Keeping libraries relevant in the Semantic Web with resource description and access (RDA)

Cataloging is not just building a catalog, but about providing users with timely access to information relevant to their needs. The task of identifying resources collected by libraries, archives and museums results in rich metadata that can be reused for many purposes. It involves describing resources and showing their relationships to persons, families, corporate bodies and other resources, thereby enabling users to navigate through surrogates to more quickly get information they need. The metadata constructed throughout the life cycle of a resource is especially valuable to many types of users, from creators of resources to publishers, subscription agents, book vendors, resource aggregators, system vendors, libraries and other cultural institutions, and end users. The new international cataloging code, RDA (resource description and access), is designed to meet fundamental user tasks in a way that produces well-formed, interconnected metadata for the digital environment.

BARBARA B TILLETT
Chief, Policy and Standards Division
Library of Congress

Introduction

Libraries have made great strides towards a web presence, but many offer only an electronic version of their card catalogs. Linear displays of citations to holdings may include a link to a digitized version of the described resource, but typically excludes machine-actionable connections. Citation-based catalogs need to describe resources by their identifying characteristics in a way that computer systems can understand and show relationships to persons, families, corporate bodies and other resources. This will enable users to navigate through linked surrogates of resources to get information more quickly. It also will better enable systems to make cataloging easier.

Since mid-2010, RDA (resource description and access) has offered an alternative to past cataloging practices. This new code for identifying resources has emerged from years of international collaborations, and it produces well-formed, interconnected metadata for the digital environment, offering a way to keep libraries relevant in the Semantic Web.

How did we get to this point?

RDA is built on the traditions of the *Anglo-American Cataloguing Rules* (AACR). The Joint Steering Committee for Development of RDA (JSC) recognized during the 1990s that AACR2 (the second edition of AACR) was not a code that would serve 21st-century users. It was structured around card catalogs and linear displays of citations, created before the internet and well-formed metadata that could be used by computer systems.

During the 1990s, the JSC received many complaints about AACR2, which:

- had become increasingly complex as updates were added, particularly to address new digital resources
- lacked a logical structure and instead focused on individual rules for each type of material rather than on commonalities and basic principles for a simplified, consistent approach
- was arranged by class of materials, which caused problems when cataloging e-resources with multiple characteristics
- did not adequately address bibliographic relationships, whereas the web is all about networks of interconnected information

■ displayed a strong Anglo-American bias, even though it is used around the world

■ segregated bibliographic data from the rest of the information community's data, in a world of its own with MARC- (MAchine-Readable Cataloging[1]) formatted records. Although MARC is widely used among libraries worldwide, it is not used by the larger information community

■ had terminology for describing materials which was a mix of content and carrier data types. These terms were irregularly applied, with North American catalogers following different practices than elsewhere.

In response to these complaints, the JSC called an international conference on the 'Principles and Future Development of AACR' for cataloging rule-makers and experts from around the world. As a result of this 1997 Toronto meeting, specific problems were identified and a strategic plan was put in place for future directions. Work began to develop AACR3, keeping the same structure as AACR2 and incorporating the recommended changes.

By April 2005, after an initial draft of AACR3 went out for worldwide comments, the JSC received a negative response. People felt the JSC had not gone far enough to embrace the new conceptual models and vocabulary emerging from the international efforts within IFLA (International Federation of Library Associations and Institutions). In particular, there were calls for more attention to IFLA conceptual models, FRBR and FRAD (*Functional Requirements for Bibliographic Records and Functional Requirements for Authority Data*)[2]. Those conceptual models brought a new perspective on describing resources to focus on the content and carriers and view persons, families, and corporate bodies associated with them in terms of their identifying characteristics. The FRBR entities and relationships and the vocabulary used to describe them were important to the international community of responders. One of the key aspects coming from the conceptual models was a focus on using the identifying characteristics in describing resources to meet basic user tasks: find, identify, select and obtain. Moreover, a call to move to an element-based approach to metadata, rather than building citations, was more compatible with metadata services for web use in the broader information community; it fitted nicely with the entity-relationship approach of IFLA's conceptual models.

