An Analysis of Seven Metadata Creation Guidelines: Issues and Implications

Dr. Jung-ran Park, Caime Lu
Jung-ran.park@ischool.drexel.edu
Drexel University

Research Needs

- Rapid proliferation of digital repositories calls for serious research on metadata quality evaluation.

- Resource discovery and exchange across ever-growing distributed digital repositories demands semantic interoperability based on accurate and consistent resource description.

Research Needs

- The critical roadblock to achieving the goal of metadata quality control and semantic interoperability across digital repositories is posed by the lack of a common data model that is sharable and interoperable across libraries.

- The development of such a mediation mechanism calls for an empirical assessment of various critical issues surrounding metadata creation practice and metadata quality control.

Research Questions

- The overarching research questions of this project are derived from issues surrounding the metadata creation process, the employment of controlled vocabulary schemes, metadata quality control measures, and new competencies and skill sets faced by cataloging professionals in this digital era, together with the consequences to LIS education.

Overarching goals

- **Goal 1**. To examine current practices in the creation of descriptive metadata elements and the use of controlled vocabularies for subject access across distributed digital repositories.
- **Goal 2**. To identify factors hindering consistent, accurate and complete metadata description, resulting in the imposition of an impediment to resource sharing and access across distributed digital repositories.
- **Goal 3**. To assess new competencies and skill sets needed by cataloging professionals in developing digital repositories.

Metadata Item Record Analysis

- A study has been conducted 659 metadata item records for digitized image collections derived from three repositories.

- DC metadata element name and its corresponding definition are examined by utilizing linguistic semantic analysis.

Criteria for examining metadata item records

- Completeness/unused DC elements
- Accuracy
- Consistency
- Local addition

Inaccurate and Inconsistent Field Names and Metadata Elements

- 'Physical description' field is either mapped onto DC *Description* or *Format*.
- Great confusion in employing the DC elements *Type* and *Format* and they are interchangeably used.
- DC elements *Source* and *Relation* are inconsistently mapped onto various cataloger-defined fields.
- DC element *Relation* is interchangeably used with cataloger-defined field names such as 'digital collection' and 'example issues.'
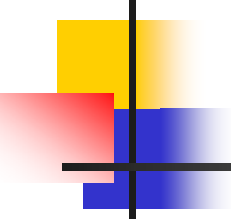
Locally Added Metadata Elements

- Accessibility and Provenance:
  - Contact information
  - Ordering information
  - Acquisition

  - Image modification
  - Full resolution, scan date, full text, note

Usage of DC Metadata Elements

Percentage of the Total Number of DC Metadata Elements Used by Three Collections								
DC Element	A n/203	% of the total number of DC elements used n/3476	B n/215	% of the total number of DC elements used n/2721	C n/241	% of the total number of DC elements used n/2606	Total n/659	% of total usage of DC
Title	203	5.8	217	8.0	241	9.2	661	100.3
Creator	196	5.6	148	5.4	30	1.2	374	56.8
Subject	580	16.7	416	15.3	448	17.2	1444	219.1
Description	203	5.8	210	7.7	263	10.1	676	102.6
Publisher	203	5.8	231	8.5	0	0.0	434	65.9
Contributor	289	8.3	100	3.7	19	0.7	408	61.9
Date	201	5.8	113	4.2	236	9.1	550	83.5
Type	0	0.0	150	5.5	235	9.0	385	58.4
Format	384	11.0	139	5.1	417	16.0	940	142.6
Identifier	265	7.6	107	3.9	7	0.3	379	57.5
Source	362	10.4	0	0.0	0	0.0	362	54.9
Language	63	1.8	0	0.0	5	0.2	68	10.3
Relation	121	3.5	98	3.6	4	0.2	223	33.8
Coverage	203	5.8	281	10.3	241	9.2	725	110.0
Rights	203	5.8	215	7.9	241	9.2	659	100.0
Non-Mapping	0	0.0	296	10.9	219	8.4	515	78.1
Total	3476	100.00	2721	100.0	2606	100.0	8803	1335.8

Most and Least Used DC Metadata Elements

- Most: subject, description, title, format, coverage (over 50%)

- Least: language, relation, source, creator and identifier

Semantic Overlaps in DC Metadata Elements

- The inherent conceptual ambiguities and semantic overlaps in some of the DC metadata elements affect semantic interoperability. Semantic overlap among certain DC metadata element names and their corresponding definitions create conceptual ambiguity and consequently hinder accurate, consistent and complete application of the DC metadata scheme.

