1-1-2008

Dublin Core Metadata Harvested Through OAI-PMH

Amy S. Jackson

Myung-Ja Han

Kurt Groetsch

Megan Mustafoff

Timothy W. Cole

Dublin Core Metadata Harvested Through OAI-PMH

Amy S. Jackson
Myung-Ja Han
Kurt Groetsch
Megan Mustafoff
Timothy W. Cole

ABSTRACT. The introduction in 2001 of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) increased interest in and awareness of metadata quality issues relevant to digital library interoperability and the use of harvested metadata to build "union catalogs"

Amy S. Jackson (E-mail: amyjacks@uiuc.edu) is the Project Coordinator for the IMLS Digital Collections and Content project and Visiting Assistant Professor of Library Administration at the University of Illinois at Urbana Champaign. Myung-Ja Han (E-mail: mhan3@uiuc.edu) is Metadata Librarian and Visiting Assistant Professor of Library Administration at the University of Illinois at Urbana Champaign. Kurt Groetsch (E-mail: kurt.groetsch@gmail.com) is Metadata Librarian and Visiting Assistant Professor of Library Administration at the University of Illinois at Urbana Champaign. Megan Mustafoff (E-mail: mustafof@ad.uiuc.edu) is Project Coordinator at the Library Research Center in the Graduate School of Library and Information Science at the University of Illinois at Urbana Champaign. Timothy W. Cole (E-mail: t-cole3@uiuc.edu) is Interim Head of Digital Services and Development, Mathematics Librarian and Professor of Library Administration at the University of Illinois at Urbana Champaign.

Address correspondence to: Amy S. Jackson, Myung-Ja Han, Kurt Groetsch, Megan Mustafoff and Timothy W. Cole, IMLS Digital Collections and Content, 052 Grainger Engineering Library, MC-274, 1301 West Springfield Avenue, Urbana, IL 61801.

The research presented here was funded in part through National Leadership Grant number LG-02-02-0281 from the Institute of Museum and Library Services. This article has been peer-reviewed.

Journal of Library Metadata, Vol. 8(1) 2008
http://jlm.haworthpress.com

doi:10.1300/J517v08n01_02

of digital information resources. Practitioners have offered wide-ranging advice to metadata authors and have suggested metrics useful for measuring the quality of shareable metadata. Is there evidence of changes in metadata practice in response to such advice and/or as a result of an increased awareness of the importance of metadata interoperability? This paper looks at metadata records created over a six-year period that have been harvested by the University of Illinois at Urbana Champaign, and reports on quantitative and qualitative analyses of changes observed over time in shareable metadata quality. doi:10.1300/J517v08n01_02

KEYWORDS. Metadata, OAI-PMH, digital collections

INTRODUCTION

The importance of descriptive practice is not a new theme in the library domain; however, the widespread adoption of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and the Dublin Core Metadata scheme has led digital library practitioners to examine the characteristics of shareable, non-MARC descriptive metadata records. The IMLS/NISO *Framework of Guidance for Building Good Digital Collections*, first published in 2001, emphasizes the importance of disseminating descriptive metadata that supports interoperability. Following the publication of this document, additional concrete advice on how to create metadata well suited for sharing has been offered in several venues (Digital Library Forum and the National Science Digital Library, 2005), (Elings & Waibel, 2007), (Hutt & Riley, 2005), (Shreeves, Riley & Milewicz, 2006a), (Zeng & Chang, 2006), (Dushay & Hillman, 2003). Less frequently discussed, however, is how institutions are implementing Dublin Core in practice (Ward, 2004). The following article discusses quantitative and qualitative observations of Dublin Core metadata records harvested by two cultural heritage service providers at the University of Illinois at Urbana Champaign (UIUC). The examination focuses on changes in metadata practices over time, as well as observations of inaccurate and inconsistent mappings to Dublin Core. Researchers originally hoped to find indications of metadata becoming more shareable as digital projects mature, but findings did not support this hypothesis.

