Moving to Linked Data: Cataloguing Outside the (MARC) Box

Ian Bigellow (Georgian College), Stacey Boileau (OCLS), Danielle Emon (Loyalist College), Rosina Leung (Seneca College)

Marina Morgan (Mohawk College), & Irene Sillius (Sheridan College)

INTRODUCTION

Linked data for libraries has been getting a lot of attention lately. Linked data offers the promise to free library metadata from its existing silos, making it available where our users are (the search engines). At the same time, linked data will incorporate the sharing of metadata in ways MARC has never allowed, bringing the opportunity for richer description, while simultaneously enhancing information retrieval through serendipitous browsing or semantic search.

With its underlying FRBR model, Resource Description and Access (RDA) places an emphasis on describing discrete metadata elements rather than records, using authorized access points, identifiers, and above all, joining them with relationships. Through this approach RDA provides a bridge for libraries to move towards linked data, however, libraries need to embrace these concepts, and many of us still exist in a MARC vacuum. Linked data and the semantic web are on the horizon for libraries, but how do cataloguers plan for this change?

Following RDA implementation, Ontario Colleges and the Bibliographic Standards Working Group (BSWG) set about updating local standards, working on an awareness campaign, training, and developing the OCLS Cataloguing Workflow to replace an existing manual.

While the OCLS Cataloguing Workflow's primary purpose is to adapt documentation for RDA, throughout the development of the workflow there was a certain tension evident between detailing current practice and planning for future changes. Ultimately it only seemed prudent to also take some small steps in adapting practice for linked data. This poster aims to outline several steps that can be taken now to prepare for linked data through the creation of descriptive metadata.

CONTEXT: RDA & LINKED DATA

RDA is here and linked data will be a new reality for libraries. This may be through BIBFRAME, or other applications (Schema.org, ALIADA, or local development).

"RDA represents a revision to AACR2 that implements FRBR concepts and incorporates FRBR terminology. The convergence of RDA and FRBR will focus cataloging, or resource description, on the resources' relationships with each other and steer the process of retrieval and access toward navigating links through a hierarchy of relationships. This change in emphasis positions libraries to participate in the emerging Semantic Web." (Szeto, 2013, p. 306)

However, RDA will only position libraries to participate in the semantic web if we embrace these aspects of RDA. Consequently, the OCLS Cataloguing Workflow was developed with future data needs in mind, in order to help us move our metadata from closed database silos,

To linked data on the semantic web:

THE TRANSITION

AACR2 in MARC/RDA in MARC/RDA in RDF

In order to prepare for future bibliographic metadata needs, we need to:
1. Acknowledge that MARC will be here for some time by continuing to provide instruction on using RDA in a MARC environment.
2. Ensure that our minimum content standards stay robust in the new linking and encoding environments.

MARC (present)

Library metadata in closed databases - very useful for resource discovery, but effectively invisible on the web.

MARC transformed into RDF

MARC living in parallel to linked data: through processes such as that used by the LibHub initiative, MARC data is transformed into RDF, making collection visible online through search engines.

Library metadata native to the web (Future)

An example of the BIBFRAME editor: library metadata native to the web, with RDA content standards preserved. This will allow for more visibility of library data, along with the potential to enrich resource discovery through linking of data sets and semantic search.

WHAT CAN WE DO NOW: A NEW APPROACH TO WORKFLOW

Step 1: Content First

RDA with underlying FRBR outlines critical content for resource description, which lends itself to usage with many metadata formats. Libraries have been using MARC for a very long time, but the future may hold the potential for library metadata in many forms, stressing the importance for content driven workflows - cataloguing outside the (MARC) box!

While previous manuals have focused on a MARC driven approach (sometimes to the point that AACR was marginalized), the OCLS Cataloguing Workflow instead directs the user on how to record the requisite content. MARC instruction and examples are provided, reflecting our current operating environment, but in a more modular way so it can be lifted out of the workflow in future.

Tools like RIMMF (RDA in many metadata formats) highlight how FRBR can help organize our data and the focus on a content driven approach. RIMMF will be explored into several formats as illustrated below.

5.1 Construct the Authorised Access Point Representing a Work RDA 6.27.1
(1XX, 240, 245, 7XX MARC Fields)

This is a requirement for the Minimum Standard of the College Union Catalogue.

Definition: An Authorized Access Point is defined as "the standardized access point representing an entity". It is not an RDA element, but a name, term, code, etc., representing a specific entity and is created by linking relevant components.

