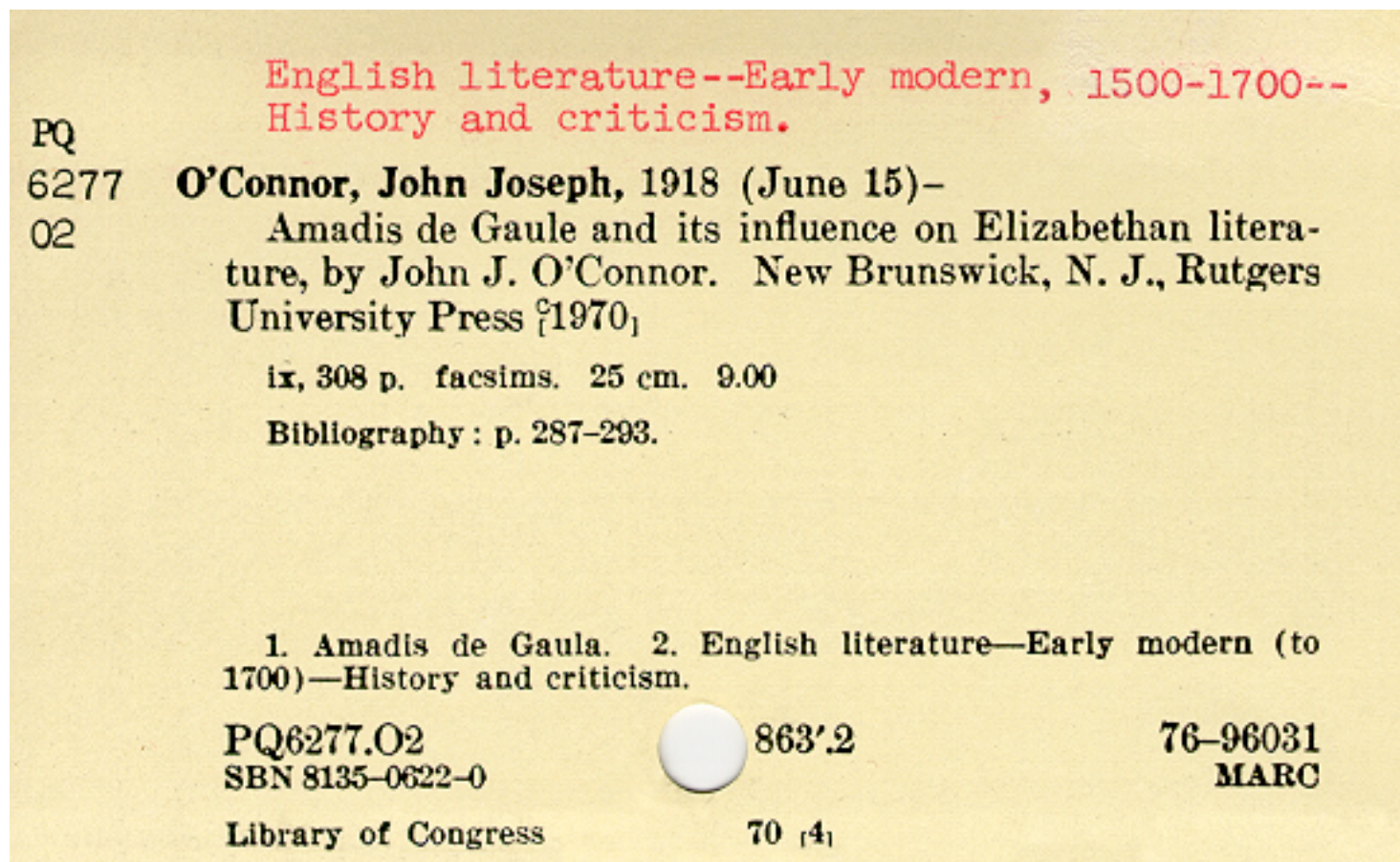
- Different disciplines

Metadata defined

- ……data about data
- …….information about data

Library catalog card

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

PQ
6277   O'Connor, John Joseph, 1918 (June 15)-
02   Amadis de Gaule and its influence on Elizabethan litera-
         ture, by John J. O'Connor. New Brunswick, N. J., Rutgers
         University Press ₍1970₎

         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   863'.2   76-96031
         SBN 8135-0622-0   MARC

         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>

Metadata Example for a specimen

☀ MS WORD dialog box and output

General | Summary | Statistics | Contents | Custom

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

Hyperlink base:

