What We Learned by Building the RDA Vocabularies ...

... and what we still need to figure out
What We Already Knew

• Traditional library practices around authorities
• Thesaurus development and standards
• Semantic Web standards (mostly W3C)
• The basic conflicts between world views that needed to be accommodated:
  • Library culture vs. SemWeb culture
  • XML vs. RDF
Model of ‘the World’ /XML

• XML assumes a 'closed' world (domain), usually defined by a schema:
  • "We know all of the data describing this resource. The single description must be a valid document according to our schema. The data must be valid."

• XML's document model provides a neat equivalence to a metadata 'record'
Model of ‘the World’ /RDF

• RDF assumes an 'open' world:
  • "There's an infinite amount of unknown data describing this resource yet to be discovered. It will come from an infinite number of providers. There will be an infinite number of descriptions. Those descriptions must be consistent."

• RDF's statement-oriented data model has no notion of 'record’ (rather, statements can be aggregated for a fuller description of a resource)
Most Important Goals (Social)

• Build the vocabularies as the JSC required
• Prepare for the future we saw coming by using forward-looking Web standards
• Build vocabularies that could also be used outside libraries as well as by libraries
• Consider issues of long-term transition from MARC 21 to RDA

• Some obvious conflicts
Most Important Goals (Technical)

• Bridge the XML and RDF worlds
• Ensure ability to map between RDA and other element sets
• Provide a sound platform for extension
• Consider methods for expressing AACR2 structures in technical ways to ease the pain of transition to RDA
The General Strategy

• The Semantic Web was our ‘mental model’
  • Made decisions that supported the creation of a “bridge” between XML and RDF
  • Used the FRBR entities as classes in a ‘FRBR in RDA’ vocabulary
• Ensured that the requirements of the library community for a FRBR-based metadata standard were met
*Decision Point*

• Once IFLA’s official FRBR was registered we faced the issue of whether the ‘FRBR in RDA’ should be replaced by the official version
  • Pros: supported the idea of ‘one FRBR’; pushed the maintenance of FRBR outside RDA
  • Cons: extant differences would need to be negotiated; dependency on IFLA for maintenance; tied to IFLA FRBR’s semantics, even if they changed in incompatible ways
  • DECISION: Keep them separate and map them to each other
Structural Rationale

- Property and value vocabularies were published via the Open Metadata Registry in a new namespace: [http://rdvocab.info](http://rdvocab.info)
  - Used RDF Schema (RDFS), Simple Knowledge Organisation System (SKOS) and Web Ontology Language (OWL)
  - Decisions oriented to favor approaches that can be generalized to make other vocabulary based standards web-friendly, available for use in applications, and easily updated by communities
The Structure, Simplified

- RDA Properties declared in two separate hierarchies:
  - an ‘unconstrained’ vocabulary, with no explicit relationship to FRBR entities
  - A subset of classes, properties and subproperties with FRBR entities as ‘domains’
- Pros: retained usability in or out of libraries; better mapping to/from non-FRBR vocabularies
- Cons: Didn’t please JSC; introduced new issues of responsibility and maintenance

Apr. 27, 2012
Property (Generalized, no FRBR relationship) → Subproperty (with relationship to one FRBR entity) → FRBR Entity → Semantic Web → Library Applications

The Simple Case: One Property--One FRBR Entity
Property (Generalized, no FRBR relationship) → Subproperty (with relationship to one FRBR entity) → Subproperty (with relationship to one FRBR entity) → Semantic Web → Library Applications

The Not-So-Simple Case: One Property—more than One FRBR Entity
Why Unconstrained Properties?

