

**IMLS Digital Collections  
and Content**

---

Amy Jackson  
amyjacks@uiuc.edu  
IMLS Digital Collections and Content project  
University of Illinois at Urbana-Champaign

**2007 LITA national forum**

---

---

---

---

---

---

---

---

**IMLS Digital Collections and Content**

- Project began December 2002 as an IMLS National Leadership Grant
  - Tim Cole, Principal Investigator
  - Amy Jackson, Project Coordinator
  - Co-Principal Investigators: Carole Palmer, Sarah Shreeves, Mike Twidale, Bill Mischo, Nancy O'Brien
  - Collaboration with UIUC Library and Graduate School of Library and Information Science
- <http://imlsdcc.granger.uiuc.edu/>

---

---

---

---

---

---

---

---

**IMLS Digital Collections and Content**

- Project Objectives:
  - Implement a human readable collection registry of digital collections created or developed with funding from IMLS National Leadership Grant (NLG) program. Expanded in 2005 to include some LSTA (Library Services and Technology Act) grant collections.
  - Use Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to implement an item-level metadata repository for items contained in NLG collections.

---

---

---

---

---

---

---

---

## IMLS Digital Collections and Content

- Research topics
  - Utility and usability of Registry & Repository
  - Current metadata practices of IMLS NLG grantees
  - Implications for interoperability (Framework of Guidance for Building Good Digital Collections)

---

---

---

---

---

---

---

---

## Collection Registry

- Collection registry
  - 180 NLG projects
  - 15 LSTA projects

Images 80%  
Text 68%  
Physical Object 29%  
Sound 20%  
Interactive Resource 10%

---

---

---

---

---

---

---

---

## Collection Registry

- Top GEM (Gateway to Educational Materials) subjects in Collection Registry
  - Social Studies 80%
    - United States history
    - State history
  - Arts 46%
    - Visual arts
    - Photography
  - Science 17%

---

---

---

---

---

---

---

---

## Collection description

- Project vs. collections
- Definition of “collection”
- Collection level schema adapted from RSLP (Research Support Libraries Programme, UK)
- Collection managers review and update collection records

---

---

---

---

---

---

---

---

## Related work

- RSLP (Research Support Libraries Programme, UK)
  - Cornucopia
- National Science Foundation
  - National Science Digital Library

---

---

---

---

---

---

---

---

## Item-level repository

- Item-level Repository
  - Harvesting 71 of 195 Collections (36%)
  - 37 Repositories
  - 310,448 records
- Item Records
  - 86% images
  - 14% text

---

---

---

---

---

---

---

---

## Item-level repository

### Top Item-level subjects

United States	Archaeology
People	Buildings
Songs with piano	Photographers
Trees	Mountains
Tennessee Valley Authority	Men
Archaeology Southern States	Archaeological site
Works Progress Administration	Insect
Cities & towns	Bodies of water
Women	

---

---

---

---

---

---

---

---

---

---

## OAI-PMH

- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
  - Low barrier protocol to enable the interoperability of digital libraries.
    - Data providers
    - Service providers (aggregators)
  - Requires that metadata be exposed in Dublin Core
  - Other formats are optional
- 5 of the 37 repositories export in schemas other than simple or qualified Dublin Core. Other schemas include MARC21, MODS, OLAC, and ETDMS.

---

---

---

---

---

---

---

---

---

---

## OAI-PMH

- Barriers to sharing metadata through OAI-PMH
  - Technical Infrastructure
  - Metadata (not shareable or no item-level metadata)
  - Institution/Project

---

---

---

---

---

---

---

---

---

---

## IMLS Digital Collections and Content

- Aggregated environment
  - Interoperability of metadata
    - Use of fields
    - Field definitions
  - Different descriptive practices (libraries, archives, museums)

Moving towards shareable metadata by Sarah L. Shreeves, Jenn Riley, and Liz Milewicz  
First Monday, volume 11, number 8 (August 2006),  
URL: [http://firstmonday.org/issues/issue11\\_8/shreeves/index.html](http://firstmonday.org/issues/issue11_8/shreeves/index.html)

---

---

---

---

---

---

---

---

## Harvested Metadata

- How has use of Dublin Core changed over time?  
Records created from January 1, 2001 through December 31, 2006.
  - Quantitative analysis
    - What measurable changes can we see in the metadata?
  - Qualitative analysis
    - How has use of fields changed over time?

