Functional interoperability and choice in RDA

Gordon Dunsire

23. seminar Arhivi, Knjižnice, Muzeji
27-29 Nov 2019, Poreć, Croatia

Published by Gordon Dunsire

Edinburgh 2021

Functional interoperability and choice in RDA

Gordon Dunsire

Recent development of RDA

RDA: Resource Description and Access is the successor to the Anglo-American Cataloguing Rules. RDA was first published in 2010 as an online resource: the RDA Toolkit [1]. The publishers of RDA and the RDA Steering Committee (RSC) who are responsible for developing the content decided in 2016 to initiate the RDA Toolkit Restructure and Redesign (3R) Project to completely review the structure and presentation of the Toolkit to meet several goals.

These goals included updating the design of the Toolkit to reflect modern approaches to online publishing and the user experience. In particular, this involved improving support for users of mobile devices and for users with sensory preferences, such as the use of screen-reader software.

The content management infrastructure for RDA also required upgrading to convert the XML mark-up used for display and functionality to the more appropriate Darwin Information Typing Architecture (DITA) standard for content authoring and publishing, and to increase the re-use of content within the Toolkit. The labels, definitions, and scope notes of RDA entities, elements, and controlled terms are separately maintained and stored in linked data format as RDA Reference, and imported to the Toolkit as required. An entry in RDA Reference is typically used in two or more different parts of the Toolkit.

At the same time, the IFLA Library Reference Model (LRM) was in the final stages of development [2]. The LRM is a consolidation of the IFLA Functional Requirements family of conceptual models for bibliographic data on which RDA is based. Consolidation into a single model was necessary to resolve differences and fill in gaps that had emerged during the decade or more of the development of the 'family' component models. The RDA Toolkit had suspended its own evolution in these areas while waiting for the IFLA work to be completed, so there was some urgency in updating RDA to ensure conformance with the LRM.

The RDA Board who oversee the development and production of RDA introduced a strategy to widen the utility of RDA to cover resources held in cultural collections beyond the library sector, and to support international and linked data communities. This also had a significant influence on the direction of the project.

The 3R Project is expected to be completed at the end of 2020. A beta version of the new Toolkit was published in 2018 to elicit feedback and to give Toolkit users the opportunity to familiarize themselves with the changes to the content and presentation of the standard.

Communities and choice

The introduction of more choice in RDA instructions was initiated in 2011 when the Deutsche Nationalbibliothek (DNB) joined the RDA governance infrastructure. The DNB was the first non-Anglophone member of the Committee of Principals (CoP) for RDA (later the RDA Board) and the Joint Steering Committee for Development of RDA (JSC) (later the RSC). Differences with traditional Anglo-American cataloguing practices were accommodated by adding alternatives and exceptions to

RDA instructions. Other national libraries and library associations indicated a desire to adopt RDA if they could participate in its management and development processes.

In 2015 the CoP announced its intention to move to a more international governance structure, and to develop RDA to meet the needs of a wider community. In particular, the new structure was designed for global interaction and expansion of RDA's coverage three strategic markets:

- International communities
- Cultural heritage communities
- Linked data communities

The CoP became the RDA Board, and the JSC was renamed as the RDA Steering Committee. The RSC realized that continuing to treat departures from the Anglo-American tradition as alternatives was neither sustainable nor desirable in an international standard. Instead, the 3R Project was an opportunity to accommodate differing practices in a more equitable way as options, with the result that by the end of 2019, most of the cataloguing tools in RDA are optional.

RDA provides a hierarchical arrangement of entities for the agents who have a responsibility for creating the works, expressions, and manifestations that comprise a information resource. The LRM restricts such agents to human beings, and offers an Agent entity with two subtypes: Person is a single human being; Collective Agent is two or more persons who act together. The existing Corporate Body, Family, and Person entities in RDA are readily accommodated in the LRM structure, with Corporate Body and Family as subtypes of Collective Agent:

> Agent
>> Collective Agent
>>> Corporate Body
>>> Family
>> Person

An application may choose to collapse the hierarchy at any level below the top: Collective Agent instead of distinguishing between Corporate Body and Family; Agent instead of distinguishing between Collective Agent and Person.