Format vs. Type

- Format is "physical or digital manifestation of the resource" — unqualified DC metadata (DCMI, 2005)
- Type: "image may include both electronic and physical representations" —qualified DC metadata (DCMI, 2005) *type vocabulary* on image

Creator, Contributor, vs. Publisher

- **Creator**: "An entity primarily responsible for making the content of the resource."

- **Contributor**: "An entity responsible for making the content of the resource."

- **Publisher**: "An entity responsible for making the resource available."

source: unqualified DC metadata (DCMI, 2005)

Source vs. Relation

- *Source* is "a reference to a resource from which the present resource is derived."—unqualified DC metadata (DCMI, 2005)

- *Relation* is "the described resource is a physical or logical part of the referenced resource." — qualified DC metadata: *Relation, is Part of*

- *Relation* is "the described resource is a version, edition, or adaptation of the referenced resource." — qualified DC metadata: *Relation, is Version of*

Source is a particular type of Relation.

Implications

- Semantic interoperability across digital collections utilizing the DC metadata scheme is hindered partially due to the drawbacks inherent in the semantics of the scheme. DC metadata scheme needs to further evolve in order to disambiguate the semantic relations of the DC metadata elements that present semantic overlaps and conceptual ambiguities.

Mechanisms of Metadata Quality Improvement

- **Metadata creation guidelines/application profile**

- **Continuing education**

- Metadata creation tools (e.g. templates, concept maps)

  Park, Jung-ran. (2007). Evolution of a Concept Network and Its Implications to Knowledge Representation. <u>Journal of Documentation</u> Vol. 63. no. 6: 963-983.

Issues/problems

- Lack of specification for content designation of DC metadata scheme
- Semantic overlaps and conceptual ambiguities (see Park, 2006)
- Differences in local needs and user groups
- Variation of DC metadata application

Empirically Data Driven Common Data Model—Shared metadata semantics

- Critical need for the building of a common data model that can be sharable across libraries.

- Metadata application guidelines and procedures for the creation of descriptive metadata elements and application of controlled vocabularies.

- Identification of criteria and reasoning behind local addition and variation of metadata element values to and from selected metadata and controlled vocabulary schemes.

- There are a lack of studies that address such needs based on empirical analysis of existing metadata guidelines and best practices.

Extracting and analyzing best practices, guidelines, documentation, application profiles

As a preliminary study, we analyzed seven local metadata creation guidelines based on the Dublin Core (DC) metadata scheme.

Criteria used in analysis

- Metadata semantics (e.g., label names, qualifiers, definitions/applications of labels)
- Coverage of DC
- Divergence from DC
- Usage of controlled/uncontrolled vocabulary
- Metadata element status (e.g., cardinality and repeatability)
- Locally added elements (emerging semantics--possible candidates for inclusion as formal metadata elements)
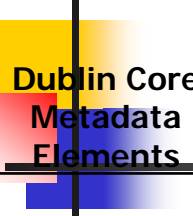
Overview of Surveyed Digital Repositories

Digital Collections	Format/Type	Subject
American Indians of the Pacific Northwest Graphic Part	Photograph	Northwest Coast and Plateau India Culture
American Indians of the Pacific Northwest Text Part	Text	Northwest Coast and Plateau India Culture
Architecture of the Pacific Northwest Database	Drawing	Pacific Northwest Architecture
Civil War Treasures from the New York Historical Society	Picture & Manuscript	History of the Civil War
Selected Civil War Photographs	Photograph	History of the Civil War
Portrait of the Ozarks	Photographed Portrait	Art & Culture
Chikanobu and Yoshitoshi Woodblock Prints	Digitized Works of Woodblock Prints	Art

Labels & Qualifiers of DC Elements.

The seven guidelines use different labels and qualifiers for the same DC element. The following table shows the corresponding label and qualifiers specified in local guidelines for each DC element.