UIUC METADATA PORTALS

UIUC provides access to descriptive metadata harvested with OIA-PMH through several portals, including the Institute of Museum and Library Services Digital Collections and Content Project (IMLS DCC), located at http://imlsdcc.grainger.uiuc.edu, and the Committee on Institutional Cooperation (CIC) Metadata Portal, located at http://cicharvest.grainger.uiuc.edu. The IMLS DCC portal harvests metadata from cultural heritage projects funded by the Institute of Museum and Library Services (IMLS). Eighty-five percent of the records in this portal represent images and 14% represent texts. The IMLS DCC project staff interacted with several data-providers regarding technical specifications and administrative information, but gave relatively little feedback to individual metadata providers regarding metadata quality. The project allowed for general presentations and publications stressing the importance of shareable metadata quality, including presentations at IMLS WebWise Conferences (Cole & Shreeves, 2004), and publications in other venues (Shreeves, Riley & Milewicz, 2006a). Conversations with data providers regarding mapping best practices were not within the scope of the project. The CIC Metadata Portal aggregates metadata describing resources held at participating CIC institutions. Most of these objects are cultural heritage resources. Construction of the CIC Metadata Portal allowed for substantial interaction between the service provider and data providers, including exchange of shareable metadata and mapping best practices, and feedback was given on a repository-by-repository basis. Table 1 provides information regarding the size of the IMLS DCC and CIC Metadata portals.

This study analyzed metadata records in the IMLS DCC portal in depth, and observations from the CIC Metadata Portal confirmed IMLS

TABLE 1. IMLS DCC and CIC Metadata Portals

	IMLS DCC Portal	CIC Metadata Portal
Number of records	300,000	630,441
Number of contributing repositories	35	28
Range of records harvested from contributing repositories	35-65,000	13-300,000
Average number of records harvested per repository	7,425	25,000
Median number of records harvested by repository	1,281	6,973

DCC findings. All records in this study were created between January 1, 2001 and December 31, 2006, and were stored and accessed on a Microsoft SQL Server. SQL queries were used for the quantitative analysis, and the qualitative analysis was performed by examining individual XML files as originally harvested.

DC AND OAI-PMH

The decision by the OAI-PMH technical committee to require Dublin Core was controversial when first made, and continues to be seen as negative in some settings (Cole & Foulonneau, 2007; Lagoze, 2004; Van de Sompel, Young & Hickey, 2003; Chavez et al., 2006). Many in the library community are concerned about its lack of richness and specificity (Lagoze, 2001). However, one of the strengths of the schema is its ability to act as a lowest common denominator among various richer schemas, and findings indicate that use of the schema is increasing in IMLS National Leadership Grant (NLG) digitization projects (Palmer, Zavalina & Mustafoff, 2007).

The Dublin Core Metadata Element Set (DCMES) has fifteen elements, all of which are optional and repeatable. These elements are Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, and Type. A previous study (Shreeves et al., 2005) and best practices published by an IMLS DCC collection identify eight of these elements as significant to the completeness and of a metadata record, and most helpful for search and discovery. These elements are title, creator, subject, description, date, format, identifier, and rights. Analysis of the IMLS DCC records indicates that at least 77% of participating repositories use these eight fields (Table 2). These elements are not always appropriate for all resources, but are more standard than the other DC elements and will be referred to as core elements.

Table 2 illustrates limited use of the 15 Dublin Core elements among the repositories harvested by the IMLS DCC project. Previous studies (Ward, 2004) show similar limited use of the Dublin Core elements. Due to the fact that not all repositories use all 15 Dublin Core elements, interoperability among records aggregated from various repositories is diminished; only title and identifier elements are used by all repositories contributing to the IMLS DCC portal.

The Open Archives Initiative released OAI-PMH in 2001 as a low-barrier specification to enable interoperability of digital libraries. In order

TABLE 2. IMLS DCC Respositories Using Each Element

Fields	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

to encourage interoperable metadata records, the specification requires that all descriptive metadata be exposed in Dublin Core format, and provides an option of exposing metadata in other formats. Of the combined 63 repositories harvested by the two UIUC portals, only 12 repositories expose metadata in a scheme other than simple or qualified Dublin Core. Other formats include MARC21, MODS, OLAC (Open Language Archives Community), and ETDMS (Electronic Theses and Dissertations Metadata Set).

Quantitative Metadata Analysis

To analyze quantitative trends in metadata practices over time, records were examined against standard measures of metadata quality that emphasize shareability (Shreeves et al., 2005; Stvilia et al., 2004). These measures look at use of core fields, number of distinct fields used, and length and repetition of fields. The date of the record was determined by the embedded date stamp, indicating the most recent date of modification or date of creation if the record was not modified.