---

---

---

---

---

---

---

---

## Quantitative analysis

- Quantitative analysis
  - Repetition of elements
  - Length of fields
  - Use of core fields (Shreeves et al. (2005))

<b>Title</b>	<b>Creator</b>
<b>Subject</b>	<b>Description</b>
<b>Date</b>	<b>Format</b>
<b>Identifier</b>	<b>Rights</b>

---

---

---

---

---

---

---

---

## Quantitative analysis

- Repetition of fields
  - Stable
- Length of fields
  - Stable
- Use of all 8 core fields
  - Declining

---

---

---

---

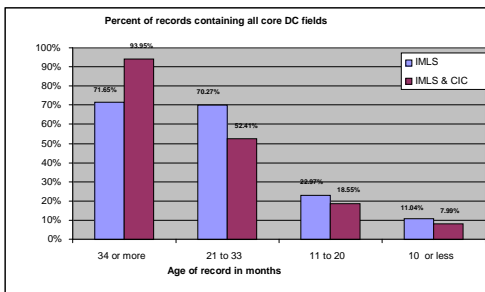
---

---

---

---

## Quantitative Analysis



---

---

---

---

---

---

---

---

## Quantitative Analysis

- Of these eight elements, the two elements most often missing are creator (used in 42% of records) and rights (52%).
- Identifier, title, and subject were each used in over 96% of all records.
- Format and description fields have shown the most significant decline in use since 2003.
- Decreased repetition and length of the description field, and an overall increase in use of the relation field.

---

---

---

---

---

---

---

---

## Item-level repository

Number of harvested collections using each DC field

Field	Number of Collections	% of Collections
Title	35	100%
Identifier	35	100%
Subject	33	94%
Type	32	91%
Creator	32	91%
Description	31	89%
Date	30	86%
Publisher	30	86%
Format	28	80%
Rights	27	77%
Language	26	74%
Relation	23	66%
Contributor	21	60%
Source	20	57%
Coverage	18	51%

---

---

---

---

---

---

---

---

---

---

---

---

## Percent of Records Containing each DC field

DC field	% of records
Identifier	100%
Title	97%
Subject	96%
Type	86%
Publisher	76%
Date	70%
Description	67%
Format	66%
Relation	53%
Rights	52%
Language	50%
Source	45%
Creator	42%
Coverage	30%
Contributor	8%

---

---

---

---

---

---

---

---

---

---

---

---

## Qualitative analysis

- Qualitative analysis
  - 225 records from 6 average repositories (time increments)
    - Document changes in practice over time
  - 600 randomly selected records
- Only 1 observed change in practice over time
  - Early records:  
<title>Frankie / Music by Neil Sedaka; words by Howard Greenfield </title>
  - Later records:  
<title>Frankie</title>  
<creator>Music by Neil Sedaka; words by Howard Greenfield</creator>

---

---

---

---

---

---

---

---

---

---

---

---

## Qualitative analysis

- Unpacking MARC tags to Dublin Core elements
  - Merged publisher and date fields (maybe) from MARC 260
  - Title field containing creator or contributor information from MARC 245
  - Confusion of type and format fields from MARC 300
  - Incorrect use of delimiter
    - <creator>Pika,</creator>
    - <creator>I.,</creator>
    - <creator>et al.,</creator>

---

---

---

---

---

---

---

---

## Qualitative analysis

- Unpacking MARC tags to Dublin Core elements
  - Merged publisher and date fields from MARC 260
  - Title field containing creator or contributor information from MARC 245
  - Confusion of type and format fields from MARC 300
  - Incorrect use of delimiter
    - <creator>Pika,</creator>
    - <creator>I.,</creator>
    - <creator>et al.,</creator>

---

---

---

---

---

---

---

---

## Qualitative analysis

- Misuse/confusion of Dublin Core elements
  - Date and coverage fields
  - Source and relation fields
  - Format and description fields
  - Format

---

---

---

---

---

---

---

---

## Qualitative analysis

### ■ Misuse of Dublin Core Elements

- Date and coverage fields

Item about the nineteenth century, published in 2007.

