On the Dublin Core front

by Norm Medeiros
Coordinator for Bibliographic and Digital Services
Haverford College
Haverford, PA

Metadata in a Global World

{A published version of this article appears in the 22:2 (2006) issue of *OCLC Systems & Services*.}

"The second day of a diet is always easier than the first. By the second day you're off it." -- Jackie Gleason

ABSTRACT
This article reviews the DLF/NSDL draft publication, "Best Practices for Shareable Metadata." The article highlights some of the areas where data providers typically fail in ensuring their metadata's interoperability. The article closes with a brief summary of the RDA Forum held at the 2006 ALA Midwinter Meeting.

KEYWORDS
metadata interoperability ; Open Archives Initiative ; OAI ; data providers ; RDA: Resource Description and Access

My wife and I spent the past few months preparing our 20-month-old daughter for the arrival of every child's favorite gift giver, Santa Claus. Our daughter was too young last year to understand Christmas, never mind the idea of a man distributing presents around the world in a sleigh powered by magical reindeer. This year, however, she knows well what gifts are ("mine, mine, MINE!") and has recently discovered that the packaging, as attractive and entertaining as it may be, is outdone by the battery-operated, noise-making, light-parading gadget that rests inside. Thus my wife and I decided to introduce Santa this year (sans magical reindeer, elves, how he manages to fit down the chimney, etc.) during our local civic association's "Santa home visit" event. At the much anticipated moment when Santa appeared at our doorstep, the months of preparation failed to quell the terror that ensued. A scarlet rash ran quickly from our daughter's neck to forehead just before her first blood-curdling shriek bellowed. Gushing tears and hyperventilation followed until my wife was able to shepherd our pride and joy away from the grizzly Santa monster.

Parent lesson #421: expectations often exceed results. The same is true of metadata interoperability. Colleagues and I have recently discussed the viability of sharing metadata within a visual resources application used within my library's local consortium. What we've learned the hard way is that the needs of local communities often conflict with the overarching goal of effective cross-searching and retrieval of information. It's a battle many fight.

BEST PRACTICES FOR SHAREABLE METADATA

As a result of these conversations, I happened upon the draft of "Best Practices for Shareable Metadata" <http://oai-best.comm.nsdl.org/cgi-bin/wiki.pl?PublicTOC>, a document sponsored jointly by the Digital Library Federation (DLF) and the National Science Digital Library (NSDL). The editors of this document, Kat Hagedorn and Sarah Shreeves, join an array of highly qualified contributors in providing a practical, easy to follow set of guidelines for facilitating interoperable descriptive and technical metadata. The document assumes data providers are adhering to the Open Archives Initiative (OAI), but recognizes that most providers want to employ a metadata semantic more descriptive than unqualified Dublin Core (DC), the baseline requirement for an OAI provider.

The document begins with a key observation; that is, high quality metadata may not be highly shareable. In order to achieve shareable metadata, the authors contend that five criteria must be achieved:

- **Proper context** (ensuring every component metadata record can stand on its own contextually)
- **Content coherence** (ensuring that inappropriate metadata – metadata that exists for local purposes only – is excluded)
- **Use of standard vocabularies**
- **Consistency**
- **Technical conformance** (ensuring XML schemas are properly encoded)

Among this list, consistency seems particularly endangered as metadata creation is distributed among staff. Yet the value of conforming to these practices seems well worth the effort. As the authors state, shareable metadata is more likely to appear in a meaningful and appropriate way when intermingled with results from other metadata repositories. They are also more effectively searched, and they increase the means by which users can access the contents they describe. These rewards are likely to occur both in structured OAI environments such as OAIster, as well as in the ever more sophisticated array of general search engines available on the Web.

MYTHS ABOUT OAI AND OTHER TALES

The authors debunk the misconception that the OAI Protocol for Metadata Harvesting deals only with Dublin Core. Although DC is required by the protocol, other richer semantic schemes are harvestable and recommended for use, particularly in light of the criticism unqualified Dublin Core has received over the past few years. Given that many repositories are built on element sets far richer than simple Dublin Core, it makes sense for data providers to offer the richer set of metadata their collection is built upon, along with the mandated Dublin Core.

The authors advise distinguishing between objects whose description should be at the item level, versus those whose attributes are sufficiently similar to allow a broader collection-level description. This suggestion seems basic enough, but it's not hard to imagine cases where data providers may become too energetic in their record creation, thus providing unnecessarily granular access.

In addition to an annotated list of metadata schemes for use with OAI, the authors describe the important practice of crosswalking, the means by which elements are mapped from one metadata scheme to another. Crosswalking may result in a one-to-one correspondence; that is, the element in one's local system matches exactly with an element from the destination system. Sometimes,

however, an element in one's local system must be parsed into separate fields within the destination system. For instance, a local database may record an author's name in a single field, while the destination system may require the given name and surname be recorded in separate fields. A third case cited by the authors occurs when a local field allows multiple values, but the destination field requires these values be placed into separate fields. For instance, if my local system had a keyword field that could be populated with multiple, comma-separated values, it's possible the destination system would require these values be placed into separate fields. Given this possibility, it's preferable to use separate elements for values, even when the system being used allows multiple values per field.

RDA UPDATE

Much of the remainder of this excellent document discusses use of content rules to encode values. At the recent ALA Midwinter Meeting in San Antonio, the ALCTS/CCS Executive Committee sponsored a forum on *RDA: Research Description and Access*. This inaugural forum, an event slated to continue at each ALA meeting until a year after *RDA* is released, provided an update on the progress of the new content rules, as well as an opportunity for members of the audience to ask questions of Jennifer Bowen and Barbara Tillett, the ALA and Library of Congress JSC representatives, respectively. Ms. Bowen's overview of the reasons for moving to *RDA* rather than *AACR3* was clear, as was her description of the process for how decisions about the new rules are made. She addressed many of the common issues concerning *RDA* before opening up the floor to questions. General comments were positive. There seems to be a realization that *RDA*, by being a simpler set of rules that are geared towards describing digital objects, may become a very valuable resource to communities outside of librarianship. Although some catalogers are concerned that *RDA*'s simplification is a disservice to library catalogs, my sense is that even those who feel this way recognize that *RDA* may be of service to other communities and thus be a powerful resource with international impact. Additional information about the scope and progress of *RDA* is available at <http://www.collectionscanada.ca/jsc/news.html>.