Findings indicate a relative stability in repetition of elements and length of fields, but a significant decline in the number of records using

all eight core fields discussed above. Both the IMLS DCC and CIC portals show the same trend (Table 3).

Of these eight elements, the two elements most often missing are creator (used in 39% of IMLS DCC records) and rights (52%); 96% of all records use identifier, title, and subject fields. Format and description fields have shown the most significant decline in use since 2003 (Table 4). Other findings of note in the IMLS DCC portal include decreased repetition and length of the description field, and an overall increase in use of the relation field.

The decline in records using all eight core fields is counterintuitive to the increase in discussions regarding quality metadata since the introduction of OAI-PMH. In order to gather additional data about this trend, researchers used information about each collection gathered through the course of the IMLS DCC project (Palmer & Knutson, 2004).

Collection Characteristics

Use of the core fields varies considerably by the size of repository, with the smallest and largest repositories creating more records using all eight fields. Table 5 illustrates the percent of records that include all

TABLE 3. Percent of Records Containing Eight Core DC Fields, Based on Age of Record

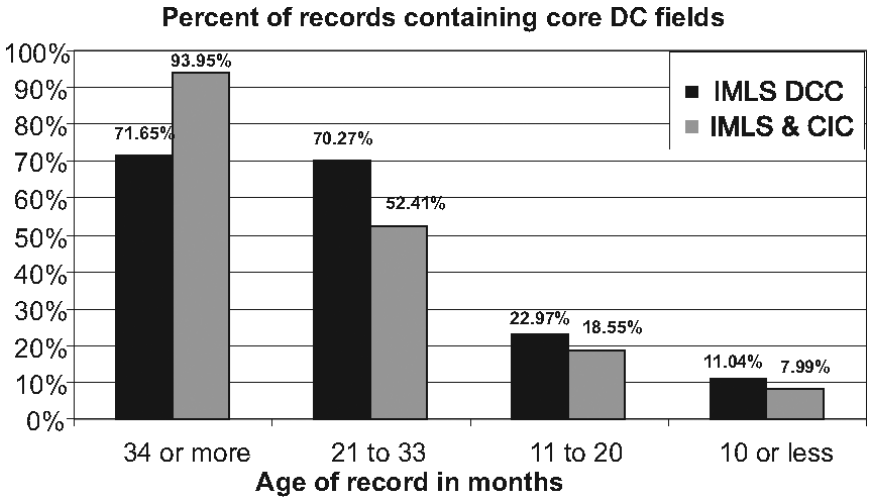


TABLE 4. Percent of All Records Containing Description and Format Fields

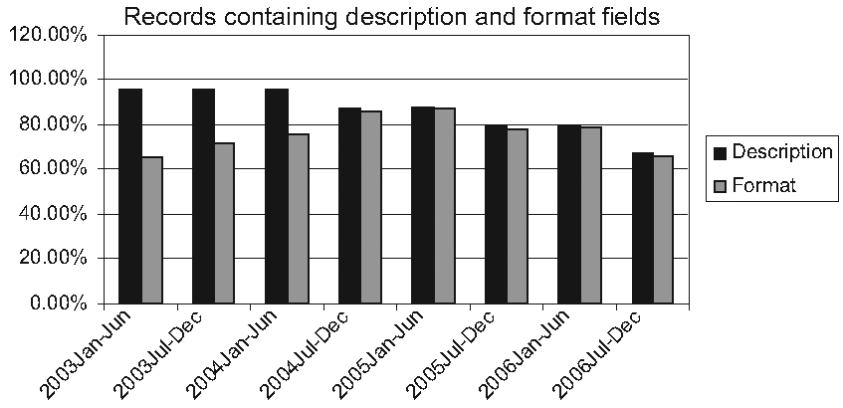
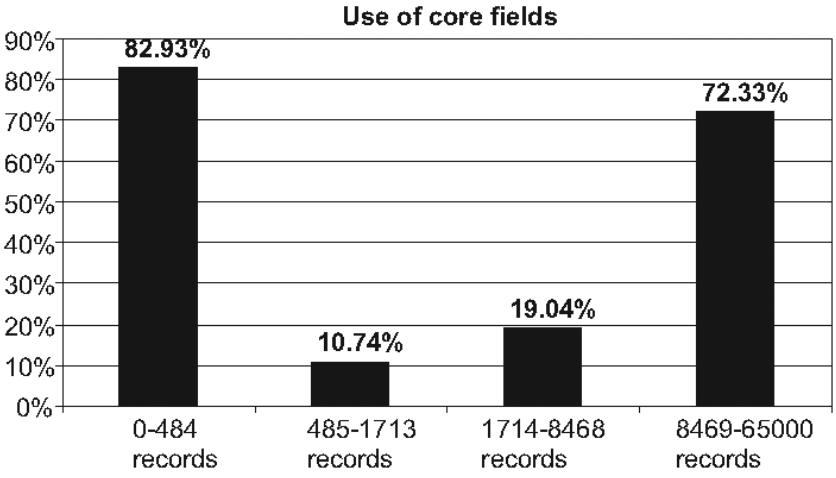