Dublin Core Metadata Elements	American Indians of the Pacific Northwest Graphic data dictionary	American Indians of the Pacific Northwest Text data dictionary	Architecture Collection Data Dictionary	Civil War Treasures from the New York Historical Society	Selected Civil War photographs	Missouri Digitization Planning Project Metadata Guidelines	The Claremont Colleges Digital Library Metadata Best Practices
Title	Title	Title	Title; Alternative Title	Title	Title; Other title	Title; Title.Alternative	Title; Alternative title
Creator	Photographer; Original Creator	Author	Architectural firm; Architects; Associate architects; Engineers; Artist; Client	Creator	Creator	Creator	Creator
Subject	Subjects	Subjects	Building Style; Subject (LCTGM); Subject (LCSH)	Subject fields	Subjects	Subject; Subject.LCSH; Subject.MeSH; Subject.TGM; Subject.Keyword	Subject; LCSH MeSH DDC LCC UDC AAT
Description	Notes	Notes	Building street address; Purpose; Representation; Descriptive notes; Building notes	Notes; Description; Inscriptions	Notes	Description	Description; Table Of Contents; Abstract

Dublin Core Metadata Elements	American Indians of the Pacific Northwest Graphic data dictionary	American Indians of the Pacific Northwest Text data dictionary	Architecture Collection Data Dictionary	Civil War Treasures from the New York Historical Society	Selected Civil War photographs	Missouri Digitization Planning Project Metadata Guidelines	The Claremont Colleges Digital Library Metadata Best Practices
Publisher	Studio Name; Studio Location	Publisher; Place of Publication				Publisher	Publisher
Contributor	Contributor	Contributor	Contributor		Related Names	Contributor	Contributor
Date	Date; Dates	Date of Publication	Date of drawing execution; Dates	Date	Date	Date.creation; Date.Current	Date
Type	Object Type; Type	Object Type; Type	Object Type; Type			Type	Digital Type
Format	Transmission Data	Transmission Data	Digital reproduction information; Physical description	Medium; Dimensions	Medium; Formats	Format	Format
Identifier	Photographer's Reference Number; Negative Number; Resource Identifier;	Resource Identifier	Negative Number	Call Number	Call number; Card #; Digital ID (or Video frame ID)	Identifier.URL; Identifer.MDI; Identifier.Local	Identifier

Dublin Core Metadata Elements	American Indians of the Pacific Northwest Graphic data dictionary	American Indians of the Pacific Northwest Text data dictionary	Architecture Collection Data Dictionary	Civil War Treasures from the New York Historical Society	Selected Civil War photographs	Missouri Digitization Planning Project Metadata Guidelines	The Claremont Colleges Digital Library Metadata Best Practices
Source	Collection; Repository	Original Source; Repository	Repository Collection; Repository		Collection; Repository	Source	Source
Language						Language	Language
Relation	Relation	Relation	Digital collection			Relation	Relation Is Version Of; Has Version; Is Replaced By; Replaces; Is Required By; Requires; Is Part Of; Has Part; Is Referenced By; References; Is Format Of; Has Format; Conforms to
Coverage	Location Depicted	Geographic Subjects	Building location			Coverage	Coverage
Rights	Restrictions	Restrictions	Restrictions			Rights	Rights Management
Source	Collection; Repository	Original Source; Repository	Repository Collection; Repository		Collection; Repository	Source	Source

Coverage of DC Metadata Elements

- Two guidelines specify all 15 DC elements. Other guidelines specify only seven of the DC elements.

- Four metadata guidelines utilize locally added non-DC elements to reflect local resource characteristics.

Locally Added Elements

Guidelines	Non-DC elements
American Indians of the Pacific Northwest Graphic data dictionary	Physical Description, Acquisition
American Indians of the Pacific Northwest Text data dictionary	Acquisition
Architecture Collection Data Dictionary	Acquisition, Repository collection guide, Earliest date, Latest date, Order number, Ordering information
The Claremont Colleges Digital Library Metadata Best Practices	Notes, Staff only, Cataloged by, Catalog date, Object file name

Status of Elements

- Four guidelines specify the status of the metadata elements.

  - each of these four guidelines recommends a different set of required/mandatory elements.

  - *title* is the only required element.

- Two guidelines explicitly specify whether a DC metadata element is repeatable in describing digital objects.

