

To be published in Technicalities, July/Aug. 2007, v. 27, no.4

Great Leaps Forward

Diane I. Hillmann

It's axiomatic that one often doesn't recognize the most important milestones representing significant progress as they're happening, but like most axioms, it isn't always true. As I write this, one of those significant moments has occurred, and having been present, I'm happy to provide some "colour commentary" as they spell it in the country where it happened. I'm speaking here about a relatively small meeting, months in the organizing, scheduled for April 30-May 1, 2007 in the British Library, London. The participants were a varied group representing a number of organizations and constituencies:

1. **Tom Baker** (Dublin Core Metadata Initiative: Chair of the Usage Board, Director Specifications and Documentation, Co-chair DC-Architecture; Co-chair, W3C Semantic Web Deployment Working Group)
2. **Tom Delsey** (RDA Editor. Consultant to: IFLA Working Group on Functional Requirements for Bibliographic Records -- FRBR (1992-1997), the Library of Congress, IFLA and many other international standards groups)
3. **Gordon Dunsire** (Centre for Digital Library Research, Strathclyde University. Member, CILIP-BL Committee on AACR; Chair, Cataloguing and Indexing Group in Scotland; Contributor, DC AP for Eprints)
4. **Diane Hillmann** (Dublin Core Metadata Initiative Liaison to ALA CC-DA; Member, Dublin Core Usage Board and Advisory Board; Editor, "Using Dublin Core")
5. **Alistair Miles** (Editor, Simple Knowledge Organisation System (SKOS); Editor, "Best Practice Recipes for Publishing RDF Vocabularies"; Member, W3C Semantic Web Deployment Working Group)
6. **Mikael Nilsson** (Co-author of DCMI Abstract Model and DC-in-RDF specifications, co-chair of DC-Architecture and of Joint DCMI-IEEE LTSC Task Force (for expressing IEEE LOM using the DCAM))
7. **Andy Powell** (Co-editor, Eprints Application Profile, Co-author, DCMI Abstract Model)
8. **Barbara Tillett** (Joint Steering Committee for RDA)
9. **Robina Clayphan** (Co-ordinator of Bibliographic and Metadata Standards at The British Library; Moderator of DCMI Libraries Community)

The idea for this meeting began at DC2006 in Manzanillo, Mexico, where Don Chatham, representing ALA Editions, the publisher of RDA, got into conversations with some of the Dublin Core group. These conversations, coupled with Don's attendance at other conference discussions led him to suggest that

there were many possibilities for cooperation on the RDA effort, and thus the idea for a meeting to discuss these was proposed. It took a few months to clear schedules and approve dates, budgets, etc., but six months later, the designated participants collected at the British Library, the host for the meetings.

The stated purpose of the meetings was to "... [examine] the fit between RDA: Resource Description and Access and models used in other metadata communities." This may sound to some a rather flimsy reason to get that many busy people together for a day and a half, but a little background might be in order.

It will hardly be a surprise to readers of this column that RDA has been the object of a great deal of criticism, both from those who would like to maintain its ties to the past, and those (like me) who are more focused on the future. As that criticism became more focused on specific issues—in particular the agglomeration of attributes, instruction and value vocabularies incorporated within the textual bondage of the RDA drafts, and the resulting opacity of identification and relationships in the product—many librarians involved in DC were becoming convinced that unless the RDA developers could shift some modicum of focus toward exposing that internal structure, there would be no significant change in the ability of librarians to operate effectively in an increasingly Semantic Web-aware world. The clamor got louder after ALA Midwinter in Seattle, with the heavy discussion about the article by Karen Coyle and myself [1] that was published just before that gathering. The subsequent appearance of the proposed "Framework for a Bibliographic Future" that appeared in Feb./Mar. focused particularly on the formal representation and identification issues. [2]

On the Dublin Core and broader Web Community side, a major focus of effort for the past few years has been on Application Profiles, which require for optimal interoperability the ability to use and reuse formally declared properties and vocabularies. Some communities developing element and property sets for use in metadata have defined such properties, but they are, for the most part, informally declared, e.g., they lack URIs and the kind of human- and machine-available documentation to support use within Application Profiles. Since one important underlying point of APs is to avoid the unnecessary proliferation of similar metadata elements, this lack of reusable properties was problematic, and the fact that RDA was sitting on a veritable treasure chest of such declarable treasures was a significant attraction. Couple that with the increasing interest by Application Profiles developers in integrating the FRBR model as part of a description set oriented focus, and it's possible to see the stars aligning. [3]

The meeting itself began with each group presenting some pertinent information about how they saw the issues at hand, and by after lunch the first day, the possibilities began to look a bit clearer, and a lot more interesting. The growing enthusiasm was such that, at a group dinner at a very nice (but noisy) Italian restaurant, the food fought for attention with the ideas for moving forward that were flying across the selfsame table. The enthusiasm for completing some concrete recommendations by meeting's end is reflected in the fact that the group voted to begin the following day an hour early, at an impressive 8 a.m.

The results of the meeting were announced by the British Library, and consisted of some basic recommendations:

"The meeting participants agreed that RDA and DCMI should work together to build on the existing work of both communities.

