Diane I. Hillmann
May 15, 2008
To be published in *Technicalities*, July/Aug. 2008

THE EMERGING CATALOGING FUTURE

I spent some time last week at the Massachusetts Library Association meeting in Falmouth, Mass., held at the Sea Crest Resort, right on the beach. Thankfully the first two days were lovely and I spent some quality time walking on that beach, thinking great thoughts. By the morning I left the rains had started, and those who had put off their beach time were out of luck. I had been invited to this conference to talk about "The Emerging Cataloging Future: RDA, DCMI and the Semantic Web," [1] and found there a receptive and engaged group of librarians, eager to figure out where they fit into the world to come.

I can't remember when I last attended a state library conference, so I spent some time hanging out in hallways and in meetings prior to my presentation and after. I was really impressed with how this one was pulled together, and the quality of the speakers and discussions. Some of the good stuff can be gleaned from the Massachusetts Library Association Conference Reports blog [2], which is a great example of a vibrant multi-user blog bringing the information from a conference back to the membership.

The day before my presentation there was a pre-conference on "The Future of the ILS" which featured a keynote by Marshall Breeding and presentations about Georgia Pines, Koha, VuFind, Endeca, Scriblio and WorldCat Local. By the time I arrived at the end of the pre-conference, the buzz was all about open source software and how libraries could think about meeting their needs in a different world. Dinner that evening was followed by a lively discussion of how Massachusetts libraries of all sizes might participate in this revolution and where the leadership for such an effort might be found in Massachusetts.

All this lead very nicely into what I was talking about. As is my usual habit I like to take questions as they come up, so I didn't manage to finish my slides, but the feedback afterwards reassured me that the important ideas had come across regardless. What worked especially well were some concrete "pictures" of how cataloging data might be put together in our future systems (not yet created, but more on that later). As usual, people need some help visualizing abstract things, but once they "get it," they are able to move forward enthusiastically.

The last session that I attended during the conference was the morning I left, when Janet Swan Hill spoke to the group about the recommendations of the Library of Congress Working Group on the Future of Bibliographic Control [3]. Janet spoke primarily about the recommendations addressed to individuals, rather than those addressed to organizations or vendors. To my mind this is one of the strengths of the report--it's not just about what "they" should be doing, it's about what "we" should be doing.

What's Still Missing From This Picture?

Interestingly enough, I got many of the same questions from this group that I get whenever I speak on these issues, and I think it might be useful to use the LC report as a way to frame the

questions. Many of the questions that arise emanate from the various activities around area 3 of the report: "Position our Technology for the Future," and its Desired Outcomes:

"Library bibliographic data will move from the closed database model to the open Web-based model wherein records are addressable by programs and are in formats that can be easily integrated into Web services and computer applications. This will enable libraries to make better use of networked data resources and to take advantage of the relationships that exist (or could be made to exist) among various data sources on the Web.

In coordination with a broad group of interested parties, especially creators of bibliographic data outside of libraries, the library community will develop a record carrier that can interact seamlessly with library data and library systems, and that can be used both by libraries and by other communities that deal in bibliographic data. The carrier format will support a variety of bibliographic control practices and resource types.

The vocabularies developed by the library community will be available for Web discovery and easy reuse by applications developers. Vocabularies will be managed in registries or other structures to facilitate more rapid updates than are possible with centrally managed lists. Knowledge organization systems will facilitate multilingual versions of vocabularies and cross-walking between them.

All data points in the networked environment will be clearly identified, primarily with Uniform Resource Identifiers (URIs). Registration of data points will include information about meaning and usage. The library community will share identifiers of authors, works, and other controlled elements of bibliographic data to enable interchange of data between different communities of use, while still allowing display and indexing of data elements to vary according to the particular needs of the communities concerned. " [3]

Given the above, two big questions come to mind: What are OCLC's plans to accommodate this kind of data? and What are the ILS vendors doing to plan for these changes? I generally have to answer that I don't really know. There's always the possibility that I've missed something, but the lack of openly discussed, specific plans understandably tends to make those in the trenches very nervous.

So what are the possibilities? Just knowing where we want to go--well stated in the LC WG report above, gives us some goals, but doesn't provide much in the way of steps to get there. Nor does it define who has the responsibility to set out those steps, either for the major players or the community at large--the WG was quite clear that they did not have the responsibility to do that. I've been heard to opine that we're headed into a long transition, without much of a map, but that's not terrifically helpful either, no matter how true it might be.

Never being one to sit around passively waiting for someone else to figure things out and provide a plan, I'm going to stick my neck out here and make some of my own recommendations. The LC WG quite clearly recognized that the leadership for the changes coming upon us needs to be far more distributed than it has been, which means that we'll not be able to sit around and wait for LC and OCLC to solve all our problems for us.

In some important respects, the ongoing efforts of the DCMI/RDA Task Group are an instance of this taking up of tasks by parties not usually involved in library leadership. [4] The Task Group itself consists of a small number of people, some funded, some not, with a wide variety of backgrounds. Task group leaders are myself and Gordon Dunsire, from the University of Strathclyde. Everything is done via email and teleconference, since we're about evenly divided between Europe and the U.S., as well as between librarians and Semantic Web folks.

Another obvious place where such "rump" leadership is emerging is in the open source alternatives to traditional library catalogs and catalog front ends. This group has created interest and buzz in the many places that they speak to librarians--most of whom are intensely frustrated by the systems they use presently. I remember distinctly at the ALA conference in Washington last summer, I spoke to a group of library vendors about the DCMI/RDA work and its potential. My enthusiasm was met by far too many blank and uninterested faces (a few "got it," but very few). One vendor representative suggested that his company would wait to see the finished RDA and then decide what to do. My thought then, not expressed out loud, was that by that time the open source guys would be eating their lunch. I still think so, and obviously, so do some of the people I met in Massachusetts last week.

One of the important things that we can all do, aside from the obvious one of keeping abreast of developments by libraries experimenting with new technologies, is keep pushing those that profess to want to provide services to libraries to detail how, exactly, they propose to do so in this changing environment. In other words, don't ask me what OCLC intends to do about RDA and the RDA vocabularies a-building on the NSDL Registry, ask OCLC. [5] Ask OCLC and other vendors about how they intend to integrate user data into the mix of information available about resources. The Library of Congress is by all accounts considering seriously the calls for making their authority data and vocabularies available for use outside traditional library systems, as web-enabled vocabularies (this was an LC WG recommendation as well). We should be asking them to provide more details about their plans for accomplishing this. One question might be: "Will there be a way for libraries using these vocabularies to be notified of changes in some way that they can use to update their metadata?" Another question that should be asked is whether there will be a broadening of the collaborative creation of this data, as the LC WG recommended. Will libraries who create things like name authority data about local institutions, authors, etc., be enabled to contribute them to the general good?

It was heartening to see the extent of librarian interest in the on-the-ground experimentation and development going on in the profession being played out in Massachusetts, and presumably at state and regional meetings all over the country. Few of us seem to be waiting around for turnkey solutions to the challenges we see before us, and a good thing, too.

[1] Hillmann, Diane I. "The Emerging Cataloging Future: RDA, DCMI, and the Semantic Web," Available at: http://www.slideshare.net/smartbroad/emerging-cataloging-future/

[2] Massachusetts Library Association Conference Reports blog. Available at: http://mlamasslib.blogspot.com/

[3] On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control. Available at: http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf

[4] DCMI/RDA Task Group. Available at: http://dublincore.org/dcmirdataskgroup/

[5] For the first pass at RDA elements, see http://metadataregistry.org/schemaprop/list/schema_id/1.html