Testing Resource Description and Access (RDA) with Dublin Core

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

Melanie Wacker
Columbia University, USA
mw2064@columbia.edu

Judith Dartt
University of Chicago, USA
jdartt@uchicago.edu

Abstract

Resource Description and Access (RDA) is a new standard for describing all types of resources. Columbia University, the University of Chicago, and the University of Illinois at Urbana-Champaign evaluated the guidelines by means of the Dublin Core element set during the U.S. National Libraries RDA Test, held from October to December 2010. This paper speaks to the issues which emerged during the test and what each institution did to address them. Test set-ups employed, and tools used, as well as a selection of problems encountered are described in the following summation of findings.

Keywords: Dublin Core; RDA; FRBR; FRAD; content standard; schema; semantic interoperability; Open Metadata Registry

1. Introduction

The Joint Steering Committee (JSC) for the Development of RDA devised Resource Description and Access as a replacement for the Anglo-American Cataloging Rules, Second Edition (AACR2), the library domain's widely used content standard (Oliver, 2010). Though its text is derived from that of AACR2, RDA's structure is based on the conceptual models of the Functional Requirements for Bibliographic Records (FRBR) and the Functional Requirements for Authority Data (FRAD) (Hart, 2010). The new standard is equipped with an element set and value vocabularies, and is not dependent on a particular metadata-encoding scheme (Hillmann et al., 2010). RDA can, therefore, be "encoded using existing schema, such as MARC 21, Dublin Core, [and] MODS, and can also be mapped to other schema, current or future ones" (Oliver, 2010, p. 2).

The JSC published a first draft of RDA in November 2008 (JSC, 2010), and after a public review period, issued a final draft in June 2010 as the RDA Toolkit (http://access.rdatoolkit.org/). Following the toolkit's release, the Library of Congress (LC), the National Library of Medicine (NLM), and the National Agricultural Library (NAL) initiated a nationwide test with several strategic goals in mind. Among these: to assure the operational, technical, and economic feasibility of RDA; and to confirm that it is independent of the format, medium, or system used to store or communicate the data (LC, 2010; U.S. RDA Test Coordinating Committee, 2011, p. 10).

The national libraries and twenty-four official partners, representing libraries, teaching institutions, and vendors took part in the U.S. National Test of RDA (JSC, 2010). These testers were charged with assessing the new standard in a variety of communication formats and schemas (LC, 2010b). At its conclusion, participating institutions had created 10,507 bibliographic records and 12,797 authority records in MAchine Readable Cataloging (MARC) format (Morris, 2011). Sixty-three records were produced employing other metadata standards, thirty-three in the Dublin Core, twenty-eight in Metadata Object Description Schema (MODS), and two with the Encoded Archival Description (EAD). The testers of Columbia University, the University of Chicago, and the University of Illinois at Urbana-Champaign created thirty-two of the thirty-three Dublin Core records submitted.

2. Test set-ups

Mindful of the national libraries' goals, the testers at the three institutions aimed to evaluate RDA as a content standard as well as some of its value vocabularies, to determine if it could be encoded using any descriptive metadata standard. The testers felt it was essential to work with Dublin Core, primarily because it is a metadata standard commonly used by academic and research libraries to describe digital resources (Ma, 2007, p. 27).

All of their Dublin Core records were encoded in eXtensible Markup Language (XML) with the XML editor oXygen. The testers chose this popular tool for two specific reasons. Firstly, it allows users to modify the existing schema straightforwardly, that is, if additional encoding schemes and further refinements are needed for local purposes; and secondly, because building valid well-formed records is a requirement of XML encoding, and can be easily accomplished.

The structure of the RDA test and the types of resources described were different within each institution. At Columbia University, the Metadata Coordinator and four volunteers from other library units took part in non-MARC testing. Two Dublin Core RDA records describing websites were produced, as were records incorporating EAD, and MODS. Employing MODS with RDA was of particular interest to this group, because it is the schema invariably used to create records for the University Libraries' digital collections.

In the University of Chicago Library, the Digitization Manager was the sole creator of non-MARC records. Prior to the test, the Manager focused on gaining proficiency with the online RDA Toolkit. Dublin Core was selected for testing with RDA since it is used to describe the Library's digitized cultural resources. As a first step, a record template employing Dublin Core elements was devised in XML format. Twenty Dublin Core RDA records describing digitized documents from Special Collections were created for the test.