TABLE 5. Percent of Records Using Eight Core Fields by Collection Size



eight fields, based on the size of the repository in number of records harvested.

Use of turnkey software (e.g., CONTENTdm) also appears to have a significant impact on repositories creating records with the core Dublin Core fields, as shown by a comparison between repositories using CONTENTdm and other contributing repositories. All CONTENTdm

repositories use all eight fields in at least 50% of records; while only 44.4% of non-CONTENTdm repositories use these eight fields in at least 50% of records (see (Park, 2005) and (Park & Park, 2005) for additional studies). The higher use of core Dublin Core fields in the CONTENTdm repositories may be due to the default Dublin Core template.

Other variables having impact on metadata practices include the number of staff members working on a digital project, previous experiences with digital collections at an institution, and the collaborative nature of the project. Analysis revealed that digital projects at an institution with no previous experience in digital projects were more likely to have records using core fields, and projects with fewer than 10 staff members were also more likely to have records with all eight fields. Records from collaborative projects were more likely to use all of the core fields than records from non-collaborative projects.

Based on analysis of the IMLS DCC repositories, the type of digital collection most likely to produce records using the eight core Dublin Core fields is a small collaborative project with a small staff using turnkey software. Collaborative projects are more likely to encourage standard use of Dublin Core rather than local metadata schemes, and projects with fewer individuals creating metadata are more likely to create consistent metadata.

The decline in use of all eight core Dublin Core fields may indicate that as digital project mature, metadata creators are becoming more discriminating in their use of Dublin Core fields in the local context. While this may make sense for local implementations, use of Dublin Core fields needs to become broader, rather than narrower, in order to maximize interoperability in the aggregated environment. Due to the current use of Dublin Core fields by repositories harvested by the IMLS DCC portal, a user can only search across all repositories by searching on the title field.

Qualitative Metadata Analysis

Although the quantitative analysis helped to determine overall trends, questions remained regarding changes in semantics and values of Dublin Core fields over time, as well as questions regarding perceived weaknesses and attempts by data providers to make metadata more shareable. A qualitative analysis of the metadata records was needed to examine these trends, and a sample of 900 records from the IMLS DCC portal was chosen for a 92% confidence limit. This confidence limit ensures

that 92% of the random samples taken from the IMLS DCC portal would contain average values, calculated using standard statistical sampling rates. Researchers examined an initial 225 records from six repositories representing average record contributions at various times throughout the project. Five records were examined from each six-month period for each repository, with researchers attempting to document changes in practice over time. An additional 90 records were selected from repositories using CONTENTdm, and 600 records were randomly selected from the entire IMLS DCC portal.

Of the initial six repositories examined, researchers only found one repository with a change in practice over time. In early records from this repository, the entire MARC 245 field was mapped to the Dublin Core title field. The change in practice occurred in late 2006 when the MARC 245 subfield c began to be mapped to the Dublin Core Creator field. The researchers did not observe any other changes in the records analyzed from these repositories, even though several mapping errors were found.