RDA provides a range of elements that are used to describe characteristics of each RDA entity and its relationship to other entities. An element is categorized as a "relationship" if its value refers to another entity, and as an "attribute" if its value refers to the entity being described. The distinction is fuzzy and depends on which entities are included in a local application. For example, if Nomen is included as an entity then an "appellation" element such as a title is a relationship; if appellations are not treated as nomen entities than the relationship is an attribute.

The first part of the alphabetical list of Item elements illustrates this fluid categorization:

- access point for item
- accompanied by item
- annotator
- appellation of item
- authorized access point for item
- autographer
- …

The ‘access point’ and ‘appellation’ elements are used to record labels for an item being described, while the other elements record labels for item and agent entities that are related to an item being described. The Nomen entity allows labels to be described in their own right as a distinct entity, and the ‘appellation’ element can be treated either as an attribute of the item being described, or as a relationship with a specific nomen.

The only element in RDA that is mandatory for all applications is “nomen string’. A description of a nomen must include the string that is the nomen itself. All other RDA elements are assigned mandatory or optional status through the specification of a local application.

RDA specifies a minimum description for each entity. The description must include an appellation of the entity, chosen from a name/title, access point, or identifier element, so that there is at least one referent label for the entity. In addition, an RDA “resource entity” (Work, Expression, Manifestation, or Item) must also include elements that relate it to the other resource entities that comprise the “WEMI stack” for an information resource.

RDA elements may have hierarchical relationships between them: an element may be broader or narrower, or both, to one or two other elements.

For example, the element “adaptation of work” is part of an element hierarchy:

Work: based on work

- Work: adaptation of work
  - Work: dramatization of work
  - Work: graphic novelization of work
  - Work: motion picture adaptation of work
  - …

The value of a narrower element is valid for all elements that are broader in the hierarchy. An application that uses RDA metadata has a choice of granularity for recording a data value, and can easily process a fine-grained value to assign it to a broader element if required.

RDA Toolkit allows four different ways of recording the value of an element:

1. Unstructured description
2. Structured description
3. Identifier
4. IRI (Internationalized Resource Identifier)

An unstructured description is “a human-readable string that is an uncontrolled full or partial description of an entity or an uncontrolled term describing an aspect of an entity”. A note, transcription, and name or title as it appears in a source of information are examples of unstructured descriptions.

A structured description is ”a human-readable string that is a full or partial description of an entity that is based on a string encoding scheme, or is a controlled term that describes an aspect of an entity”. An access point and a term taken from a vocabulary encoding scheme are examples of structured description.

An identifier is “a machine-readable string that is assigned to an entity in order to differentiate the entity from other entities within a local domain, or a notation for a term from a controlled

vocabulary that is assigned to an aspect of the entity". An identifier or notation taken from a vocabulary encoding scheme are examples of identifier values.

An IRI is "a machine-readable string that is assigned to an entity in order to differentiate the entity from other entities, or to an aspect of an entity, within the global domain of the semantic web and open linked data".

Most of the RDA elements can use any of the recording methods, although there are restrictions for some elements. For example, an access point for an entity is always recorded as a structured description or as an IRI (if it is treated as a nomen), but not as an unstructured description or identifier. The recording methods are approximately aligned with four RDA data implementation scenarios:

- Scenario A: Linked open data
- Scenario B: Relational or object-oriented data
- Scenario C: Bibliographic/authority data
- Scenario D: Flat file data

The implementation of linked open data requires the IRI recording method. The implementation of a relational database uses local identifiers to relate tables. Bibliographic and authority data implementation architectures usually relate authority and bibliographic descriptions with structured and controlled access points, while flat-file data architectures do not require controlled or structured links. Operational RDA applications are likely to use a hybrid approach that mixes two or more scenarios.

Nearly all of the specific instructions in RDA Toolkit are optional. A single option consists of instructions for a single course of action. A set of options may be mutually exclusive, so that only one option can be chosen, or allow for more than one option to be applied in the same description; this is indicated by the content and context of the instructions.

