

Creating Metadata Best Practices for CONTENTdm Users

Myung-Ja Han
University of Illinois, USA
mhan3@illinois.edu

Sheila Bair
Western Michigan
University, USA
sheila.bair@wmich.edu

Jason Lee
OCLC, USA
leeja@oclc.org

Abstract

The OCLC CONTENTdm Metadata Working Group was formed in response to research demonstrating the need for guidelines and best practices for creating quality Dublin Core metadata, which is useful to the primary user community and “shareable” outside of the local context. The CONTENTdm Metadata Working Group has worked since August 2009 in a test environment to develop best practices for creating Dublin Core metadata in CONTENTdm and mapping to MARC to share in WorldCat via the WorldCat Digital Collection Gateway, a self-service OAI harvesting tool. The best practice document identifies 12 core elements and four recommended elements, and provides guidelines for field content standardization and mapping to other metadata standards. Concerns such as adherence to the Dublin Core one-to-one principle regarding the recording of original and digital dates and publishers are also discussed, along with recommendations for configuring metadata for the Digital Collection Gateway to increase the shareability of metadata.

Keywords: Dublin Core; CONTENTdm; metadata; shareable metadata; metadata mapping; best practices; digital library development; OAI-PMH; WorldCat; MARC

1. Background

According to *SPEC Kit 298 Metadata*, CONTENTdm is one of the most widely used digital resource management tools other than locally developed software and DSpace (Ma, 2007, p.26). For digital collection curators, there are two notable advantages for using CONTENTdm as a tool for presenting their digital collections to users, especially in terms of creating metadata. First, it is easy to customize the metadata elements for local practice. Collection curators can use any elements that best describe the items in the collection without being bound to any metadata standard. Secondly, metadata is readily available to aggregators. For example, metadata can easily be enabled for harvesting via OAI-PMH and can also be disseminated in two other ways, i.e., the Digital Collection Gateway allows the export of metadata to WorldCat, and the metadata can be retrieved through the Z39.50 information retrieval protocol (see fig. 1).

However the metadata quality in the service providers’ environments or WorldCat is often different from the local environment. Research has shown issues with losing richness and meaning of native metadata when sharing Dublin Core (DC) metadata in aggregated environments (Shreeves et al., 2006; Jackson, Han, Groetsch, Mustafoff, & Cole, 2008). DC is a simple, flexible scheme promoting data sharing. This flexibility and lack of content guidelines, however, has been problematic as local semantic interpretation, or misinterpretation, of elements can cause challenges when sharing data (Shreeves, Kaczmarek & Cole, 2003; Hutt & Riley, 2005). Though CONTENTdm use facilitates data sharing by making it easy to send data to aggregators, the use of locally defined fields, variations in adherence to the DC one-to-one principle, and semantic problems with mapping can lead to loss of information and impede interoperability (Park, 2005; Han, Cho, Cole & Jackson, 2009). In particular, mapping DC metadata to Machine Readable Cataloging (MARC) standard has long been shown to be challenging (Caplan & Guenther, 1996, Jackson et al., 2008; Beisler & Willis, 2009). Consequently, many researchers have discussed the importance of creating shareable metadata and metadata mapping practices for service providers.

Creating interoperable metadata is highly desirable for both end-users and institutions because it creates wider exposure to unique local collections and allows for the aggregation of subject-specific, but distributed, collections (Shreeves et al., 2005; Shreeves, Riley & Milewicz, 2006; Riley & Shreeves, 2006). Best practices for creating consistent and shareable metadata have been created for using the DC (Hillman, 2005; Bachli, Moser & Vince, 2005; CDP, 2006; Utah Academic Library Consortium, 2006; Wisser, 2007) and for using DC in CONTENTdm (University of Washington Libraries, 2009). Although the best practices for creating shareable metadata have been defined (Digital Library Federation, 2007), there is a need for guidelines in using DC in CONTENTdm with the specific goal of creating shareable metadata (Beisler & Willis, 2009, p. 71).