In order to gain a better understanding of the observed mapping errors, researchers compared harvested Dublin Core metadata records to the corresponding metadata records available on project web sites. Because many collections use other schemas than Dublin Core in the local context, harvested metadata is often mapped to Dublin Core, rather than created natively in Dublin Core. Harvested records were examined to determine if any changes in semantics and values had been made when the native metadata was mapped to Dublin Core. As an experiment, native metadata records were remapped to Dublin Core by researchers to see which fields were most often incorrectly mapped. Findings indicated that many harvested Dublin Core metadata records had element values that were incorrectly mapped, and value strings mapped to single elements that should be divided between several different elements (Hutt & Riley, 2005). Publicly available crosswalks (e.g., Library of Congress' MARC to Dublin Core Crosswalk) do not always account for semantic values of elements, and may provide misleading mappings. Among the fifteen simple Dublin Core elements, description, format, subject, and type fields show the most significant changes in numbers when remapped from the original harvested records.

Multiple value strings in one element instance in the original records caused the increase in description and subject fields. Although it is acceptable to repeat elements, researchers found many records that have one description or subject field with many value strings delimited by semicolons, colons, or periods. The same problem was observed in the contributor field.

Example:

Incorrect: <*subject*>Information Science; Artificial Intelligence; Computer Science; Engineering, Chemical</*subject*>

Correct: <*subject*>Information Science</*subject*> <*subject*>Artificial Intelligence</*subject*> <*subject*>Computer Science</*subject*> <*subject*> Engineering, Chemical</*subject*>

Incorrect mappings to the type and format fields caused changes in the number of these fields in the experiment (examples included below).

While examining the records for changes in practice over time, four significant issues were observed: incorrect mappings from MARC to Dublin Core, misuse of Dublin Core elements, confusion in descriptive metadata and administrative metadata, and information lost in the mapping process.

Incorrect Mapping from MARC to Dublin Core

One MARC field can hold many different kinds of information. For example, information in the MARC 245 field can be mapped to several different Dublin Core elements:

- MARC 245 subfield *a* maps to <dc:title>
- MARC 245 subfield *b* maps to <dc:title>
- MARC 245 subfield *c* maps to <dc:creator> or <dc:contributor>
- MARC 245 subfield *f* maps to <dc:date>
- MARC 245 subfield *g* maps to <dc:date>
- MARC 245 subfield *h* maps to <dc:format>
- MARC 245 subfield *k* maps to <dc:type>
- MARC 245 subfield *n* maps to <dc:description> or <dc:title>
- MARC 245 subfield *p* maps to <dc:description> or <dc:title>

Researchers found a significant number of Dublin Core elements that contained more information than the element should hold. The most common mapping problems happened in following areas.

Merged publisher and date fields: MARC field 260 contains place of publication (subfield a), publisher (subfield b), and publication date (subfield c). In many records mapped from MARC to Dublin Core, all the information was mapped to the publisher field instead of publisher and date fields (e.g., ***<publisher> New York : Aldon Music, Inc., 1951</publisher>***).

Title field containing creator or contributor information: In some cases, it was also found that the Dublin Core title element contained creator information due to subfield c in the MARC 245 field (e.g., ***<title>Frankie/Music by Neil Sedaka; words by Howard Greenfield </title>***).

Incorrect use of delimiter: A rather surprising observation was incorrect use of a comma delimiter in the creator field. This error has not been fixed over the course of the project, indicating a lack of quality control for exposed metadata (e.g., ***<creator> Pica, </creator> <creator>l., </creator> <creator>et al.</creator>***).

Confusion of type and format fields: The MARC 300 field contains physical descriptions that could correctly map to type or format, depending on the item the record describes, and the information contained within each subfield, when the item is not a monograph or serial (subfield a for type and c for format). Analyzed records did not always make this distinction (e.g, ***<type>1 photographic print: b&w; 3 x 4 3/4 in.</type>***).

Misuse of Dublin Core Elements

Previous discussions have focused on misuses of Dublin Core elements, including increases in description and subject fields, and confusion of type and format fields (Godby, Park, Hutt & Riley, 2003). Analysis performed for this survey showed frequent misunderstandings of date, coverage, source, relation, type, format, and description fields.

Date versus coverage fields: Date is often confused with coverage. In some metadata records, publication dates are mapped to the coverage field instead of the date field. The date in the coverage field should refer to the coverage of the intellectual content of the object, rather than the date of the object. An example is a book about the nineteenth century published in 2007. The coverage date should be 1800-1899, and the date value should be 2007.