Authorised access point for works include the following components:
1. Authorised access point for the preferred title RDA 6.2.2
2. Additions to the preferred title as instructed under RDA 6.27.1.9
3. Authorised access point for the creator RDA 19.2.1.1

Note: In order to construct the authorised access point for works, you will need to first develop AAPs for creators/preferred titles.

RDA, with its underlying FRBR model, addressed the description of standard elements for library metadata. In a linked environment this means that key FRBR elements (Work, Expression, Manifestation, et cetera) can be more readily collocated for enhanced resource discovery and facetting, and linked to other data sets that can further enhance the research process. In this light, the OCLS Cataloguing Workflow uses a FRBR WEMI (Work, Expression, Manifestation, Item) structure to outline the process of resource description, while covering the description of FRBR group 1,2, & 3 entities throughout.

Key Workflow FRBR Elements

Step 5: Work	Authorized Access Point for Work	Group 1 & 2
	Record Other Creators responsible for Work	Group 2
	Other persons, families, corporate bodies associated with work	Group 2
	Related works	Group 1
	Relationships to concepts, objects, events and places	Group 3
Step 6: Expression	Authorized access point for expression	Group 1
	Record AAP for contributors to expression, e.g.,	Group 2
	Related expressions	Group 1
Step 7: Manifestation	Title and statement of responsibility	Group 1
	Publication/Distribution/Manufacturing/Production and Copyright	Group 1
	Record manifestation level relationships as needed	Group 2
Step 8: Item	Local practices as needed.	Group 2

Group 1 Entities:	Work, Expression, Manifestation, Item
Group 2 Entities:	Person, Corporate Body
Group 3 Entities:	Concept, Object, Event, Place

Step 2: WEMI

Initial efforts in the development of the OCLS Cataloguing Workflow focused on describing the manifestation in hand first, however, the workflow was changed to a WEMI order so that:
1. Describing the work first requires the cataloguer to ask some of the big questions about a resource upfront, describing AAP and relationships for the work, then Expression, that are often neglected in a manifestation driven approach.
2. A WEMI approach is in close alignment with the development of BIBFRAME.

Step 3: AAP, Identifiers, and Relationships

With the development of the OCLS Cataloguing Workflow, in addition to emphasizing RDA's underlying FRBR model, it became clear that in order to prepare for upcoming metadata needs, RDA's usage of authorized access points, identifiers and relationships (AIR) would need to be stressed.

The Resource Description Framework (RDF) used by linked data, organizes data into a series of triples, in the form:

"Catalogers must create data in MARC that will allow it to be batch processed into RDF automatically with the least amount of difficulty. Here again they can look to the fundamentals of data quality and require that the data produced during batch processing will be discrete, consistent, and semantically unambiguous." (Seeman & Goddard, 2015, p. 335)

Overall, the focus on describing resources needs to shift to describing a number of related elements, rather than bibliographic records.
1. Emphasize the use of authorized access points
2. Use identifiers whenever possible
3. Leverage the use of relationships in RDA and MARC wherever possible

As a result, the OCLS Cataloguing Workflow is driven by FRBR Group 1 elements (WEMI), and seeded with ample opportunity to build relationships with group 2 and 3 elements at every stage. By outlining the description of AAP, identifiers and relationships relevant to each respective WEMI element set, the cataloguer will begin to see resource description as a set of interrelated elements joined by relationships, rather than as a record. AIR elements are emphasized throughout the OCLS Cataloguing Workflow as outlined below:

```
024 7#  $ahttp://experiment.worldcat.org/entity/work/data/1840249565$2uri
050 #4  $aZ666.73.L56$b.H66 2014
100 1#  $aHooland, Seth van,$eauthor.$0http://id.loc.gov/authorities/names/no2014097600
245 10  $aLinked data for libraries, archives and museums :$bhow to clean, link and publish your
         metadata
```

While we are still working within MARC, in this example the following elements can be drawn out:

Work	"Linked data for libraries"
Work identifier	http://experiment.worldcat.org/entity/work/data/1840249565
Relationship designator	"author"
From linked data set - has identifier	http://id.loc.gov/vocabulary/relators/aut
Person	Hooland, Seth van
Identifier for person	http://id.loc.gov/authorities/names/no2014097600

This is starting to look like a triple!