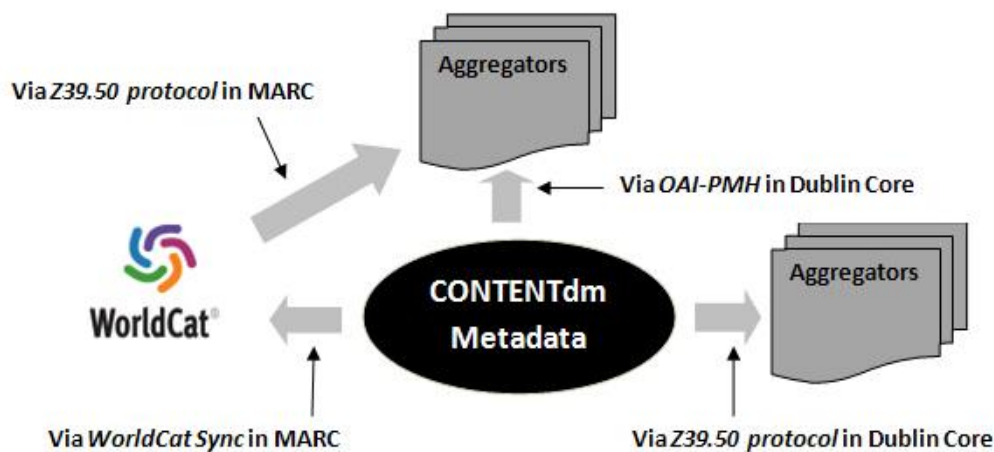


FIG 1. How CONTENTdm metadata travels.

2. Creating Best Practices

Creating sharable metadata has been one of the most frequently discussed topics in CONTENTdm user group meetings¹ and on the ListServ because of its direct impact on resource discovery and retrieval in broad environments. The issue of mapping local metadata to DC has especially been the focus of many discussions since OAI-PMH requires simple DC as a minimum requirement. (CONTENTdm uses qualified DC as its default OAI metadata format to provide richer metadata to service providers.) In addition, CONTENTdm version 5.1 available since spring 2009 added the WorldCat Digital Collection Gateway service, which allows metadata from local digital collections to be directly exported to WorldCat in MARC format by a new function called WorldCat Sync. This raises another mapping issue for CONTENTdm users, i.e., mapping local metadata to MARC. As a result, the need for having a community-wide best practice for metadata was brought up in the Midwest CONTENTdm User Group Meeting in April 2009 and the discussions evolved into a 34 member, and still growing, Metadata Working Group that was convened in August 2009.

2.1. Objectives

The Metadata Working Group had two objectives when creating a metadata best practice document. The first objective was to provide detailed usage guidelines for DC elements that are useful to both new and existing CONTENTdm users in configuring metadata. To ensure the consistent metadata creation, all of the DC elements in the best practice document include

¹ Currently there are four CONTENTdm regional user groups that have hosted annual user group meetings since 2006, and there is a new user group that will have its first meeting in December 2010.

recommended content standards. The second objective was to identify best practices for mapping local fields to DC elements that would improve the interoperability of metadata. Though the best practices recommend the use of DC elements in local environments, the working group realizes that locally created and defined elements may be necessary to best serve different communities and user groups within the local environments. By accomplishing these two objectives, the best practice allows collection curators to use local elements but sustain minimum loss of contextual information when data is harvested or mapped to MARC. Finally, CONTENTdm would implement the document as a default metadata template that helps CONTENTdm users in mapping local fields to DC. The best practice includes two example templates for image and archival collections to show different metadata elements used for describing different types of resources. (See Appendix C of Best Practices Draft 1.9.²)

The group hopes that the best practice document will serve not only the CONTENTdm user group, but also other digital collection creators who use DC as their metadata standard to increase the interoperability of metadata in the DC user community.

2.2. Challenges in Creating Sharable Metadata

In order to preserve the local metadata quality in aggregated environments, collection curators make decisions on what information can and should be shared and how to map local fields to other metadata standards. However, the loss of information in mapping is inevitable (Attig, Copeland & Pelikan, 2004). Mapping from DC to MARC is particularly challenging because of DC's one-to-one principle and because the mapping is done from a semantically simple standard to a semantically rich standard (Library of Congress, 2008). Local collection creators are encouraged to use qualified DC elements to distinguish what manifestation the record describes. In the case of <date>, it is easy to distinguish manifestations since <date> has qualifiers, such as <created> and <issued>. However other elements such as <type>, <format>, and <publisher> do not have qualifiers for different manifestations. In the local CONTENTdm environments, the problem is relatively simple since local field names can contain manifestation specific information, such as <original> and <digital>, but this information can be lost in aggregated environments including WorldCat.