Source versus relation fields: *F*indings in this survey regarding confusion of source and relation fields are similar to findings in other studies (Yen & Park, 2006). In some of the records analyzed, inconsistent practices were found from single repositories when mapping information regarding the physical collection. Both source and relation fields contained this information in records from different digital collections in the same repository. This information should be placed in the relation field. The source field should only be used for information identifying the original object from which a digital reproduction was created. Additionally, the source of the title from MARC records, usually in the

MARC 500 field, was incorrectly mapped to the source field in Dublin Core records.

Format versus description fields: The MARC 500 field is used for any information related to the object, including medium of the object (more appropriately placed in MARC 300, subfield c). When this information is mapped to Dublin Core, the description field is often used, rather than the format field, due to the semantic meaning of the MARC 500 field, "note."

Example:

Notes Material: Whale Bone

Incorrect:

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

Correct:

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

If following the Dublin Core One-to-One Principle (Hillmann, 2005), this record would describe a physical artifact and not a digital surrogate. See (Digital Library Federation/National Science Digital Library, 2005) for additional discussion of the One-to-One Principle.

Researchers also found many records with incorrect use of the format field. These records use the format field to describe the means of accessing the digital object, rather than the format of the object. Examples of bad practice include:

<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>

Confusion in Descriptive Metadata and Administrative Metadata

Some Dublin Core records observed in this study contain not only descriptive metadata but also administrative metadata. Administrative metadata is usually found in format and description fields, and the values

range from the software used for digitization to master file format and storage equipment. Administrative metadata is useful for data providers to maintain, but is meaningless in the aggregated environment. Best Practices (Digital Library Federation/National Science Digital Library, 2005) advise that administrative metadata should not be exposed in OAI-PMH, and that exposed metadata should be seen as only one "view" of the entire metadata record.

Lost Information

Researchers also observed information that would be helpful in the aggregated environment not being exposed in OAI-PMH. When harvested Dublin Core records were compared to the native metadata records available from the collections, information was found that would have been helpful for search and discovery in the aggregated environment. This problem was observed in several CONTENTdm collections, due to the option of choosing no mapping for local fields. The example below shows that the local element physical description could have been mapped to format and type fields in a Dublin Core record and exported to a service provider.

Example:

Local field:

<*physical description*>xvi, 202 p.: maps, facsims; 20 cm.<*/physical description*>

Correct mapping to Dublin Core fields:

<*format*>xvi, 202 p.<*/format*>
<*format*>20 cm.<*/format*>
<*type*>maps<*/type*>
<*type*>facsims.<*/type*>

Qualitative findings indicate that correct mappings from MARC to Dublin Core would increase the availability of the eight core fields identified in the quantitative section of this article. When comparing harvested Dublin Core records to records in their native context, evidence of mapping semantic values of metadata fields to Dublin Core fields was found, rather than mapping value strings to appropriate Dublin Core fields. Various metadata schemas, including MARC, also contain information in one field that should be mapped to various different Dublin Core fields.

CONCLUSIONS

Although discussions regarding quality metadata and shareable metadata have been prevalent in the professional literature in the past decade, results suggest little change in practice, at least as evidenced in metadata records harvested using OAI-PMH. Based on both quantitative and qualitative analysis of metadata records harvested at UIUC, positive changes in metadata practices over time have not been observed.

Qualitative findings indicate that 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; and correct mapping could improve metadata quality significantly. It is not clear yet why mapping errors are so prevalent. It may be that OAI-PMH mapping practices are poor from the start, or it may be that OAI-PMH mapping practices are not keeping up as local metadata creation practices improve. The lack of change in metadata practices may be due to the fact that OAI-PMH has not been in use long enough for practices to change, or that metadata creators are not receiving strong enough impetus from OAI service providers to change practice. Feedback regarding metadata quality is difficult to give or receive in the current closed system where edited records cannot be replaced by anyone other than the original data provider.

UIUC researchers recommend that data providers publicly document crosswalking and metadata creation practices in order to provide service providers with information appropriate for enhancing harvested records for interoperability in the aggregated environment. Other suggestions include exposing native metadata in addition to Dublin Core metadata so that the service providers can map elements appropriately for their aggregated environment.

Finally, metadata creators should receive appropriate training in order to create shareable metadata. The digital library community has invested much time and effort in discussions regarding shareable metadata since the advent of OAI-PMH, and efforts should be made to ensure that the information is reaching individuals responsible for metadata creation.