The University of Illinois participated as a teaching institution, collaborating with its Graduate School of Library and Information Science. A Practicum was set up for students who volunteered to test RDA. The program focused on the teaching aspects of the test. Each student was taught to apply cataloging rules, to use the RDA Toolkit, and the tools employed to create records. These consisted of Voyager, the Library's cataloging system, and the Online Computer Library Center (OCLC) for the MARC portion of the test. The student who worked with Dublin Core learned the standard's semantics and schema, in addition to XML. Ten Dublin Core RDA records describing an LC example set, a digital collection, and a digitized image were generated.

3. Findings

3.1. Semantic mapping from RDA to Dublin Core

Though the organization of testing, and the resource types differed, testers at the three institutions found common ground, and exchanged information during the test about record creation and evaluation. When building their Dublin Core RDA records, the testers looked to the *Draft mapping RDA to Dublin Core* and *Notes on the mapping* for basic guidance. The documents were posted to the DC-RDA ListServ in February 2010, and an associated mailing noted that because of remaining unresolved issues, the first release of RDA would not contain an official RDA/DC mapping. Thus, when using these rough copies, the testers were heedful of the notice, and of the additional caveats, such that, "this was a working draft…not…subsequently developed, and… based on a draft of RDA which differs from the final text" (Danskin, 2010).

According to this mapping, the semantic relationship between the RDA elements and those of Dublin Core is many-to-one, much like the relationship between MARC and Dublin Core. In some cases, therefore, an unwieldy array of RDA elements aligns with one or two elements of Dublin Core. For example, eleven RDA elements can be mapped to <dc:title> or <dcterms:alternative>, instances of which are parallel title, variant title, and earlier variant title

(See Table 1). Three RDA elements map to <dc:type>, and nineteen correspond to <dc:format>. By way of contrast, the RDA elements publication date and copyright date were easily accommodated within Dublin Core's <dcterms:created> and <dcterms:dateCopyrighted>. With this use of the element refinements of <dc:date>, the testers felt they were able to provide users with a clear understanding of the data.

It was not possible to satisfy all the requirements of RDA, for example, to describe resources with its required core set of elements. A question representative of how difficult that accommodation might be to make, is to ask how one can add a statement of responsibility (RDA 2.4) to a Dublin Core RDA record of an unpublished archival item.

TABLE 1: Mapping table of <dc:title> and RDA element

Dublin Core element	RDA element
<dc:title>	2.3 Title
	2.3.2 Title proper
<dcterms:alternative>	2.3.3 Parallel title
	2.3.4 Other title information
	2.3.5 Parallel other title information
	2.3.6 Variant title
	2.3.7 Earlier variant title
	2.3.8 Later variant title
	2.3.9 Key title
	2.3.10 Abbreviated title
	2.3.11 Devised title

3.1.1. Roles; Type elements; Contextualization

Because of its dependence on FRBR, RDA, as Diane Hillmann has observed, does place "emphasis on relationships and roles" of entities (Hillmann, 2009, sl. 8). The solution in a MARC record, and one long accepted, is to add relator terms, several examples of which are creator, author, and illustrator. These terms can be appended to the relevant personal and corporate name fields (100, 110, 600, 610, 700, 710) by using the designated subfield e.

Though wanting to add role information specific to the personal and corporate names in their Dublin Core RDA records, the testers found the process somewhat complicated. The efforts of the Dublin Core Metadata Initiative and the Library of Congress have made it possible to employ a number of MARC relator terms in records utilizing the Dublin Core element set. This is achieved by expressing the terms as properties, that is, as elements or element refinements (Using Dublin Core-Appendix, Roles, 2005). To add this kind of record enhancement, however, each institution would have had to edit their local schemas to incorporate appropriate namespaces. This avenue was not pursued during the test, but a discussion concerning local schema extensions is continuing.

When integrating the RDA elements carrier type, media type, and content type into their records, the testers encountered another challenge. The value vocabularies used to record this information, computer and online resource, to name just two, lose meaning in a Dublin Core RDA record due to the many-to -one relationship of the element sets. For the MARC standard, the Library of Congress addressed this issue by introducing three new fields, 336 (content type); the 337 (media type); and, 338 (carrier type). So that the source of the vocabulary is not lost, the data can be captured in subfield 2 of each field (See Table 2). In attempting to provide contextual information within portions of their records, the testers varied their approaches. The methods included: placing type names as qualifiers; linking controlled vocabulary terms and element names together with a colon; or simply putting in the single term.

