The Future is Now: Considering the Impact of RDA Post-Implementation and What’s Next

Saturday, April 27th, 10:30-12:00

RDA and the New BIBFRAME: Is Linked Data the Solution?
Philip Schreur
Stanford University

Slide 1 - Introduction

- take a look at where we've come from and why what we have now is incompatible with today’s needs
- then at a high level take a look at RDA and BIBFRAME and if they can resolve the issues
- we live in truly revolutionary times, periods of transition are never easy
- *We live in the best of times, we live in the worst of times*
- *it is the age of wisdom, it is the age of foolishness*
- *it is the epoch of belief, it is the epoch of incredulity*
- *it is the season of Light, it is the season of Darkness*
- *it is the spring of BIBFRAME, it is the winter of MARC*
- *we have everything before us, we have nothing before us*
- *we are all going direct to Heaven, we are all going straight to Hell*
- our transition from AACR2 to RDA, from MARC to linked data ... what more could be said?
Slide 2 & 3 – AACR2

- AACR first published 1967, almost 50 years ago
- Born in the time of the card catalog
- 2nd edition Michael Gorman in 1978
- Unifies British and American version, in line with ISBD
- AACR2 = “designed for the construction of catalogues and other lists in general libraries of all sizes”
- Remember transition to AACR2, music in 1981 same year as the release of the Howling, “Imagine your worst fear a reality”
Slide 4 – MARC

- In 1963 Council on Library Resources (CLIR) issued a grant on feasibility of automating the library systems at LC.

- After initial studies the first MARC pilot project was from 1966-1968 and by 1969 the MARC distribution service started.

- MARC extraordinary job on allowing us to communicate what AACR2 was designed to create ... catalogues for card catalogs and lists.

- Two major downsides in today's changing environment
  - Focus is on communicating strings of text collected into a single record
    - Strings of text used to communicate content as in a contents note or for linking and control as in controlled headings
    - Also, you need to see the record as a whole to understand the meaning of the text
  - It is used exclusively by the library community

- So why are these two points so crucial to our discussion? ... worthwhile taking a closer look at a typical MARC record.
Slide 5 – Bibliographic example

- typical MARC catalog record for a sound recording: impressive!

- It gives a description of the medium, the contents, the years of performance, controlled subject headings, analytical entries for all the individual musical works it contains, and displays the information in an easily digestible structure for the eye.

- it represents a recording of Fritz Kreisler performing a selection of violin music. The musical works are clearly articulated and responsibilities are clear from glancing at the record as a whole. But what would a machine make of this record?

- Much of the semantic meaning in this example can only be derived from the bibliographic record as a whole.
  o The human eye can easily see that the main entry is Fritz Kreisler and that he is a violinist,
  o piece by Joseph Sulzer is for violin and piano and
  o subject heading Violin and piano music,
  o from contents note Mozart Violin concerto is accompanied by the London Symphony Orchestra conducted by Sir Landon Ronald.

- This dependence on a complete bibliographic record for semantic meaning is a holdover from AACR2 the card catalog days.

- The MARC format allowed these records to be transformed into electronic documents and shared internationally, but they are still bibliographic records and to be understood must be evaluated as a whole.

- It’s often difficult for us to realize how much information our minds supply.
  o From author field we see Fritz Kreisler is listed as a creator
  
  o From the Participant Note we see that he is a violinist
  
  o From the Contributor fields we see the recording includes Efrem Zimbalist
  
  o From the Contents note we see that he is also a violinist
  
  o From the Contents Note we see that Kreisler performs a piece by Tchaikovsky (Chant sans paroles) that was originally for piano
  
  o From the Included Works Note we see that this piece is from Tchaikovsky’s work “Souvenir de Hapsal”
From the Subject Notes we see that the correct LCSH subject term for this work is Violin and piano music, Arranged.

There is nothing in the bibliographic record itself, though, that links these bits of information together. It is the human mind that makes these logical associations.

AACR was created to convey, static, self-contained lists of data encapsulated into a bibliographic record. The MARC format itself was created to clearly communicate the information encoded in this data, and in that it has been very successful. Although perpetuating the concept of the bibliographic record, it very clearly articulates and differentiates all the elements in the record. The MARC format is used, however, almost exclusively by the library community and much of its semantic meaning created through the use of AACR2 is lost to machine understanding. In the semantic world of linked data, these MARC records themselves are inarticulate. The shift to the Web as a primary source of information is unarguable. And as it is impossible for us to encompass the entirety of the Web in our library catalogs, our catalogs must move intelligently to the Web. Our millions of bibliographic records and the resources they represent are one of the truly great treasures we have to offer the web of data. The care with which we have created, maintained and enhanced them have made them a primary focus of the Semantic Web, but the way in which the data has been recorded in MARC prevents any intelligent, automated manipulation or linking.
Slide 6 – RDA

- so why has the development of RDA so crucial at this point?