Simultaneously, IFLA's work towards *International Cataloguing Principles*[3] was well underway to review the basic 1961 'Paris principles'. Five regional conferences were held between 2003-2007 with rule-makers and cataloging experts worldwide to develop the new *International Cataloguing Principles* of 2008, which are part of the foundation for RDA.

RDA emerged in response to worldwide comments from and beyond the Anglo-American community of libraries and other information agencies: publishers, book dealers, archives, museums, developers of web services, and more. It is built on the idea of reusing identifying information coming from publishers and vendors, building on descriptions and making relationships not just by libraries but all stakeholders in the information chain.

Collaborations

The JSC initiated collaborations with special communities:

■ concern about AACR2 dealing inadequately with seriality resulted in the harmonization of ISBD, ISSN and AACR2 standards; those discussions were set to be resumed during 2011 in the light of RDA

■ content-, media-, and carrier-type terminology was addressed with the publishing community. The result was the RDA/ONIX Framework and a plan for ongoing review and revision of that controlled vocabulary to share consistent data

■ controlled vocabularies were addressed by representatives from the JSC, Dublin Core, IEEE/LOM and Semantic Web communities. This resulted in the DCMI/RDA Task Group to develop a registry of the RDA vocabularies and a library application profile for RDA. The controlled vocabularies and element set from RDA are now available as a registry on the web, as a first step to making library data accessible in the Semantic Web environment

■ the JSC also met with various library and archive communities to initiate discussions about more principle-based approaches to describing their collections. An example of changes resulting from those discussions was the approach to identifying the *Bible* and books of the *Bible,* so they could be better understood

by users and more accurately reflect the contained works. The JSC is resuming discussions with the law, cartographic, religion, music, rare book and publishing communities to propose further improvements to RDA.

Technical developments

FRBR-based systems have existed, been tested and used worldwide for over a decade to enable collocation and navigation of bibliographic data. Some examples are systems developed by the National Library of Australia, the VTLS Virtua system, the linked data services of the National Library of Sweden, and the music catalog of Indiana University's Variations 3 project. The Dublin Core Abstract Model is built on the FRBR foundation. RDA positions libraries to enter that realm. Recent research articles reaffirm the use of FRBR as a conceptual basis for cataloging in the future.[4]

It is important that libraries join the rest of the information community on the web to share their expertise, multi-lingual controlled vocabularies and organizational skills. The element-based approach of RDA facilitates identifying persons, families, corporate bodies and works in a manner that machines can more easily use. Controlled vocabularies for RDA are posted on the web as 'registries' along with other controlled vocabularies from our traditional authority files. For example, freely available authority data from hundreds of national libraries and other institutions now resides in the Virtual International Authority File[5]. VIAF includes names and identifying data for persons, corporate bodies/conferences and uniform titles, and demonstrates how library metadata can be reused and packaged in new ways. It provides a multilingual, multiscript base that has the potential to serve as a switching mechanism to display the language and script a user prefers, assigning a distinctive universal resource identifier (URI) to each entity. Although VIAF can manipulate authority data from various schema or communication formats, having the data clearly identified (as RDA does) will make it easier for services like VIAF and future linked data systems to use the specific identifying characteristics to describe persons, corporate bodies, works, etc. RDA will make it easier for machines to use that data to link related information and to display information users want.

The RDA registries include terms for description and access elements, such as title proper, date of publication and extent, as well as values for specific elements, such as the terms to use when describing types of carriers (e.g., computer disc, volume, microfiche, videodisc). These are posted on the Open Metadata Registry[6], giving URIs for all of the terms, which then can be used in the Semantic Web to enable greater use by web services. This positions the library community to move access to its resources out of the silos of data used only by other libraries to the web.

So what is different?