Received: September, 2007
Revised: November, 2007
Accepted: December, 2007

AUTHOR NOTE

Amy S. Jackson holds an MS in Library and Information Science from Simmons College, a Masters of Music from the Peabody Conservatory of the Johns Hopkins University, and a Bachelors of Music from the University of Iowa. Myung-Ja Han holds an MS in Library Science from the University of Illinois at Urbana Champaign and a BA in Korean Language and Literature from Sook-Myung Women's University in Seoul. Kurt Groetsch holds an MS in Library Science from the University of Illinois at Urbana Champaign, a Masters in Media Communications from Webster University in St. Louis, Missouri, and BA in Political Science from Augustana College, Rock Island, Illinois. Megan Mustafoff holds an MS in Public Health with a specialization in Epidemiology from the University of Illinois at Urbana Champaign. Timothy W. Cole holds an MS in Library Science from the University of Illinois at Urbana Champaign.

REFERENCES

Bruce, Thomas R. & Diane I. Hillmann (2004). The continuum of metadata quality: Defining, expressing, exploiting. In Diane Hillman & Elaine L. Westbrooks (eds.) *Metadata in Practice*. Chicago: American Library Association.

Bui, Yen & Jung-ran Park (2006). An Assessment of Metadata Quality: A Case Study of the National Science Digital Library Metadata Repository. In Haidar Moukdad (Ed.) *CAIS/ACSI 2006 Information Science Revisited: Approaches to Innovation*. Proceedings of the 2005 annual conference of the Canadian Association for Information Science held with the Congress of the Social Sciences and Humanities of Canada at York University, Toronto, Ontario. June 1 - 3, 2006. http://www.cais-acsi.ca/proceedings/2006/bui_2006.pdf

Chan, Lois Mai & Marcia Lei Zeng (2006). Metadata Interoperability and Standardization – A study of Methodology Part I. In *D-Lib Magazine* 12(6). http://www.dlib.org/dlib/june06/chan/06chan.html.

Chavez, Robert, Timothy W. Cole, Jon Dunn, Muriel Foulonneau, Thomas G. Habing, William Parod, & Thornton Staples (2006). DLF-Aquifer Asset Actions Experiment: Demonstrating Value of Actionable URLs. In *D-Lib Magazine* 12(10). http://www.dlib.org/dlib/october06/cole/10cole.html.

Cole, Timothy W. & Muriel Foulonneau (2007). *Using the Open Archives Initiative Protocol for Metadata Harvesting*. Third Millennium Cataloging, ed. Susan Lazinger and Sheila Intner. Westport, CT: Libraries Unlimited.

Cole, Timothy W. & Sarah L. Shreeves (2004). The Basics of OAI: An Introduction to the Protocol for Metadata Harvesting. Pre-Conference Workshop at Web-Wise 2004: Sharing Digital Resources. Chicago, IL. March 3, 2004

Digital Library Federation/National Science Digital Library (2005). Best Practices for OAI Data Provider Implementations and Shareable Metadata. http://webservices.itcs.umich.edu/mediawiki/oaibp/?PublicTOC

Dushay, Naomi & Diane I. Hillmann (2003). Analyzing metadata for effective use and re-use. In *DC-2003: Proceedings of the International DCMI Metadata Conference and Workshop*. [United States]: DCMI. http://www.siderean.com/dc2003/501_Paper24.pdf

Elings, Mary W. & Gunter Waibel (2007). Metadata for All: Descriptive Standards and Metadata Sharing Across Libraries, Archives and Museums. In *First Monday*, 12(3). http://firstmonday.org/issues/issue12_3/elings/index.html

Godby, Carol Jean, Devon, Smith & Eric Childress (2003). Two Paths to Interoperable Metadata. DC-2003: *Supporting Communities of Discourse and Practice-Metadata Research & Applications*. http://www.oclc.org/research/publications/archive/2003/godby-dc2003.pdf

Hillmann, Diane (2005). Using Dublin Core. http://dublincore.org/documents/usageguide/index.shtml

Hutt, Arwen, & Jenn Riley (2005). Semantics and Syntax of Dublin Core Usage in Open Archives Initiative Data Providers of Cultural Heritage Materials. In *Fifth ACM/IEEE-CS Joint Conference on Digital Libraries 2005* (262-270). New York: ACM Press.