REFERENCES

ALEMU, G., STEVENS, B., ROSS, P., & CHANDLER. (2012). LINKED DATA FOR LIBRARIES: BENEFITS OF A CONCEPTUAL SHIFT FROM LIBRARY-SPECIFIC RECORD STRUCTURES TO RDF-BASED DATA MODELS. NEW LIBRARY WORLD, 113 (11/12), 549. doi:10.1108/03074801211282920

AMERICAN LIBRARY ASSOCIATION. (2010). RDA TOOLKIT. CHICAGO, ILL: AMERICAN LIBRARY ASSOCIATION. RETRIEVED FROM HTTP://ACCESS.RDATOOLKIT.ORG/

BIGELOW, I., BOILEAU, S., EMON, D., LEUNG, R., MORGAN, M., & SILLIUS, I. (2015). OCLS CATALOGUING WORKFLOW. TORONTO, ON: OCLS. RETRIEVED FROM: https://www.ocls.ca/sites/default/files/attachments/OCLS_CAT_Workflow_23Oct_1.pdf

CLARKE, R. I. (2015). BREAKING RECORDS: THE HISTORY OF BIBLIOGRAPHIC RECORDS AND THEIR INFLUENCE IN CONCEPTUALIZING BIBLIOGRAPHIC DATA. CATALOGING & CLASSIFICATION QUARTERLY, 53(3/4), 286-302. DOI:10.1080/01639374.2014.960988

COYLE, K. (2013, JULY 23). LINKED DATA FIRST STEPS & CATCH-21 [BLOG POST]. RETRIEVED FROM: HTTP://KCOYLE.BLOGSPOT.CA/2013/07/LINKED-DATA-FIRST-STEPS-CATCH-21.HTML

GONZALES, B.M. (2014). LINKING LIBRARIES TO THE WEB: LINKED DATA AND THE FUTURE OF THE BIBLIOGRAPHIC RECORD. INFORMATION TECHNOLOGY & LIBRARIES, 33(4), 10-22. 13P. RETRIEVED FROM: http://ejournals.bc.edu/ojs/index.php/ital/article/download/5631/pdf

HILARIO, A.T., FERNANDEZ, T.R. AND CAMPO, D.M. (2014). FROM BIBLIOGRAPHIC RECORDS TO DATA: CHANGES IN THE LIBRARY ENVIRONMENT WITH THE APPLICATION OF LINKED OPEN DATA TECHNOLOGIES. INFORMATION RESOURCES MANAGEMENT JOURNAL, 27.3 (JULY-SEPTEMBER 2014).

JING, J., TRIPP, E., BAGLOW, L., BOILEAU, S., & EMON, D. (2015, SEPTEMBER). THE IMPACT OF LINKED DATA IN DIGITAL CURATION AND APPLICATION TO THE CATALOGUER'S WORKFLOW. PRESENTATION AT THE ACCESSYYZ CONFERENCE, TORONTO, CANADA.

MCCALLUM, S. (2015). BIBFRAME UPDATE. Library of Congress. https://www.youtube.com/watch?v=0-B9sUhGnKo

MILLER, E. (2014, DECEMBER). MOVING FROM MARC: HOW BIBFRAME MOVES THE DATA IN LIBRARIES CONVERSATION TO LARGE-SCALE ACTION. PRESENTATION AT THE SEMANTIC WEB IN LIBRARIES CONFERENCE, BONN, GERMANY. SLIDES RETRIEVED FROM HTTP://SWIB.ORG/SWIB14/SLIDES/MILLER_SWIB14_67.PDF

SEARLS, D. (2009). SILOS. FLICKR. RETRIEVED FROM: https://www.flickr.com/photos/docsearls/5500714140/

SEEMAN, D., & GODDARD, L. (MAY 15, 2015). PREPARING THE WAY: CREATING FUTURE COMPATIBLE CATALOGING DATA IN A TRANSITIONAL ENVIRONMENT. CATALOGING & CLASSIFICATION QUARTERLY, 53, 331-346

SCHMACHTENBERG, M., BIZER, C., JENTZSCH, A. & CYGANIAK, R. (2014). LINKING OPEN DATA CLOUD DIAGRAM. RETRIEVED FROM: http://lod-cloud.net/

STAHMER, C. (2015). LINKED DATA IN THE LIBRARY WORKFLOW ECOSYSTEM. https://www.youtube.com/watch?v=9-6MEyLLQshJY

SZETO, K. (2013). POSITIONING LIBRARY DATA FOR THE SEMANTIC WEB: RECENT DEVELOPMENTS IN RESOURCE DESCRIPTION. JOURNAL OF WEB LIBRARIANSHIP, 7(3), 305-321. DOI:10.1080/19322909.2013.826576