Some of the differences between AACR and RDA can be summed up as follows:

- AACR2 said it was based on principles but never specified what those principles were. RDA is based on IFLA's *International Cataloguing Principles* and describes the principles for each section of elements. For example, RDA follows the ICP principle of representation, instructing catalogers to transcribe what they see (e.g., title proper, statement of responsibility, publication statement). This saves time and builds on existing creator, publisher or vendor metadata

- the principle of common usage means no more Latin abbreviations, such as s.l., s.n., and no more English abbreviations, such as col. and ill., which users do not understand

- RDA relies on cataloger's judgment to make decisions about how much description or access is warranted. For example, AACR2's 'rule of 3' provides up to only three authors, and that now is an alternative rather than the main instruction in RDA. Thus RDA encourages access to the names of more persons, corporate bodies and families important to users. RDA ties every descriptive and access element to the relevant FRBR user tasks: find, identify, select, obtain, in order to develop cataloger's judgment to know not only *what* identifying characteristic to provide, but *why* they are providing it (to meet a user need)

- RDA requires that catalogers name the contained work and expression as well as the creator of the work when appropriate. The concept of 'main entry' disappears

■ RDA provides authority data instructions, which were not covered in AACR2. RDA states the 'core' identifying characteristics that must be given to identify persons, families, corporate bodies, works, expressions, etc. In addition, other characteristics may be provided when readily available: the headquarters location for corporate bodies, or the content type for expressions, such as text, performed music, still image, cartographic image

■ identifying characteristics or elements are separate from the authorized access points that may need to be created while the MARC-based environment persists. RDA describes how to establish authorized access points, but it does not require them, instead, looking toward a future where the identifying characteristics needed to find and identify an entity can be selected for the context of a search query or display of results

■ important for the web which is all about relationships, RDA provides relationship designators to explicitly state the role a person, family, or corporate body plays with respect to the resource being described. It enables description of how various works are related, such as derivative works to link motion pictures or books based on other works, musical works and their librettos, and to link textual works and their adaptations. It connects the pieces of serial works in successive relationships through title changes. The inherent relationships connect the contained intellectual and artistic content to the various physical manifestations, such as paper print, digital and microform versions.

The US RDA test

Although the Library of Congress (LC) publicly committed to implementation of RDA in 2007 in a joint statement with the British Library, the Library and Archives Canada and the National Library of Australia[7], that commitment had to be postponed. In response to the 2008 report to LC from the Working Group on the Future of Bibliographic Control[8] recommending all work on RDA be stopped, LC, the National Library of Medicine and the National Agricultural Library instead launched a US test of RDA to explore whether or not to implement the new code. This included gathering

information about the technical, operational and financial implications of implementation.

In preparation for the test, LC provided 'train-the-trainer' modules[9] and examples, which are freely available as Webcasts, PowerPoint presentations, and Word documents[10]. LC's Policy and Standards Division also set up an e-mail address that remains available at LChelp4rda@loc.gov for anyone to ask questions about the RDA instructions and LC policies for RDA. Initial policy decisions for the test were established and posted on the website as well as in the RDA Toolkit[11] that supplies RDA instructions. Those LC policy decisions are now being adjusted, informed by the test results and feedback from participants in conjunction with discussions with the Program for Cooperative Cataloging and preliminary suggestions from the Library and Archives Canada and the National Library of Australia regarding their implementation decisions.

Twenty-six US RDA test participants included many sizes and types of libraries, as well as archives, museums, book dealers, library schools, system vendors, consortia and funnel projects. They created 10,570 bibliographic records and 12,800 authority records and filled out more than 8,000 surveys. The analysis of that data provided helpful feedback for needed improvements to the RDA Toolkit, to the language used to convey the instructions, and suggestions for moving beyond the current MARC format.

The report from that test recommended implementation no sooner than January 2013 provided certain conditions were met.[12] Those conditions were stated as recommendations to the JSC, the American Library Association (ALA) Publishers who created the RDA Toolkit, system vendors, the Program for Cooperative Cataloging, and the senior managers at the Library of Congress, the National Library of Medicine and the National Agricultural Library.