Correct: `<date>2007</date>`  
`<coverage>1800-1899</coverage>`

Incorrect: `<date>1800-1899</date>`

---

---

---

---

---

---

---

---

## Qualitative analysis

### ■ Misuse of Dublin Core Elements

- Source and Relation fields

Source: "The resource from which the described resource is derived."

Relation: "A related resource."

(From: *Dublin Core Metadata Element Set, Version 1.1*,  
<http://dublincore.org/documents/dces/>)

Example: Series title belongs in relation not source.

---

---

---

---

---

---

---

---

## Qualitative analysis

### ■ Misuse of Dublin Core elements

- Format and Description fields

Example:

Notes Material: Whale Bone (MARC 500)

Incorrect: `<description>Material: Whale Bone</description>`  
`<description>9 in. x 6 in.</description>`

Correct: `<format>Material: Whale Bone; 9 in. x 6 in.</format>`

---

---

---

---

---

---

---

---

## Qualitative analysis

### ■ Misuse of Dublin Core fields

- Format "The file format, physical medium, or dimensions of the resource." (From: *Dublin Core Metadata Element Set, Version 1.1*, <http://dublincore.org/documents/dces/>)

Incorrect:

<format>Available via the World Wide Web</format>

<format>web browser</format>

<format>Any machine capable of running graphical Web browsers, 640x480 minimum monitor resolution</format>

Correct:

<format>image/jpeg</format>

---

---

---

---

---

---

---

---

---

---

## Qualitative analysis

### ■ Confusion in Descriptive metadata and Administrative metadata (too much information exposed)

- Administrative metadata is not helpful for discovery in the aggregated environment.
  - software used for digitization, master file format, storage equipment
- Exposed metadata as one "view" of all associated metadata.

---

---

---

---

---

---

---

---

---

---

## Qualitative analysis

### ■ Lost information (not enough information exposed)

- Additional metadata fields helpful for discovery could be mapped to Dublin Core and exposed.
- Some DL software allows local fields to be mapped or not mapped to Dublin Core fields for exporting through OAI-PMH; be sure that helpful information isn't being hidden from the service provider.

---

---

---

---

---

---

---

---

---

---

## Conclusions regarding metadata

- Conclusions
  - Native metadata records are rich in meaning in their own environment, but lose richness in the aggregated environment due to mapping errors and misunderstanding and misuse of Dublin Core fields.
  - Mapping is often based on semantic meanings of metadata fields rather than value strings.
  - Correct mapping could improve metadata quality significantly.

---

---

---

---

---

---

---

---

## Metadata Recommendations

- Recommendations
  - Publish local metadata practices
  - Publish crosswalking practices
  - Expose native metadata in addition to Dublin Core
  - Ensure that metadata creators receive proper training

---

---

---

---

---

---

---

---

## Project findings

- Virtual aggregations of digital resources require clear and complete collection development and collection selection policies.
- Both content provider and service provider share responsibility for the successful aggregation of digital resources.
- The definition of digital collection is evolving as digital projects become more sophisticated.

---

---

---

---

---

---

---

---

## Future work for aggregators

- Create clear collection development plans.
- Analyze relationships between collection-level metadata and item-level metadata.
- Combine metasearch and OAI-PMH
- Improve interface representation of content for browsing and searching.

---

---

---

---

---

---

---

---

## IMLS Digital Collections and Content

Reference: Shreeves, S.L., Knutson, E.M., Stvilla, B., Palmer, C.L., Twidale, M.B., & Cole, T.W. "Is 'quality' metadata 'shareable' metadata? The implications of local metadata practice on federated collections." In H.A. Thompson (Ed.) *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries*, April 7-10 2005, Minneapolis, MN . Chicago, IL: Association of College and Research Libraries. p.223-237.

Amy Jackson  
Project Coordinator, IMLS Digital Collections  
and Content  
University of Illinois at Urbana Champaign  
amyjacks@uiuc.edu

---

---

---

---

---

---

---

---