For example, the instructions for an access point for any RDA entity include options for selecting the value from a vocabulary encoding scheme (VES) or constructing the value using a string encoding scheme (SES). A well-formed VES uses an SES to construct an access point, so the two options cover the selection and addition of entries in an authority file.

Choice vs Consistency

Local values of an element may vary for several reasons. There may be variation in values found in different sources of information; there may be differences because of language, naming conventions, and other cultural aspects; the values recorded may reflect policies associated with the application, such as its intended audience. This is a strong indication that a one-size-fits-all approach to cultural metadata cannot be as effective as accommodating a range of choices. In particular, a top-down imposition of a "universal" method for ensuring the consistency of metadata values across a range of applications is not suitable for an international content standard.

RDA assumes that the values recorded for different applications will not be the same. Instead, RDA relies on the semantic consistency of its entities and elements to ensure that RDA guidance and instructions can be applied across the range of implementation scenarios, recording methods, and local requirements. This guarantees a specific level of interoperability of RDA data from multiple sources and re-use of RDA data between applications. For example, the different values "Nineteen-sixty" and "1960" refer to the same timespan; the first value is an unstructured description that is

transcribed from a manifestation, and the second is a structured description based on the ISO 8601 standard for representing dates and times. These timespan values may appear at any level of granularity of a related timespan element.

The semantic consistency of RDA entities and elements is built on conformance with the LRM. The LRM "is a high-level conceptual model and as such is intended as a guide or basis on which to formulate cataloguing rules and implement bibliographic systems … A wide range of decisions made in cataloguing rules can be accommodated … ".

All LRM entities have an equivalent RDA entity, and all RDA elements are assigned to a specific entity. All RDA relationship elements are also assigned a "target" entity. All RDA elements can be mapped as a subtype of a high-level LRM element; all RDA elements are distinct with clear definitions. The new RDA guidance and instructions cover data provenance, such as the agency who creates the metadata and the controlled terminologies used as a source of element values, to support the processing of RDA metadata from multiple agencies.

As a result, RDA supports the use of local strings (element values) to describe global things (entities).

Application profiles

Most applications that use RDA will place restrictions on the choices that should be made, to ensure that only the entities, elements, recording methods, and controlled values required for the application are used.

This is essential for the effective and efficient use of RDA Toolkit; a single practical application of RDA does not need to use the whole Toolkit, and the sheer number of options can be overwhelming for someone who is not familiar with all aspects of modern resource description and access.

There are several methods for specifying which optional components of RDA Toolkit should be used by a specific application:

- Application profile
- Policy statements
- Workflow and other procedural documentation
- Toolkit annotation

The RDA glossary defines an application profile simply as "A specification of the metadata that is used in an application". It is the most structured, and therefore machine-readable, of the methods.

The concept of application profiles was originally developed by Dublin Core Metadata Initiative [3]. A Dublin Core profile can mix entities and elements from multiple standards, but an "RDA application profile" is confined to the components of RDA Toolkit: only RDA entities, elements, etc. are specified. The RDA entities and elements are comprehensive, and it is usually not necessary to incorporate non-RDA components in an application.

An application profile typically specifies:

- The elements to be recorded as a metadata description set for an entity
- The mandatory and repeatability status of each element
- The vocabulary encoding scheme to be used as a source of data for an element
- The string encoding scheme to be used to assemble the data for an element

An application profile can be used to support several functions:

- Removing unnecessary choice and reducing the time taken by a cataloguer to decide what options to use.
- The basis of an interface for data input and maintenance. Only the profile entities and elements are displayed in a cataloguer's screen.
- Data validation for metadata conformance and quality control. Recorded data is matched against the specifications of the profile.
- Data extraction from external metadata sources. Extracted data is matched against the profile.

As well as restricting the profile to RDA components, an RDA application profile can specify additional parameters that are unique to RDA:

- Links to RDA instructions at multiple levels using URLs.
- The recording method to be used for an element, where a choice is available
- The optional instruction to be applied to an element
- The policy statement to be applied to an element

The most common way of presenting an application profile is as a comma-separated variable spreadsheet. Parameters such as repeatability, display label, and source of values are recorded as columns; the parameters for specific element are recorded as rows. An application profile can be published anywhere because the RDA components are accessed via URLs.