- RDA developed as part of the JSC’s strategic plan (2005-2009)

- From 2004 to its publication in 2010, RDA’s editorial team headed by Tom Delsey worked on developing a model that was consistent, flexible, and extensible

- It is based upon two international, conceptual models: the Functional Requirements for Bibliographic Records (FRBR) and Functional Requirements for Authority Data (FRAD)

- What about it is important for this shift:
  - Explicit inclusion of description and access to digital objects
  - Equally as crucial, as opposed to a self contained card catalog, RDA designed for use in digital environments (Internet, Web OPACs), it speaks the language of the Web and is meant to be used with it
  - Designed to be flexible enough to be of use to a number of different communities not just libraries, the world of information and data is immense, we must not only learn to be a part of it but make use of it as well
  - Focus was not just on just creating a list of data points as AACR2, but also major emphasis on relationships between those data points, both to external authorities through the use of linked data and internally to make the record semantically understandable

- When testing was announced for July-December 2009, Stanford was one of the first volunteers – Why?
  - Developed intense interest in linked data and it’s ability to bring our data to the web and interconnect it – AACR2/MARC with its focus on records and static lists was not compatible – RDA and its emphasis on relationships and linked data elements was perfect match

- Our local implementation decisions focused on this linked data aspect:
  - trace all creators as essential linking nodes (no more rule of three)
  - relationship designators for all people/corporate bodies to clearly articulate their relationship to the work being described
  - emphasis on the controlled vocabularies at the Open Metadata Registry

- For test, we made the decision to have all of our original catalogers take part, not just one or two, totally committed
Once testing was complete, we decided to continue on with RDA, it was the future and we were confident that the broader decision to adopt RDA was forthcoming.

This shift to a new conceptual model of cataloging has been a difficult one, but one we are extremely pleased we made but I'd like to focus now on how RDA is communicated.
Slide 7 – BIBFRAME

- so what is BIBFRAME exactly and how does it fit into all this?

- BIBFRAME, short for the Bibliographic Framework Initiative, was officially launched in May of 2011

- It is designed as the replacement for MARC, that is, it is a communication format only

- BIBFRAME: “aims to re-envision and implement a new bibliographic environment for libraries that makes the network central and interconnectedness commonplace”

- Designed to integrate with and engage in the wider information community while also serving the very specific needs of its maintenance community, that is, US

- So we see from the start, BIBFRAME is designed to overcome the two major flaws mentioned earlier, incompatibility with the web and exclusive use by the library community

- The BIBFRAME model is divided into Works, Instances, Authorities, and Annotations.

- It has two main divisions: work and instance
  - Work = the conceptual cataloging item, it is an abstract entity, in the case of our sound recording, it would be both the sound recording as a whole, and also associated works for the separate pieces performed like the Mozart violin concerto

  - Instance = a material embodiment of a BIBFRAME work, each BIBFRAME instance is an instance of one and only one BIBFRAME work. In other words, this is parallel to the FRBR manifestation

  - Authorities are key concepts such as people, places, topics or organizations and may be associated with a work or an instance, in the slide we see that concepts such as creator or subject can be associated with the work, while publisher or format can be associated with the instance
Annotation is basically everything else attached to the work or instance, from the slide we see the fact that an institution holds a particular instance is an annotation to that instance, or, the review of a particular work can be associated to that work through an annotation at the work level.

So why is BIBFRAME such an essential part of our RDA revolution?

1st, The focus is on the web, we’ve seen earlier that our records expressed in MARC are unintelligible to the web, however, it’s where our patrons are going and we must be there. BIBFRAME’s expression as linked data speaks the language of the Web and integrates with it.

2nd, It is designed to be able to be used by many communities, not just the library world, it will incorporate us into the much wider world of information exchange. It will embed our rich data into the web and allow us to make use of the wealth of data others have created.

3rd, It is expressed through linked data so all those relationships at the heart of RDA can be expressed meaningfully.

4th, through the use of BIBFRAME’s annotations we can associate all those concepts such as subjects and performer, to the correct work in the record. No more dependence on your eyes and the entire record to make sense of the information.

and 5th, because BIBFRAME is based in linked data, we can make use of all the controlled vocabularies at id.loc.gov and the Open Metadata Registry to unambiguously link concepts to works or instances, finally breaking free from dependence on accurate text strings in an authority file for correct linking.

So with all these unarguable benefits, why is BIBFRAME still so controversial?

1st, BIBFRAME is still in a very early stage of development and there are some key areas to be developed:

BIBFRAME is a communications format meant to be used by many communities, so it cannot be tailored to RDA and FRBR alone, FRBR has 4 levels, WEMI, but BIBFRAME may be called upon to communicate conceptual models that have 2 levels or 16 levels.
All must be accommodated so is frustrating to the library community that sees a disconnect between BIBFRAME and RDA/FRBR

- BIBFRAME is meant for communications only! Not like MARC which has confused communication and storage and is also the basis of our ILS. Perhaps our ILS may still use MARC for a while for internal functioning, but then data must be able to be converted back and forth between the two.

- Definition of a work in the BIBFRAME sense is still unclear at times, is a new performance an instance or a work? And if they are instances how do you go about relating different instances?

- As we've seen earlier, our MARC data is very unclear to a machine. That means there cannot be a foolproof, complete conversion of data from MARC to BIBFRAME, all those correct associations we'd like to make between subjects or performers in a resource that contains multiple works will have to be done by humans over time.
Slide 9 – Conclusion

- As our RDA revolution, shifts to a less bloody stage, we can once again return to an era of enlightened discourse

- As we’ve seen, AACR2 and the MARC format are clearly not designed for today’s world, its limited use to the library community and lack of compatibility with the web make it a dangerous dead-end

- BIBFRAME, however, with its roots in linked data and emphasis on relationships and compatibility with the broader, worldwide exchange of data is an ideal medium to communicate RDA between ourselves and others

- But BIBFRAME is at an early stage of development and can still be swayed in many different directions

- it is essential for everyone from all communities to become involved with the electronic salon of BIBFRAME discourse

- the time for input is now before the new model is formalized and concerns from our many communities become difficult to incorporate