

To: Joint Steering Committee for Development of RDA

From: Gordon Dunsire, CILIP Representative

Subject: Machine-actionability and interoperability of RDA value vocabularies: a discussion paper

Introduction

This paper discusses a number of ways of improving the machine-actionability of the RDA value vocabularies through their representation in RDF in the RDA namespace. Processing of these vocabularies by machine leads to improved interoperability of RDA data within RDA and with external data.

A value vocabulary “represents a controlled list of allowed values for an element”.¹ This includes the sets of terms explicitly given in the RDA toolkit and recommended for use as the content of RDA elements.

Each set of terms is currently represented in RDF as a separate value vocabulary in the Open Metadata Registry (OMR). The value vocabularies are listed in the OMR.²

The RDF representations use properties from SKOS (Simple Knowledge Organization System) :

skos:prefLabel	Preferred value.
skos:altLabel	Alternative value; equivalent to a <i>see from</i> reference.
skos:definition	Definition of the term.
skos:scopeNote	Scope of the term.

Separation of definition and scope note

Scope information includes examples and notes on what is specifically included within or excluded from the definition. Such information can be identified by phrases such as “for example”, “includes”, “including”, “excludes”, etc.

Discussion: It is good practice to avoid mixing scope information with the definition, and vice-versa. This improves the clarity of the definition and scope note, allows the scope note to be amended by the addition or removal of examples without changing the definition, and reduces the size of the definition and/or scope note.

Existing RDA definitions and scope notes generally follow good practice. In the RDA Toolkit glossary, the definition precedes the scope note. However, some existing definitions and scope notes can be improved further.

¹ W3C Library Linked Data Incubator Group report, available at: <http://www.w3.org/2005/Incubator/llid/XGR-llid-vocabdataset-20111025/>

² Listed as “RDA vocabularies” at: <http://rdvocab.info/>

For example, “tactile notated music” has the definition “Content expressed through a form of musical notation intended to be perceived through touch” and scope note “Includes braille music and other tactile forms of musical notation”. Here, the phrase “and other tactile forms of musical notation” in the scope note is redundant, as it is covered by the definition and preferred term.

Recommendation: Clearly separate the scope note from the definition and avoid redundancy for new and existing value vocabularies.

Dependencies between value vocabularies

RDA/ONIX Framework

The RDA Carrier Type, RDA Content Type, and RDA Media Type vocabularies are based on the RDA/ONIX Framework for resource categorization³. The RDA Toolkit states “RDA also conforms to the RDA/ONIX Framework for Resource Categorization”, but this is the sole reference to the Framework. The Toolkit does not explicitly state the dependency of any of the value vocabularies to the Framework.

There does not appear to be a follow-up to 6JSC/ALA Rep/1 Revision to: Categorization of content and carrier⁴, and the proposed 6JSC/RDA/Section 1/Categorization/Rev has not yet been published. This means that there is very little documentation about the current relationship between RDA value vocabularies and the Framework, which inhibits public understanding of the derivation of the vocabularies and their intended utility.

Recommendation: Prioritize the publication of up-to-date documentation on the relationship between RDA value vocabularies and the RDA/ONIX Framework.

The Framework is intended to improve interoperability between content and carrier categories as well as provide a clear distinction between content and carrier. Use of the Framework for interoperability of ISBD and RDA content and carrier value vocabularies is at the core of the proposed methodology in 6JSC/Chair/5 Mapping ISBD Area 0 vocabularies to RDA carrier and content⁵. The mapping itself is now available⁶, but it cannot be made machine-actionable until the Framework elements have been represented in RDF. Note that there is no complete one-to-one mapping between the ISBD and RDA value vocabularies, so use of the Framework as a mapping hub is essential. The same is likely to be true for resource categories in other vocabularies, such as schema.org, and the lack of an RDF representation of the Framework will impede RDA interoperability in general.

Recommendation: Develop and publish, as a priority, an RDF representation of the RDA/ONIX Framework, and publish an RDF representation of the mappings between the RDA value vocabularies and the Framework.