Template: Normal.dot

☐ Save preview picture

OK   Cancel

xmlns:st1="urn:schemas-microsoft-com:office:smarttags"⟨o:DocumentProperties⟩
⟨o:Subject⟩automatic metadata generation⟨/o:Subject⟩
⟨o:Author⟩Jane Greenberg, Kristina Spurgin, Abe Crystal⟨/o:Author⟩
⟨o:Keywords⟩metadata, generation, automatic, semi-automatic⟨/o:Keywords⟩
⟨o:Description⟩This is a draft circulated for editing⟨/o:Description⟩
⟨o:LastAuthor⟩Kristina M Spurgin⟨/o:LastAuthor⟩
⟨o:Revision⟩2⟨/o:Revision⟩
⟨o:TotalTime⟩13⟨/o:TotalTime⟩
⟨o:LastPrinted⟩2004-11-06T18:41:00Z⟨/o:LastPrinted⟩
⟨o:Created⟩2004-11-12T16:50:00Z⟨/o:Created⟩
⟨o:LastSaved⟩2004-11-12T16:50:00Z⟨/o:LastSaved⟩
⟨o:Pages⟩1⟨/o:Pages⟩
⟨o:Words⟩9160⟨/o:Words⟩
⟨o:Characters⟩54690⟨/o:Characters⟩
⟨o:Category⟩AMeGA project⟨/o:Category⟩
⟨o:Manager⟩Jane Greenberg⟨/o:Manager⟩
⟨o:Company⟩UNC-CH⟨/o:Company⟩

☀ 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>

<dc:identifier> http://www.... 2501Paper24.pdf</dc:identifier>

**<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

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

(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
- 1988 Anglo-American Cataloigng Rules, 2nd ed. *rev*.
- **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**

Titel (edit in slide master)

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*

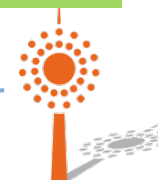• Canberrra qualifiers (meta-meta)

Dublin Core Elements, version 1.1

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type

- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights Management

•Dublin Core Element Set: http://dublincore.org/documents/dces/
•DCMI Metadata Terms: http://dublincore.org/documents/dcmi-terms

Property	Refinement	Outside standard
Title	Alternative title Subtitle	
Date	Created Modified	W3CDTF YYYY-MM-DD
Subject		LCSH, AAT, LCC

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

DC 6: November, 1998, Washington, D.C.

•   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

DC 11: Sept. 28-Oct.3, 2003, Seattle, WA
•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

Simplicity

050 00 Z693$b.W94 1991

082 00 025.3$220

100 1 Wynar, Bohdan S.

245 10 Introduction to cataloging and classification /$cBohdan S. Wynar.

250 8th ed. /$bArlene G. Taylor.

260 Englewood, Colo. :$bLibraries Unlimited,$c1992.

300 xvii, 633 p. :$bill. ;$c24 cm.

440 0 Library science text series

504 Includes bibliographical references (p. 591-599) and index.

650 0 Cataloging.

650 0 Subject cataloging.

650 0 Classification$xBooks.

630 00 Anglo-American cataloguing rules.