The test had not specifically focused on the MARC format, but responses from the participants made it clear that the MARC format was seen as a barrier to achieving the potential benefits of the code to move libraries onto the web. As a result, one of the recommendations was to show credible progress towards a replacement for MARC. Work is well underway towards that end through the new Library of Congress initiative, 'Transforming the Bibliographic Framework'.[13]

Implementation of RDA

Eight institutions participating in the test decided to continue to use RDA, regardless of the test recommendations. Their bibliographic and authority records are being added to bibliographic utilities, such as SkyRiver and OCLC, and are available for copy cataloging.

The Library of Congress had about 50 catalogers engaged in the test. They will resume using RDA in November 2011 in order to assist with training and writing proposals to improve the code, as well as to inform related policy decisions.

Many Europeans also expressed interest in learning more about RDA. Several countries joined EURIG, the European RDA Interest Group, which held conferences before the 2010 and 2011 IFLA meetings to share news. These interested parties are also expected to submit proposals to improve RDA, and the JSC had, at time of writing, already received one such proposal for review in 2011.

Translations of RDA are underway so people will be able to read RDA in their own language. Translations are expected for Spanish, French and German, among others. Anyone interested in translating RDA into another language should contact Troy Linker at ALA Publishing (tlinker@ala.org).

In recognition of the international intentions for RDA, the governance for the JSC will be expanded to include one to three new members from countries that intend to implement the code. Those interested in participating should contact a member of the Committee of Principals (CoP), the group that oversees JSC activities. The CoP includes representatives from ALA, Canadian Library Association, the Chartered Institute of Library and Information Professionals (CILIP), Library of Congress, Library and Archives Canada, the British Library and National Library of Australia.

Conclusion

Libraries are in danger of being marginalized by other information delivery services as they have not had a strong presence with other services in the information community on the web. Bibliographic control is based on the MARC format, which is not suited to the Semantic Web environment. For example, MARC is not granular enough to distinguish among different types of dates, and it puts many types of identifying data into a general note which cannot easily be parsed for machine manipulation.

Current online catalogs are no more than electronic versions of card catalogs with similar linear displays of textual information. Yet, the metadata libraries provide could be repackaged into interesting visual information, such as time-lines for publication histories and maps of the world to show places of publication, similar to VIAF displays. Librarians could build links between works and expressions – like original works and their translations or novels that form the basis for screenplays – to navigate these relationships rather than rely on textual notes that are not machine-actionable. Libraries can make their data more accessible on the web.

In order to help reduce the costs of cataloging, librarians need to reuse publisher and vendor metadata. Libraries must share metadata more than they have to reduce the costly, redundant creation and maintenance of bibliographic and authority data. RDA positions libraries for a linked data scenario of sharing descriptive and authority data through the web to reuse for context-sensitive displays that meet user needs for languages/ scripts they can read.

By providing well-formed metadata that can be packaged into various schema for use in the web environment, RDA offers a data element set for all types of materials. It is based on internationally agreed principles, incorporating the entities and relationships from IFLA's conceptual models. It focuses on the commonalities across all types of resources while providing special instructions when there are different needs for types of resources, such as music, cartographic, legal, religious and rare materials and archives, or refers to specialized manuals for more granular description of such materials.

Vendors and libraries around the world are being encouraged to develop better systems that build on RDA. Once RDA is adopted, systems can be redesigned for today's technical environment, moving libraries into linked data information discovery and navigation systems in the internet environment and away from online public access catalogs (OPACs) with only linear displays of textual data.

This is a transition period when libraries want and need to move bibliographic data to the web for

use and re-use. RDA is not the complete solution, but its role as a new kind of content standard may smooth the path in that direction.

Two other components are also needed: firstly, an encoding schema that maintains the integrity of RDA's well-labeled metadata – the aforementioned transition from MARC – and, secondly, systems that can accommodate RDA to harness its full potential to express relationships among resources.