³ <http://www.rda-jsc.org/docs/5chair10.pdf>

⁴ <http://www.rda-jsc.org/docs/6alarep1.pdf>

⁵ <http://www.rda-jsc.org/docs/6JSC-Chair-5.pdf>

⁶ <http://www.rda-jsc.org/docs/6JSC-ISBD-Discussion-2-Mapping.pdf>

Other internal dependencies

A preliminary analysis shows that other RDA value vocabularies can be associated with the Framework:

RDA value vocabulary	Framework element
Applied Material	Carrier: AppliedMaterial
Base Material	Carrier: BaseMaterial
Base Material for Microfilm, Microfiche, Photographic Film, and Motion Picture Film	Carrier: BaseMaterial
Broadcast Standard	Carrier: EncodingFormat
Digital Representation of Cartographic Content	Carrier: EncodingFormat
Emulsion on Microfilm and Microfiche	Carrier: AppliedMaterial
Encoding Format	Carrier: EncodingFormat
File Type	Carrier: EncodingFormat
Generation for Microform	Carrier: Generation
Generation for Motion Picture	Carrier: Generation
Generation for Videotape	Carrier: Generation
Generation of Audio Recording	Carrier: Generation
Generation of Digital Resource	Carrier: Generation
Groove Pitch	Carrier: FixationMethod?
Groove Width	Carrier: FixationMethod?
Mode of Issuance	Content:ExtensionTermination
Polarity	Carrier:EncodingFormat?
Presentation Format	Carrier:EncodingFormat?
Production Format	Carrier:FixationMethod
Production Method for Manuscripts	Carrier:FixationMethod
Production Method for Tactile Resource	Carrier:FixationMethod
Recording Medium	Carrier:FixationMethod
Reduction Ratio	Carrier:EncodingFormat?
Special Playback Characteristics	Carrier:EncodingFormat
Track Configuration	Carrier:EncodingFormat?
Type of Recording	Carrier:EncodingFormat
Video Format	Carrier:EncodingFormat

Further analysis of these associations may reveal intrinsic relationships between the value vocabularies. These include hierarchical (broader, narrower) and associative (see, see also) relationships which can be represented in RDF using SKOS.

This can be supported by matching terms; for example, the word “audio” appears in multiple RDA value vocabularies (Carrier Type, Encoding Format, Extent, File Type, Media Type), indicating some kind of semantic relationship between the terms and vocabularies.

Recommendation: Carry out further analysis of matches between RDA value vocabularies, terms and the RDA/ONIX Framework to identify semantic relationships for further improvement of interoperability between vocabulary terminology and semantics.

Extension of the Framework

A couple of extensions to the RDA/ONIX Framework (ROF) were proposed in 6JSC/ALA Rep/1. It is possible that the analysis recommended above will suggest further extensions. 6JSC/Chair/5 identifies further issues to be discussed, including:

- Mapping of ROF properties and classes to other namespaces, including RDA and ISBD.
- Use of qualified resource categories.
- Development of one or more DC Application Profiles or OWL ontologies for the ROF, specifying the base property sets.
- Creation of a management and development group for the ROF.

This should lead to improved machine-actionability and interoperability of data for RDA users.

The first three issues can be included in the analysis recommended above.

Recommendation: Create a management and development group for the RDA/ONIX Framework.

Recommendations

1. Clearly separate the scope note from the definition and avoid redundancy for new and existing value vocabularies.
2. Prioritize the publication of up-to-date documentation on the relationship between RDA value vocabularies and the RDA/ONIX Framework.
3. Develop and publish, as a priority, an RDF representation of the RDA/ONIX Framework, and publish an RDF representation of the mappings between the RDA value vocabularies and the Framework.
4. Carry out further analysis of matches between RDA value vocabularies, terms and the RDA/ONIX Framework to identify semantic relationships for further improvement of interoperability between vocabulary terminology and semantics.
5. Create a management and development group for the RDA/ONIX Framework.

Gordon Dunsire
1 October 2012