TABLE 2: RDA elements represented in MARC and Dublin Core

Type information represented in MARC	Type information represented in Dublin Core
336 - - $a text $2 rdacontent	<dc:type>text (content type)</dc:type>
337 - - $a computer $2 rdamedia	<dc:format>computer (media type)</dc:format>
338 - - $a online resource $2 rdacarrier	<dc:format>online resource (carriertype)</dc:format>

3.2. Local schema development

The testers used Qualified Dublin Core schemas formulated at their respective institutions to validate their records. The University of Illinois at Urbana-Champaign Library's schema supports a number of encoding schemes, and it is used in record creation for other digital collection contents. Because of test findings, the libraries of the University of Chicago and Columbia University modified their schemas after the fact. The integration of "ucdcterms" into the schema allows Chicago's practitioners to enrich descriptive records with terms from additional vocabulary encoding schemes, for example, those of the Getty Art and Architecture Thesaurus, <dc:subject xsi:type= "ucdcterms:AAT">documentary papyri </dc:subject>.

Columbia University Libraries' testers thought it important to indicate the source of value vocabularies in their Dublin Core RDA records, especially <rdamedia> and <rdacarrier>, which map to <dc:format>. But such an extension to the official schema was not possible without editing it. Though they mocked up several records to illustrate what the addition might look like, the official test records submitted did not contain this encoding change. Columbia University Libraries Digital Program Division created a local schema which imports the qualified Dublin Core XML schema and adds three locally defined attributes including "vocabURI", and "resourceURI", to enable the use of URIs to specify the values of Dublin Core elements as well as the vocabularies in which the URIs are defined. For <dc:format> they used the RDA vocabularies available through the Open Metadata Registry (http://metadataregistry.org/).

TABLE 3: RDA Type vocabularies represented in <dc:format>

Record submitted to LC	Record after the schema change
<dc:format>computer</dc:format> <dc:format>online resource</dc:format>	<dc:format cul:vocab="RDA Media Type" cul:vocabURI="http://RDVocab.info/termLIst/RDAMediaType" cul:resourceURI="http://RDVocab.info/termLIst/RDAMediaType/1003"> computer </dc:format> <dc:format cul:vocab="RDA Carrier Type" cul:vocabURI="http://RDVocab.info/termList/RDACarrierType" cul:resourceURI="http://RDVocab.info/termList/RDACarrierType/1018"> online resource </dc:format>

3.3. Describing relationships between FRBR entities

RDA derives its concepts, language, and categories from the conceptual models FRBR and FRAD. The assemblage of FRBR entities are: work, expression, manifestation, and item (Group 1); person, family, and corporate body (Group 2); and concept, object, event and place (Group 3). These and the relationship roles of the models are central to the application of RDA (Oliver, 2010). Therefore, in using this standard, a cataloger will need to know what entity the record describes and what level of record to create. FRAD is used to formulate access points and authority records (JSC 2010b, 10.1), and the latter is a critical part of the process when defining the relationships required by RDA rules.

While describing still images in a Dublin Core RDA records, one of the testers realized that designating FRBR entity relationships for images is less of a challenge because images as works link to the self-same images as items. For other types of resources, all of the testers sought ways

to express relationships just as straightforwardly, because these associations provide users with a wealth of contextual data. Yet, fashioning connections was not a simple task in Dublin Core RDA records, because authority record creation is not part of the Dublin Core structure. The testers briefly discussed the feasibility of using or expanding the <dc:relation> element as a means to describe group entities relationships. But because of the test's time constraints, the proposal was not pursued.

4. Looking ahead

The authors applied RDA as a content standard and guidance text with the data structure standard Dublin Core during the nationwide test of RDA. The scope of their testing was restricted to semantic comparisons, and offers proof that the semantic interoperability between RDA and Dublin Core is comparable to that of MARC and Dublin Core, a relation that results in the loss of contextual information.