700 10 Taylor, Arlene G.,$d1941-

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
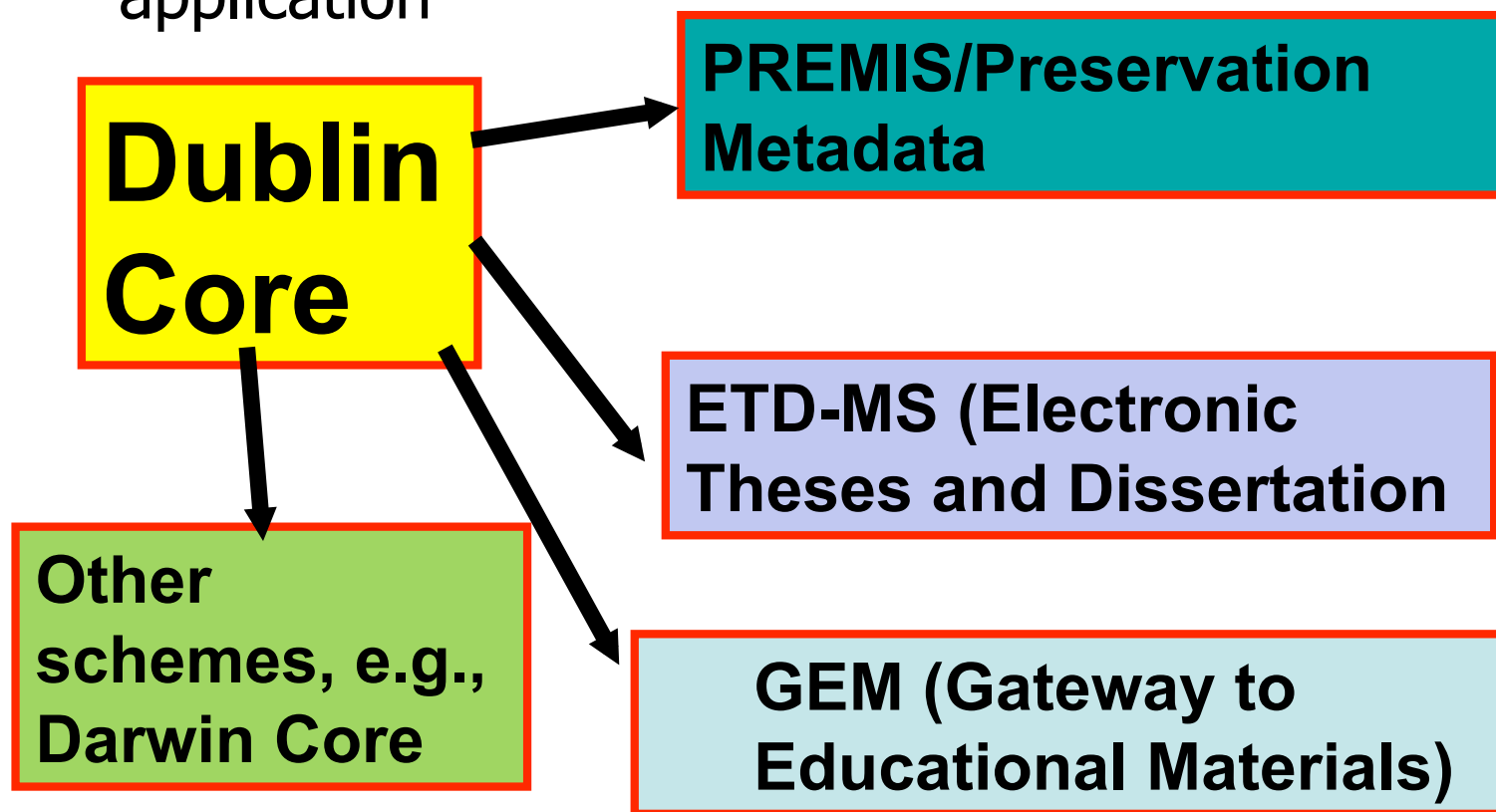
RELATION: Library science text series

Zeng, M (2004)

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 (Zeng, M., 2004, JG *rev., 2006*)

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

Other schemes, e.g., Darwin Core

GEM (Gateway to Educational Materials)

Modularity (Weibel discusses as Legos)

☀ **Additional principles**

- *Guiding metadata creation*

Dublin Core principles

- Dumb-down

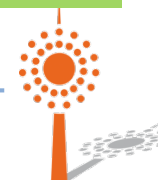- The one-to-one principle

- 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)

☀ **Property**	**Refinement**	**Outside standard**
Title	Alternative title Subtitle	
Date	Created Modified	W3CDTF **<YYYY-MM-DD>**
Subject		LCSH, AAT, LCC

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 paining.. *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, <u>but</u> . . .

- Remember that your metadata may be interpreted by machines <u>and</u> people, <u>so</u> . . .

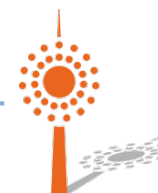- Consider whether the values you use will aid discovery outside your local context