Library administrators need to understand that the full benefits of investment in these components now will not be realized immediately, but the investment is critical to the future health and role of libraries.

RDA makes library bibliographic descriptions and access data more internationally acceptable. There is still more work to be done, but the direction is set.

References

1. *'The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form'*. MARC Standards at:
   http://www.loc.gov/marc/ (accessed 26 September 2011).

2. *Functional Requirements for Bibliographic Records, Final Report*. IFLA Study Group on the Functional Requirements for Bibliographic Records. Approved by the Standing Committee of the IFLA Section on Cataloguing, September 1997 as amended and corrected through February 2009, p. 79. PDF available at:
   http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf (accessed 26 September 2011).
   and
   Functional Requirements for Authority Data, a Conceptual Model. Final report, Dec. 2008. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2009, Munich, Saur.

3. *IFLA Cataloguing Principles:* the Statement of International Cataloguing Principles (ICP) and its Glossary in 20 languages, ed. Tillett, B B and Cristán, A L , 2009, Munich, Saur, p. 28.

4. Pisanski, J and Žumer, M, Mental Models of the Bibliographic Universe. Part 1: Mental Models of Descriptions, 2010, *Journal of Documentation*, 66(5), 643–667 and Pisanski, J and Žumer, M, Mental

Models of the Bibliographic Universe. Part 2: Comparison Task and Conclusions, 2010, *Journal of Documentation*, 66(5), 668–680.

5. VIAF. Available at:
   http://viaf.org (accessed 26 September 2011).

6. Open Metadata Registry, RDA vocabularies at:
   http://metadataregistry.org/rdabrowse.htm (accessed 26 September 2011).

7. *Joint Statement of Anglo-Heritage National Libraries on Coordinated RDA Implementation*, 22 October 2007. Available at:
   http://www.rda-jsc.org/rdaimpl.html (accessed 26 September 2011).

8. *On the Record.* Report of the Library of Congress Working Group on the Future of Bibliographic Control, January 2008. PDF available at:
   http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf (accessed 26 September 2011).

9. RDA test 'Train the Trainer' (training modules). Presented by Kuhagen, J and Tillett, B B, 15 January 2010; Northeastern University, Boston, Mass.
   Modules 1-9
   http://www.loc.gov/bibliographic-future/rda/trainthetrainer.html (accessed 26 September 2011).
   PowerPoint files of the Modules (with speaker's notes) and accompanying material are freely available at:
   http://www.loc.gov/catdir/cpso/RDAtest/rdatraining.html (accessed 26 September 2011).
   - Module 1: What RDA Is and Isn't
   - Module 2: Structure
   - Module 3: Description of Manifestations and Items
   - Module 4: Identifying Works, Expressions, and Manifestations
   - Module 5: Identifying Persons
   - Module 6: Identifying Families (*filmed at the Library of Congress, 1 March 2010*)
   - Module 7: Identifying Corporate Bodies
   - Module 8: Relationships
   - Module 9: Review of Main Concepts, Changes, Etc.

10. US RDA test website is known as 'Testing Resource Description and Access (RDA)':
    http://www.loc.gov/bibliographic-future/rda/ (accessed 26 September 2011).

11. RDA Toolkit:
    http://www.rdatoolkit.org (accessed 14 October 2011).

12. *Report and Recommendations of the US RDA Test Coordinating Committee*, 9 May 2011, revised for

public release 20 June 2011. PDF available at:
http://www.loc.gov/bibliographic-future/rda/
rdatesting-finalreport-20june2011.pdf (accessed
14 October 2011).

13. *Bibliographic Framework Transition Initiative*. Available
at:
http://www.loc.gov/marc/transition/ (accessed
14 October 2011).

■ **Barbara B Tillett, PhD**
Chief, Policy & Standards Division
Library of Congress
101 Independence Ave., SE
Washington, D.C. 20540-4260
USA
Tel: +1 (202) 707-4714
E-mail: btil@loc.gov