Lagoze, Carl. (2001). Keeping Dublin Core Simple: Cross-Domain Discovery or Resource Description? In *D-Lib Magazine*, 7(1). http://www.dlib.org/dlib/january01/lagoze/01lagoze.html

Lagoze, Carl (2004). The relationship between OAI-PMH and Dublin Core: Required, recommended or other? In *Proceedings CERN Workshop on Innovations in Scholarly Communication: Implementing the Benefits of OAI* (OAI3), CERN (Geneva, Switzerland), February 12-14, 2004. http://eprints.rclis.org/archive/00001000/

Library of Congress. Network Development and MARC Standards Office (2001). MARC to Dublin Core Crosswalk. http://www.loc.gov/marc/marc2dc.html

NISO Framework Advisory Group (2004). *A Framework of Guidance for Building Good Digital Collections*. 2nd edition. Bethesda, MD: National Information Standards Organization. http://www.niso.org/framework/framework2.pdf

Palmer, Carole L., & Ellen M. Knutson (2004). Metadata practices and implications for federated collections. In *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology*, edited by Linda Schamber and Carol L. Barry (456-62). Medford, N.J.: Information Today, Inc.

Palmer, Carole L., Ellen M. Knutson, Michael Twidale, & Oksana Zavalina (2006). Collection Definition in Federated Digital Resource Development. In *Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology*. Vol. 43. Austin, TX.

Palmer, Carole L., Oksana Zavalina, & Megan Mustafoff (2007). Trends in metadata practices: A longitudinal study of collection federation. In *Proceedings of the 2007 Conference on Digital Libraries*. International Conference on Digital Libraries, Vancouver, BC, Canada (386-395). New York, NY: ACM Press.

Park, Jung-ran (2005). Semantic interoperability across digital image collections: A pilot study on metadata mapping. In *CAIS/ACSI 2005 Data, Information, and Knowledge in a Networked World*, edited by Liwen Vaughan. Proceedings of the 2005 annual conference of the Canadian Association for Information Science held with the Congress of the Social Sciences and Humanities of Canada at the University of Western Ontario, London, Ontario, June 2 - 4, 2005. http://www.cais-acsi.ca/proceedings/2005/park_J_2005.pdf

Park, Jung-ran & Sang Joon Park (2005). Digital collection management software employed by libraries and museums: Evaluation of metadata semantic mapping

functionality. Presented at the poster session at ALISE (Association/Library and Information Science Education) annual conference, January 11-14, 2005 in Boston, Massachusetts. http://hdl.handle.net/1860/524

RLG (2004). *Descriptive Metadata Guidelines for RLG Cultural Materials*. http://www.rlg.org/en/pdfs/RLG_desc_metadata.pdf

Shreeves, Sarah L., Ellen M. Knutson, Besiki Stvilia, Carole L. Palmer, Michael B. Twidale, & Timothy W. Cole (2005). 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, (223-237). http://acrl.org/ala/acrl/acrlevents/shreeves05.pdf

Shreeves, Sarah L., Jenn Riley, & Liz Milewicz (2006a). Moving Towards Sharable Metadata. In *First Monday*, 11(8). http://firstmonday.org/issues/issue11_8/shreeves/index.html

Shreeves, Sarah L., Jenn Riley & Liz Milewicz (2006b). Creating Shareable Metadata. Web-Wise 2006 Pre-Conference. Los Angeles, CA. Feb. 15, 2006.

Stvilia, Besiki, Les Gasser, Michael Twidale, Sarah L. Shreeves, & Timothy W. Cole (2004). Metadata quality for federated collections. In *Proceedings of ICIQ04 – 9th International Conference on Information Quality* (111-125). Cambridge, MA.

Van de Sompel, Herbert, Jeffrey A. Young, Thomas B. Hickey (2003). Using OAI-PMH . . . Differently. In *D-Lib Magazine* 9(7/8). http://www.dlib.org/dlib/july03/young/07young.html

Ward, Jewel (2004). Unqualified Dublin Core usage in OAI-PMH data providers. *OCLC Systems and Services: International Digital Library Perspectives*, 20(1), (40-47).

Zeng, Marcia Lei & Lois Mai Chan (2006). Metadata Interoperability and Standardization – A study of Methodology Part II. In *D-Lib Magazine* 12(6). http://www.dlib.org/dlib/june06/chan/06chan.html

doi:10.1300/J517v08n01_02