Guidelines

- *"a simple scheme in a complex world!"*
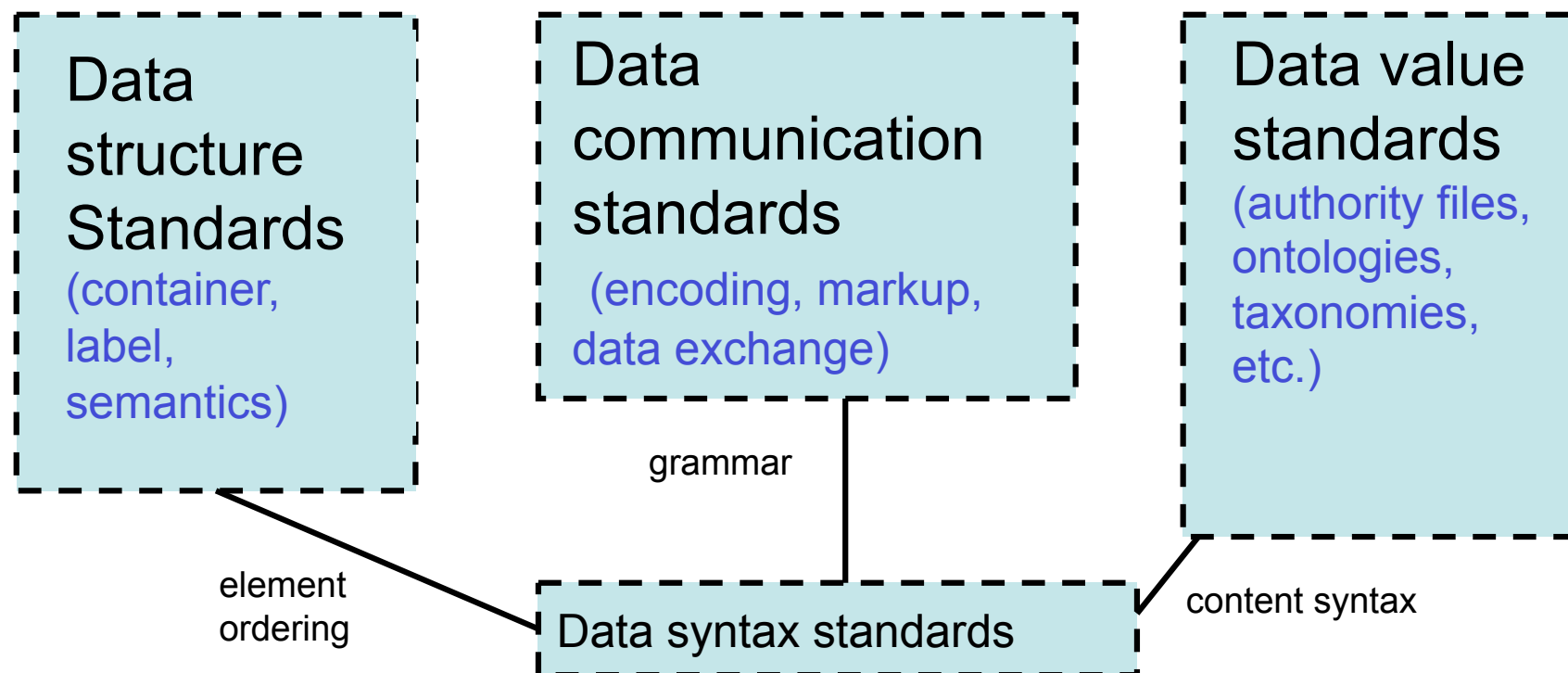
Standardization

- 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
  - US NISO Standard Z39.85-2001
  - ISO international standard 15836-2003
- Updated versions:
  - Internet Draft RFC5013, 2007
  - Revision US NISO Standard Z39.85-2007
  - 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.)

grammar

element ordering

Data syntax standards

content syntax

Guidelines, recommendations-1

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.

4. Using Dublin Core: http://dublincore.org//documents/usageguide/

Encoding guidelines "in use"

1. Expressing Dublin Core in HTML/XHTML meta and link elements: http://www.dublincore.org/documents/dcq-html/

2. DCMI Recommendation Guidelines for implementing Dublin Core in XML: http://dublincore.org/documents/dc-xml-guidelines/

Guidelines, recommendations-2

Dublin Core Abstract Model

1. http://dublincore.org/documents/2007/06/04/abstract-model/

New encoding guidelines

1. Expressing DC metadata using HTML/XHTML meta and link elements (DCMI Recommendation, 2008-08-04): http://dublincore.org/documents/2008/08/04/dc-html/

2. Expressing DC Description Sets using XML (DC-DS-XML) (Proposed Recommendation, 2008-09-01): http://dublincore.org/documents/2008/09/01/dc-ds-xml/

3. Expressing DC metadata using RDF (DCMI Recommendation 2008-01-14): http://dublincore.org/documents/2008/01/14/dc-rdf/
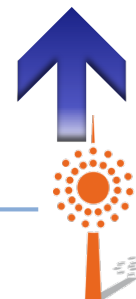
Toward a Semantic Web

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