• We think that the ‘bounded’ properties should be seen as the official JSC-defined RDA Application Profile for libraries
• Extensions and mapping can be built more usefully from the unconstrained properties
  • Unconstrained vocabularies more acceptable in domains where FRBR not assumed or inappropriate
  • Mapping from vocabularies not using the FRBR model directly to ones that do (and back) not necessarily a good idea
*Decision Point*

- Preliminary decision to separate the unconstrained and bounded vocabularies
  - Relationships can remain as is
  - New namespace required for the unconstrained vocabularies (which requires updating of relationship URI links after a move)
  - Discussions on synchronization and formalizing maintenance responsibilities still ongoing
Roles: Attributes or Properties?

• In 2005, the DC Usage Board worked with LC to build a formal representation of the MARC Relators so that these terms could be used with DC
  • This work provided a template for the registration of the role terms in RDA (in Appendix I) and, by extension, the other RDA relationships
• Role and relationship properties are registered at the same level as elements, rather than as attributes (as MARC does with relators, and RDA does in its XML schemas)
*Decision Point*

- Roles and relationships as properties should be usable in XML as well as RDF
- If separate attribute vocabularies for roles and vocabularies is really desired, they can be added to the set
Aggregated Statements

• RDA (and MARC 21) include aggregated statements that have traditionally been handled as a group, ex.: Publication, Distribution, Manufacture and Production statements, etc.

• Some users will prefer to deal with the issue at the display level, by using separate properties and configuring the display to manipulate the order

• To deal with retrospective, transitional and prospective needs, we must address the desire to continue this practice
*Decision Point*

- We viewed those aggregations as ‘Syntax Encoding Schemes’ (as defined in the DCAM), and accommodated them within the bounded properties only
  - An SES covers both inclusion and order, so that the data can be validated properly
  - Right now the inclusion is covered in the RDA Vocabularies, but the order exists only in a textual note
  - Technically it requires a defined datatype with an associated XML schema for validation to be possible
Pre-coordinated Statements: Structure

Aggregated Statement (no domain or range)

Aggregated Statement Subproperty

Range: RDA Syntax Encoding Scheme (Subclass of RDF Datatype)

Range: [Specific] Encoding Scheme (Subclass)

General Property (no domain or range)

Subproperty

Domain: FRBR Entity
Pre-coordinated Statements: Example

Publication Statement (no domain or range)

Publication Statement (Manifestation)

Range: RDA Syntax Encoding Scheme (Subclass of RDF Datatype)

Range: Publication Statement Encoding Scheme (Subclass)

Place of publication (no domain or range)

Place of publication (Manifestation)
Application Profiles

• The RDA ‘bounded’ properties are most of what is needed for an RDA AP
  • What’s still lacking is the addition of the necessary constraints: datatypes, cardinality, associated value vocabularies
  • These can be exposed for human readable documentation, but not yet for machine-readable implementation
    • Still lacking an implementation neutral constraint language
*Decision Point*

- Current review of the Dublin Core Abstract Model should help move the process along
- OMR building capability to document, maintain and share APs
- Specialist communities will be able to develop APs that use RDA properties as well as specialized extensions
Vocabulary Extension

• The inclusion of unconstrained properties provides a path for extension of RDA into specialized library communities and non-library communities
  • They may have a different notion of how FRBR ‘aggregates’ (For example, a colorized version of a film may be viewed as a separate work)
  • They may not wish to use FRBR at all
  • They may have additional properties to use, that could have a relationship to the RDA properties
RDA:adaptedAs

hasSubproperty

RDA:adaptedAsARadioScript
Extension using Generalized Properties
Extension using Generalized Properties
Mapping

- Building relationships between vocabularies is not necessarily a proprietary activity or tied to particular applications (not ‘crosswalking’ as we have understood that term)
- [http://MARC21rdf.info](http://MARC21rdf.info) provides a beginning point for mapping library data that is easily shared and modified to meet a variety of needs
- New approaches go beyond ‘sameAs’ and suggest the potential for more granular, nuanced relationships
Links and Contact Info

RDA Vocabularies: http://rdvocab.info
MARC 21 in RDF: http://MARC21rdf.info

Diane: metadata.maven@gmail.com


A special Thank You to the British Library and Siderian Software for supporting this work!