

Linked Data and RDA: Considerations for PCC Policy Committee

Prepared by PCC Linked Data Advisory Committee
July - August, 2018

Introduction

The Program for Cooperative Cataloging has a strong history of interpreting and implementing RDA; PCC's measures of quality and attempts at interoperability are tied closely to the adoption of RDA. We see this in PCC's RDA policy statements and corresponding practices. Much of the work anticipated over the next year (likely longer) within the PCC will be in reaction to changes occurring in RDA, along with the imminent move to linked data for early adopters within the PCC. In order to respond consistently to RDA it is important to articulate the position of RDA within PCC practice and acknowledge certain realities. Because of the need for conversion from MARC to linked data during this transition period, MARC and linked data practices should strive to be consistent as they relate to RDA. We also know from a pragmatic view, that libraries are (and will be more) gluing data from wildly different data models. Instead of saying we do away with RDA, in a linked data environment we know that there will not be total agreement on models and need to position our decisions accordingly.

For the metadata produced by the PCC, we will need to know what level of commitment to RDA we aspire to. Ideally PoCo would articulate this by exploring the following questions and stating RDA's role in the work that we do. With this clarified, the Standing Committee on Standards [more grounded questions](#) should be addressed.

Framing Questions

- ***Does there need to be a clear tie-in with RDA to every decision/workflow made by the PCC?***
 - For example: with the ISNI Pilot work, is there a requirement to figure out, record, justify practice through an RDA lense/recording method? Or should we figure out what we want/need to do before concerning ourselves with RDA?
 - Possible answer: Most developments should not be seen as "or" - that is, accepting a non-RDA practice does not replace the RDA practice - they can exist side-by-side if both are deemed useful. The point is to make or find points of linkage so that two parallel practices are connected and, depending on needs, systems can branch into one or the other.
- ***Does the same decision need to be made in each different workflow?***

- We know this isn't happening now, e.g. Digital Collections; limitations on implementing RDA in MARC
- Can we support the floor and still encourage more "nice to haves"?
- Possible answer: You may want to look at the way that BIBFRA.ME is handling this, as well as the work-flows that are allowed in the RDA Toolkit. Those illustrate a way to provide views or profiles that address different needs while working within a single vocabulary. Another option is to embrace "application profiles" and a "mix-n-match" environment. In that option, you can have a core vocabulary that can be extended for different needs using elements from other vocabularies. (We can provide examples here as links.) Mixing elements from different namespaces is not new; this was introduced in the XML environment. Profiles are a way to create a new set of elements by combining the best from existing metadata vocabularies.
- ***What if there's something we justifiably disagree with from the RDA model?***
 - For example: Fictional Agents
 - Possible answer: If you wish to share with RDA-adherent partners, you may need to create data that obeys RDA to the letter. That doesn't prevent you from ALSO making use of elements from other vocabularies that better meet your needs. The complication with this particular example is that "person" is being defined differently in different vocabularies, and that makes it hard to create an equivalency between statements across vocabularies. An RDA person may not be the same as an ISNI person because of the differences in definition. There are ways to indicate that elements from different vocabularies are used in analogous ways even though they are not identical.
- ***What if there's something not addressed in the RDA model?***
 - E.g. Events, bf:Contributions, deep treatment of Custodial History?
 - Possible answer: If you find this information useful, when library data transitions to LoD it should be possible to use any elements of other vocabularies. These will become part of the graph you are creating, but they can be ignored by anyone not interested in that data. Because there are no records, just graphs, the default for applications is to select what you need and ignore the rest. Extra links shouldn't be a problem. If, on the other hand, there are many people needing to express this information, and those folks are using RDA, it might be the case to add these elements to RDA. Ideally, RDA would not duplicate these elements but would use what exists, and that could be seen as an authoritative extension to RDA.
- ***Is our goal to implement RDA/RDF? Or... Is thinking of RDA as a loose content standard enough of a goal?***
 - Possible answer: We tend to see the future as deviating (for better or worse) from the ideal of a single set of cataloging rules and a single bib record standard. The fact is that we already work in an environment with multiple rules, multiple practices, and more than one strict record type. OCLC processes hundreds of different sources that are not identical. One developer, Kiraly, has a dozen

routines just to deal with "MARC" records. Denying these differences makes the work of sharing harder. We could see the future as one in which we acknowledge and openly work with allowing differences but mitigating their impact by actively seeking to create links. Thus (finally getting to the point) RDA in RDF is one option should it ever come to fruition, but there is no reason to privilege that solution if others work for us.

Heterogeneous Workflows Through Linking

This is an example of two sets of data, ISNI and BIBFRAME 1.0, and how each can use its own data elements and workflow, and at the same time they can link together. When data is on the open web, linking is simply a matter of using the same URI. This is the case because each identifier exists only once in any given data space, and the web is a single data space.

When applications make use of data in graph form they select the portions of the graph they are designed to work with. A BIBFRAME or RDA application will select elements known to it. Applications can ignore any statements, such as `rdfs:seeAlso`, that they are not equipped to process. Other applications that are designed to make use of those statements can select and make use of any of the portions of the graph that the `rdfs:seeAlso` opens up for processing.