Other challenges mapping DC to MARC include incomplete Leader information, inability to map <creator> and <contributor> to proper MARC fields, and the inability to properly reflect controlled vocabularies in MARC indicators. Because of these reasons the mapping practice from CONTENTdm local fields, or DC elements, to MARC should be examined further to best leverage the local metadata in a union cataloging environment.

In the end, the group acknowledged that decisions on how to make local metadata available to WorldCat in MARC format records should be subject to collection curators' best judgment. The group recommended that collection curators should consider mapping the local metadata fields that are useful for resource discovery and retrieval, and not map local metadata fields that are only useful to local users or collection administrators.

3. Outcomes

The latest document *Best Practices' for CONTENTdm Users Creating Shareable Metadata version 1.9*³, released in June 2010, is now available for download through the Metadata Working Group Wiki and CONTENTdm User Support Center.⁴ The document includes 12 core elements and four recommended elements that could be used when appropriate (see table 1). Each element has a section for name and definition of the element, a corresponding DC element, a list of possible controlled vocabularies to use, and best practices describing content standards. Also

² <http://contentdmmwg.wikispaces.com/file/view/BPG1.9.pdf>

³ <http://contentdmmwg.wikispaces.com/Best+Practices>

⁴ Access to CONTENTdm User Support Center < <http://contentdm.com/USC/resources.asp> > requires user id and password.

included are short quotations from local best practices gathered from group members and data dictionaries to provide practical viewpoints on element usage.

TABLE 1: Core elements for CONTENTdm Collections

Title	Subject	Publisher	Rights
Creator	Identifier	Format	Contributor
Description	Language	Type	Date
Source*	Relation-isPartOf*	Coverage-spatial*	Coverage-temporal*

*Recommended as appropriate

Recommendations for element content standardization include the following preferences:

- **Title:** When there is no given title as defined in traditional cataloging rules, it is not recommended to use explanatory or qualifying symbols (e.g., brackets) to indicate a cataloger-supplied title since most items in image collections do not have titles. The brackets may impede the search and browse functions in both local and aggregated environments.
- **Creator:** When the creator is unknown, instead of adding values' (e.g., "Unknown"), consider leaving the element blank. When the creator has a specific role, it is appropriate to qualify named entities with the role in brackets, (e.g., "[photographer]"). There are mixed views on using qualifiers in a creator field since qualifiers can be indexed together with the creator's name. The group also recommends using qualifiers as field names in local environments and making detailed use of field names available in project pages.
- **Publisher:** When describing a publisher for digital images, add a text prefix, such as "digitized by," to qualify the value and add context, thereby avoiding confusion in the aggregated environment. It is also recommended that if an original publisher is available, map original publisher information to DC <publisher>.

The document also includes three appendices – A. *Moving towards Marketing Metadata*, B. *Using <date> in CONTENTdm*, and C. *Metadata Schemes for Photographic and Archival collections* – that enable CONTENTdm users to not only create sharable metadata but also present a better way to enhance metadata and make collections readily available to the users and aggregators around the world.

4. Future Plans

The group reconvened in July 2010, added new members, and reorganized into emergent task groups to address a number of topics including, but not limited to, qualified DC-to-MARC crosswalk, compound object configuration in CONTENTdm, consortial challenges in the OAI harvesting environment, and the Digital Collection Gateway user interface development and enhancement. The outcomes of discussion topics will be integrated into the Best Practices Guide in support of Version 2.0, scheduled for release in fall 2010. The group will also follow the development of the Resource Description and Access (RDA) standard since it will work as a content standard for the metadata creation once it is implemented.

In addition, the expansion of the Metadata Working Group will also serve as a model of collaborative framework for delivering concrete recommendations for use of the Digital Collection Gateway to a broader constituency.

Acknowledgements

We wish to thank members of Metadata Working Group who shared their experiences and expertise of metadata creation for the project. We also would like to thank Geri Ingram, Taylor Surface and Judy Cobb from OCLC Digital Collection Services for their help throughout the process.