The participants recommend that the RDA Committee of Principals and DCMI seek funding for work to develop an RDA Application Profile -- specifically that the following activities be undertaken:

- * development of an RDA Element Vocabulary*
- * development of an RDA DC Application Profile based on FRBR and FRAD*
- * disclosure of RDA Value Vocabularies using RDF/RDFS/SKOS”*

The announcement pointed out some benefits of this activity:

- * the library community gets a metadata standard that is compatible with the Web Architecture and that is fully interoperable with other Semantic Web initiatives*
- * the DCMI community gets a libraries application profile firmly based on the DCAM and FRBR (which will be a high profile exemplar for others to follow)*
- * the Semantic Web community get a significant pool of well thought-out metadata terms to re-use*
- * there is wider uptake of RDA [4]*

All this sounds very well and good, but what does it really mean? The important changes in structure and approach represented by this agreement will, for the most part, be hidden behind cataloger tools, so most changes perceived by catalogers will largely be a result of changes in RDA instructions, and the work that has gone on to re-organize the way instructional material is presented. It may well be that the current model of working directly within a MARC-based template will be the most obvious change, given that MARC cannot at this stage cope with the separation of FRBR Group I entities (work, expression, manifestation and item).

In a recent post to the RDA-L list, Bernhard Eversberg noted:

“Don't forget that MARC tags are the very language of catalogers' conversations, as opposed to the carefully defined terms in the rulebook.” [5]

He goes on to note that many catalogers rely far more on MARC documentation or local documentation than on the AACR2 rules themselves, particularly catalogers whose training has been on the job, rather than through educational channels emphasizing the philosophical basis for cataloging. This reality will be important to consider as the relationship between AACR2 and MARC21 is examined, disentangled, and reconstructed for future use.

What MARC says about itself is instructive. The introduction to *The MARC 21 Formats: Background and Principles* starts with:

1.1. The MARC 21 formats are standards for the representation and communication of bibliographic and related information in machine-readable form.

1.2. A MARC record involves three elements: the record structure, the content designation, and the data content of the record.

1.2.1. The structure of MARC records is an implementation of national and international standards, e.g., Information Interchange Format (ANSI Z39.2) and Format for Information Exchange (ISO 2709).

1.2.2. Content designation, the codes and conventions established to identify explicitly and characterize further the data elements within a record and to support the manipulation of those data, is defined in the MARC 21 formats.

1.2.3. The content of most data elements is defined by standards outside the formats, e.g., Anglo-American Cataloguing Rules, Library of Congress Subject Headings, National Library of Medicine Classification. The content of other data elements, e.g., coded data section 9 below), is defined in the MARC 21 formats. [6]

The original intent of MARC, to ‘represent and communicate’ bibliographic information is certainly still important and valid, but as 1.2.3. above indicates, MARC also contains content, codes and vocabularies. Some of this information is clearly administrative in nature (about the metadata, not about the item described). It’s the parts that don’t express directly what will be in the cataloging instructions that really require analysis and integration with the ‘new’ RDA. As libraries disentangle the legacy of forty years and take advantage of a broader variety of alternatives for encoding and communicating library metadata, we begin to see the potential for engaging with other data communities, enriching their work and ours in the process.

One thing I should say straight out. For some reason, because this agreement was initiated in large part by the Dublin Core community, and most people still [sigh] think of Dublin Core as those !@#\$\$% fifteen elements, there is often a concern that by association with DC, the RDA standard will be “dumbed down.” This is absolutely not the case. One basic requirement for RDA is the ability to express the full complexity of our bibliographic understanding. Another is to be able to bring our legacy data with us into this new world. BOTH those requirements will be far easier to meet with the changes in structure made possible by the agreement between RDA and DCMI. This is no dumbing down—it’s a significant smartening up.

There was a lot of excitement coming out of the London meeting, and we firmly believe that the promise and potential will be met as we move forward with the real work. In London we understood the importance of what we’d done, reflected in the official last act before breaking up and going home. We took a group picture, posed in front of a portrait of Sir Anthony Panizzi, the revered librarian who is considered the father of cataloging. [7]

[1] Coyle, Karen and Diane Hillmann. “Resource Description and Access (RDA) Cataloging Rules for the 20th Century.” DLib Magazine, Jan./Feb. 2007. Available at: <http://www.dlib.org/dlib/january07/coyle/01coyle.html>

[2] “Framework for a Bibliographic Future.” Available at: <http://futurelib.pbwiki.com/Framework>

[3] The ePrints Application Profile is a good example of this. For an overview of the application profile, see Allinson, Julie, Johnston, Pete and Powell, Andy. 'A Dublin Core Application Profile for scholarly works'. Ariadne, vol. 50, January 2007 (<http://www.ariadne.ac.uk/issue50/allinson-et-al/>).

[4] Data Model Meeting. British Library, London 30 April – 1 May 2007. Available at: <http://www.bl.uk/services/bibliographic/meeting.html>

[5] Post by Bernhard Eversberg ev@BIBLIO.TU-BS.DE to Resource Description and Access RDA-L@INFOSERV.NLC-BNC.CA, on Mon, 14 May 2007

[6] The MARC 21 Formats: Background and Principles. Available at: <http://www.loc.gov/marc/96principl.html>

[7] http://en.wikipedia.org/wiki/Antonio_Panizzi [See the Flickr photo: <http://www.flickr.com/photos/14882099@N00/495093866/>]