Mandatory Elements in Four Metadata Guidelines

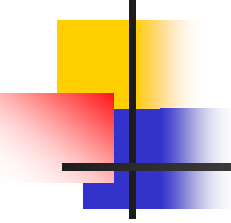| Metadata Guidelines	Mandatory Elements	
American Indians of the Pacific Northwest Graphic data dictionary	Title, Date, Object Type, Contributor, Relation, Resource Identifier, Dates	
Architecture Collection Data Dictionary	Title, Object Type, Type, Digital Collection, Earliest Date, Latest Date , Ordering Information	
Missouri Digitization Planning Project Metadata Guidelines	*Item Level*	Title, Creator, Subject, Description Date, Format, Identifier, Relation
	Collection Level	Title, Creator, Subject, Description, Identifier, Relation
The Claremont Colleges Digital Library Metadata Best Practices	Title, Subject, Description, Digital Type, Relation, Creator (if available), Publisher, Rights Management, Date, Format, Identifier	

Controlled Vocabulary

- All the surveyed guidelines recommend using controlled vocabulary for *Subject* and *Creator.*

- *Library of Congress Thesaurus for Graphic Materials I: Subject Headings* and the *Library of Congress Subject Headings* are the most frequently suggested controlled vocabularies for *Subject.*

- *Library of Congress Name Authority File* is the most frequently recommended controlled vocabulary scheme for *Creator.*

- Elements that apply controlled vocabularies:
  Subject, Creator, Type, Format, Date Coverage.

	Subject	Creator	Type	Format	Date	Coverage
American Indians of the Pacific Northwest Graphic data dictionary	LC TGM I, LCSH	LCNAF	LC TGM II, DCMI Type Vocabulary			LCNAF
American Indians of the Pacific Northwest Text data dictionary	LCSH	LCNAF				
Architecture Collection Data Dictionary	AAT, LC TGM I, LCSH	LCNAF	AAT, DCMI Type Vocabulary			LCNAF
Civil War Treasures from the New York Historical Society	LCSH, LC TGM (2nd ed), LCNAF, AACR2	LCNAF, AACR2 (for Heading)		LC TGM II		
Selected Civil War Photographs	LCSH, LC TGM I LCNAF			LC TGM II		
MISSOURI Digitization Planning Project Metadata Guidelines	LCSH, AAT, MeSH, NGL, LC TGM, LCNAF		DCMI Type Vocabulary	Internet Media Types [MIME]	ISO 8601 [W3CDTF]	
The Claremont Colleges Digital Library Metadata Best Practices	LCSH, MeSH, DDC, LCC, UDC, AAT	LCNAF, AACR2, Getty Union List of Artists Names for artwork	DCMI Type Vocabulary	Internet Media Types [MIME]	ISO 8601 [W3CDTF]	TGN, GNIS

Controlled Vocabulary

- LC TGM I: Thesaurus for Graphic Materials I: Subject Headings;
- LC TGM II: Thesaurus for Graphic Materials II: Genre and physical characteristic terms;
- LCSH: Library of Congress Subject Headings;
- AAT: Art and Architecture Thesaurus;
- LCNAF: Library of Congress Name Authority File;
- NGL: Newspaper Genre List;
- TGN: Thesaurus of Geographic Names;
- GNIS: USGS Geographic Names Information System

Preliminary Conclusion

- Results of the analysis show great divergence in the application of the Dublin Core metadata scheme across the surveyed digital repositories.

- Each set of guidelines utilizes different labels and DC qualifiers to describe local digital resources.

- Elements such as *Title, Subject* and *Type* tend to be relatively consistent in the usage of labels and qualifiers.

- Divergence across the surveyed guidelines appears in labels for the following elements: *Creator, Description, Format and Identifiers*

   (see also Park, 2006).

- Metadata semantics of locally added elements—provenance, technical & administrative information (see also Park, 2006)

Preliminary Conclusion

- DC metadata scheme offers flexibility and extensibility built directly into the framework. Differences across local guidelines and best practices evidence such flexibility. It is this flexibility that enables libraries to make adjustments and modifications correspondent to local needs.

- Divergence in application of the DC metadata scheme may impede semantic interoperability and resource sharing across DC-based digital repositories.