Broader usage of the new guidelines may be supported if a number of the proposals set forth in the final report of the U.S. RDA Test Coordinating Committee are realized. These include a development of RDA record examples for non-MARC schemas, and mapping of RDA elements to other metadata schemas such as Dublin Core (U.S. RDA Test Coordinating Committee, 2011, pp. 13, 16).

The authors maintain that improving support, such as training, for metadata practitioners who want to apply RDA to the Dublin Core element set (and other standards) would be beneficial. Although *Using Dublin Core* (http://dublincore.org/documents/usageguide/) offers guidance appropriate to the element set, it has been the authors' experience that in-depth rules for descriptive record creation often are essential. Of necessity, they have turned to the standards, such as Cataloging Cultural Objects (CCO) and Describing Archives: A Content Standard (DACS) to name just two.

RDA is more than a content standard, however, and the authors submit that its full potential will never be appreciated if its use is limited, with Dublin Core or with any non-MARC standard. Moving forward, the authors believe the work of the DCMI/RDA Task Group should be looked to, and utilized more fully. This Task Group has focused its efforts on registering RDA elements and value vocabularies in the Open Metadata Registry, thus making them available to other communities and usable in a semantic Web environment (Hillmann et al., 2010). The Task Force's charge also includes the development of an RDA Dublin Core Application Profile. In addition, RDA testing in the eXtensible Catalog (U.S. RDA Test Coordinating Committee, 2011, pp. 179-182) warrants closer attention.

References

Danskin, Alan. (2010, February 22). G'day and question regarding RDA mapping to DC [ListServ]. Retrieved from the DC-RDA Archives: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1002&L=DC-RDA&F=&S=&P=1293

Hart, Amy. (2010). The RDA Primer: A guide for the occasional cataloger. Santa Barbara, Calif.: Linworth, 9.

Hillmann, Diane. (2009). Getting real with RDA [PowerPoint slides]. Retrieved from SlideShare website: http://www.slideshare. net/confirm/MzA4ODE2ODM7bWhhbjU=/1373946-abca1ecef7f263b25098bc8cdf64637ba467b3bc-slideshow

Hillmann Diane, Karen Coyle, Jon Phipps, and. Gordon Dunsire. (2010). RDA Vocabularies: Process, outcome, use. D-Lib Magazine, 16(1/2). Retrieved from D-Lib website: http://www.dlib.org/dlib/january10/hillmann/01hillmann.html doi: 10.1045 /january2010-hillmann

Joint Steering Committee for Development of RDA. (2007). Retrieved from RDA to FRBR mapping website: http://www.rda-jsc.org/docs/5rda-frbrmapping.pdf

Joint Steering Committee for Development of RDA. (2010). Retrieved from Resource Description and Access website: http://www.rda-jsc.org/rda.html#background

Joint Steering Committee for Development of RDA. (2010b). Retrieved from Resource Description and Access website: http://www.rda-jsc.org/rdafaq.html#10-1

Library of Congress. (2010). Retrieved from Testing Resource Description and Access (RDA) website:http://www.loc.gov/bibliographic-future/rda/

Library of Congress. (2010b). U.S. National Libraries RDA Test Partners. Retrieved from http://www.loc.gov/bibliographic-future/rda/test-partners.html

Ma, Jin. (2007). Metadata. Washington, D.C.: Association of Research Libraries.

MARS Authority Control. (2011). [Web log message]. FRBR, FRAD, and RDA what they all mean. Retrieved from http:// ac.bslw.com/community/blog/2011/03/frbr-frad-and-rda-what-they-all-mean/

Morris, Susan. (2011, February 18). US RDA National Library Test: Summary of record Collection. Retrieved from https: //rdatesting.basecamphq.com/projects/3524033/file/73154064/Record%20collection%20by%20the%20numbers_final.doc

Oliver, Chris. (2010). Introducing RDA. a guide to the basics. Chicago: American Library Association, 13;2; 65.

U.S. RDA Test Coordinating Committee. (2011). Report and Recommendations of the U.S. RDA Test Coordinating Committee, 10, 179-182. Retrieved from http://www.loc.gov/bibliographic-future/rda/rdatesting-finalreport-20june2011.pdf

Using Dublin Core-Appendix, Roles (2005, December 07). Retrieved from the Dublin Core Metadata Initiative website: http://dublincore.org/documents/usageguide/appendix_roles.shtml