References

- Attig, John, Ann Copeland, and Michael Pelikan. (2004). Context and meaning: The Challenges of metadata for a digital image library within the university. *College & Research Libraries* 65, no. 3: 251-261.
- Bachli, Kelley, Judy Moser, and Pat Vince. (2005). *Dublin Core metadata elements best practices, Version 1.1*. Claremont Colleges. Retrieved February 26, 2010 from <http://ccdlib.org/libraries.claremont.edu/cgi-bin/showfile.exe?CISOROOT=/adl&CISOPTR=29&filename=30.pdf>.
- Beisler, Amalia and Glee Willis. (2009). Beyond theory: Preparing Dublin Core metadata for OAI-PMH harvesting. *Journal of Library Metadata*, 9, 65-97.
- Caplan, Priscilla and Rebecca Guenther. (1996). Metadata for Internet resources: The Dublin Core metadata elements set and its mapping to USMARC. *Cataloging & Classification Quarterly*, 22(3/4), 43-58.
- CDP (Collaborative Digitization Program) Metadata Working Group. (2006). *Dublin Core metadata best practices, Version 2.1.1*. Retrieved February 26, 2010 from <http://www.bcr.org/dps/cdp/best/dublin-core-bp.pdf>
- Digital Library Federation. (2005). *Best practices for shareable metadata*. Retrieved February 26, 2010 from <http://webservices.itcs.umich.edu/mediawiki/oaibp/index.php/ShareableMetadataPublic>
- Han, Myung-Ja, Christine Cho, Timothy W. Cole, and Amy S. Jackson. (2009). Metadata for special collections in CONTENTdm: How to improve interoperability of unique fields through OAI-PMH. *Journal of Library Metadata*, 9, 213-238.
- Hillman, Diane I. (2005). *Using Dublin Core*. Retrieved February 26, 2010 from <http://dublincore.org/documents/usageguide/>
- Hutt, Arwen and Jenn Riley. (2005). Semantics and syntax of Dublin Core usage in Open Archives Initiative data providers of cultural heritage materials. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, Denver, Colorado, June 7-11* (pp. 262-270). New York: ACM Press.
- Jackson, Amy. S., Myung-Ja Han, Kurt Groetsch, Megan Mustafoff, and Timothy W. Cole. (2008). Dublin Core metadata harvested through OAI-PMH. *Journal of Library Metadata*, 8, 5-21.
- Library of Congress. (2008). *Dublin Core to MARC crosswalk*. Retrieved June 30, 2010 from <http://www.loc.gov/marc/dccross.html>
- Ma, Jin and Association of Research Libraries. (2007). *Metadata. SPEC kit*. Washington, D.C: Association of Research Libraries.
- Park, Jung-ran. (2005). Semantic interoperability across digital image collections: A Pilot study on metadata mapping. In V. Liwen (Ed.). *CAIS/ACSI 2005 Data, Information, and Knowledge in a Networked World. 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 -4, 2005*. Retrieved February 26, 2010 from http://www.caisacsi.ca/proceedings/2005/park_J_2005.pdf
- Riley, Jenn and Sarah L. Shreeves. (2006) *Metadata for you and me: Moving towards shareable metadata*. [PowerPoint]. Ohio Valley Group of Technical Services Librarians Conference, May 10-12, 2006. Retrieved February 26, 2010 from http://www.dlib.indiana.edu/~jenlrile/presentations/ovgtsl2006/shareableMD_OVGTSLS.pdf
- Shreeves, Sarah L., Joanne S. Kaczmarek, and Timothy W. Cole. (2003). Harvesting cultural heritage metadata using the OAI Protocol. *Library High Tech*, 21, 159-169.
- Shreeves, Sarah L., Ellen M. Knutson, Besiki Stvilia, Carole L. Palmer, Michael B. Twidale, and Timothy W. Cole. (2005). Is 'quality' metadata, 'shareable' metadata? The implications of local metadata practices for federated collections. *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries*.
- Shreeves, Sarah L., Jenn Riley, and Liz Milewicz. (2006). Moving towards shareable metadata. *First Monday*, 11. Retrieved February 26, 2010 from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1386/1304>
- University of Washington Libraries. (2009). *Metadata guidelines for collections using CONTENTdm*. Retrieved February 26, 2010 from <http://www.lib.washington.edu/msd/mig/advice/default.html#setting>
- Utah Academic Library Consortium. (2006). *Metadata guidelines for the Mountain West Digital Library*. Retrieved February 26, 2010 from http://mwdl.org/public/mwdl/metadata_MWDL2006.pdf
- Wisser, Katherine M. (2007). *NC ECHO Dublin Core implementation guidelines*. Retrieved February 26, 2010 from <http://www.ncecho.org/documents/ncdc2007.pdf>