Table 1 shows a simple RDA application profile for the Person entity.

Element	**Display label**	**Mandatory?**	**Source of values**
identifier for person	VIAF identifier	No	VIAF
name of person	Name	Yes	LCNAF
date of birth	Born	Yes	ISO 8601
date of death	Died	No	ISO 8601

Table 1: Application profile for person

The underline indicates a hyperlink to the RDA element using the Toolkit URL or to a VES that is a source of values for an element. RDA provides VESs for over 40 elements, and accommodates values from appropriate VESs maintained by other agencies. Sources do not need to be online or have URLs.

A large RDA community may require multiple profiles to cover the range of materials to be described. There is likely to be significant overlap in what elements, etc. are included in each profile. For example, a basic profile for all kinds of materials may be part of each profile for specific kinds of materials such as music and legal publications. This suggests that the efficiency and effectiveness of a set of application profiles could be improved by developing a mechanism for combining or augmenting existing profiles to create new ones. For example, an application profile for music materials might combine a basic profile for ensuring a description meets RDA's minimum conditions with a general profile reflecting a community's choices of Toolkit options and augmenting the result with elements that are specific to music.

This idea is being explored by RIMMF (RDA in many metadata formats), which is an external took for orienting and training users of RDA that can produce pure RDA metadata for small-scale applications [4]. The latest version, RIMMF4, includes a facility for creating templates that populate a data entry interface. This has been developed to accommodate additional parameters required by an application profile, and to allow the data entry screens to import a template to augment the data

fields already in use. The augmented input template has evolved into an application profile. RIMMF4 is freely available for download, and is usually up-to-date with the current release of the beta Toolkit.

The new Toolkit continues to accommodate policy statements which specify the choices favoured by large cataloguing agencies such as the British Library, DNB, and the Library of Congress. A policy statement can be attached to any part of the RDA guidance and instructions and displayed in context. A set of policy statements can be an effective substitute for a full application profile, albeit with less structure and flexibility. Policy statements are embedded within the infrastructure of RDA Toolkit, unlike application profiles which are typically presented outside of the Toolkit. Policy statements are usually optimized for human readability, while application profiles can be structured for machine-processing to support data input, validation, and extraction functions.

The new Toolkit provides alternatives to policy statements and structured application profiles. A cataloguer can add personal notes to most areas of the Toolkit. These can be used as a crude method for specifying simple choices, for example “This element is mandatory” or “Do not use this recording method”. The Toolkit also provides a range of document structures to accommodate workflows and other local documentation that can be used to specify profile parameters and choices. These are intended for use by agencies and can be shared with other Toolkit users.

Conclusion

Choice is essential to meet the requirements of local applications in a wider context. An international content standard cannot impose a single set of instructions to ensure consistency and interoperability without impairing its application in specific cultural and bibliographic environments.

Instead, global consistency of metadata applications is achieved by functional interoperability, embracing many views in a common framework. The values recorded in a metadata description can vary as widely as within a language and culture, or between languages and cultures. “What's in a name? That which we call a rose by any other name would smell as sweet” [5]. Functional interoperability is based on things, such as “a flower with a sweet smell”, and not their labels. RDA recognizes this and provides choice in what values are recorded but guarantees interoperability by using a rigidly defined set of containers for those values: the entities and elements that are based on the coherent semantics of the LRM.

References

[1] RDA Toolkit. Available at: https://www.rdatoolkit.org/ [RDA Tookit is fee-based, but free short-term registration is available]

[2] IFLA Library Reference Model. Available at: https://www.ifla.org/publications/node/11412

[3] Guidelines for Dublin Core application profiles. Available at: https://www.dublincore.org/specifications/dublin-core/profile-guidelines/

[4] RIMMF4. Available at: https://www.marcofquality.com/wiki/rimmf4/doku.php

[5] Shakespeare, William. Romeo and Juliet. Act 2, scene 2. Available at: http://shakespeare.mit.edu/romeo_juliet/romeo_juliet.2.2.html