

RDA for acquisitions librarians

Gordon Dunsire

This article is based on a presentation given to the National Acquisitions Group conference held in Manchester, 7-8 Sep 2011, and brings it up to date.

Introduction

The article introduces some of the principal features of RDA: resource description and access, including Functional Requirements for Bibliographic Records (FRBR) on which it is based. It describes relationships between RDA and other standards and discusses the application of RDA to linked data and the Semantic Web. It describes some of the features of the RDA Toolkit, and current activity in developing the standard. It ends by discussing briefly the potential impact on library acquisitions.

RDA

RDA: resource description and access is a new standard for creating bibliographic metadata. It is a successor to the Anglo-American Cataloguing Rules (AACR) which have been essentially evolving since 1841 when Anthony Panizzi created his rules for the catalogue of the library of the British Museum. Panizzi's cataloguing rules formed the basis of the two editions of AACR and the International Standard Bibliographic Description (ISBD) developed during the 20th century. RDA was initially intended to be the third edition of AACR, taking into account the latest ideas of the purpose and function of catalogue data in the digital environment. A major influence on this activity was the model of Functional Requirements for Bibliographic Records (FRBR) published in 1998 and developed and maintained by the International Federation of Library Associations and Institutions (IFLA). It proved impossible to adapt AACR2 to FRBR, so the decision was taken to create a new standard, ensuring that it was as backward-compatible with AACR2 as possible. RDA also uses the model of Functional Requirements for Authority Data (FRAD) and the Statement of International Cataloguing Principles (ICP), both also developed by IFLA during the first decade of the 21st century.

RDA provides a set of guidelines for determining the content of bibliographic metadata; that is, descriptions of the items found in the collections of libraries, archives, and to a certain extent, museums, and of associated entities such as the persons, corporate bodies or families who are the creators, owners or publishers of such items. The guidelines specify where information is to be found and how it is to be recorded, so that the metadata produced by different cataloguers for a wide range of resources is coherent and consistent. RDA breaks the descriptions into a set of discrete elements which are based on, but often refine, the attributes outlined in the FRBR model. Those attributes are mapped to a set of user tasks identified by FRBR: to find, identify, select, and obtain resources relevant to the user's needs. RDA is thus much more focussed on users of bibliographic metadata than previous cataloguing codes and content standards.

RDA supports catalogue structures that are based on FRBR. These are intended to provide metadata to all types of user, including end-users who need to find, identify, select, obtain and use information, and intermediaries who manage and organize information bibliographically. The structures cover all types of media, including print-based, digital, textual, visual, and other resources. RDA gives equal, even treatment to resources irrespective of type, giving in turn more control to the user in finding and choosing the most appropriate resources for their needs.

FRBR replaces the monolithic description of a resource with four main parts: Work, Expression, Manifestation, and Item. In reverse order, these go from concrete to abstract. Broadly speaking, an item is an individual copy of a resource; a manifestation is the physical carrier of a resource, common to all copies; an expression is the content of a resource; and a work is the abstract notion that shapes the content. A work can have multiple expressions, including translations. An expression can have multiple manifestations, for example when the content is carried on print and digital formats, but is associated with only one work. A manifestation can obviously have many items. An item can be associated with only one manifestation. This is not strictly a tree-like structure, because a manifestation can be associated with more than one expression, for example when a translation is bound with the original expression in a single volume. However, this does offer a significant economy over a set of monolithic descriptions in a large catalogue, when a work has multiple expressions which have multiple manifestations which have multiple items. The work, expression, and manifestation metadata does not have to be repeated in every description for all of the items; the metadata in the four parts can be linked to avoid unnecessary and expensive duplication. Likewise, a user does not have to be faced with repeating information in every description retrieved in a resource discovery system: the work description can be opened-up or drilled-down by the user to show the related expressions, and then expressions of, say, a specific language translation can lead to all the manifestation formats available. More simply, a user of a large catalogue is not faced with several pages of titles for, say, Shakespeare's Macbeth with little information to distinguish them. Instead, a single work-level description can be displayed, from which the distinct expressions and then manifestations can be explored under the user's control. Having said that, most current online catalogues can only process monolithic records, often in a MARC (machine-readable catalogue) format, so they do not benefit from this feature. The two main MARC formats, MARC21 and UNIMARC, have been amended to accommodate RDA elements.

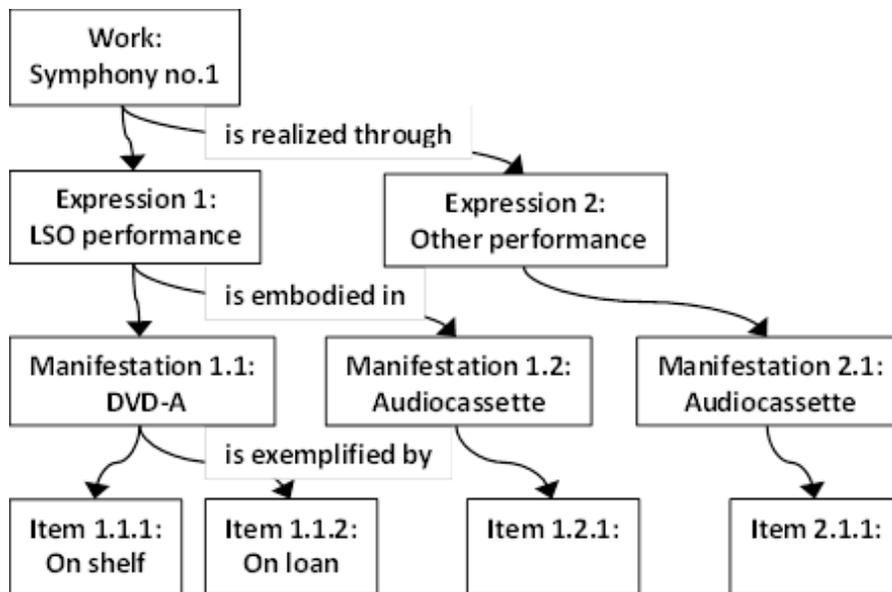


Figure 1: Simple FRBR example, showing relationships between work, expression, manifestation and item.

Figure 1 shows a simple example of the FRBR structure. It includes the basic relationships between a work, expression, manifestation and item. The hierarchical numbering is for readability and is not part of FRBR. A new manifestation of Expression 2, say a digitization of the analogue audiocassette of Manifestation 2.1 onto a recordable CD, can be related to Expression 2 with a new “is embodied in” link.

FRBR also includes a set of relationships between multiple works, expressions, manifestations and items which help the user to navigate through the increasingly diverse bibliographic universe. For example, a work may be linked to another work which is a summary or other derivation, or a successor or predecessor, and so on. RDA builds on and refines these relationships to support the rich network of links required in large resource discovery systems like union catalogues and other metadata aggregations.

There are many other user-centred features of RDA. One of the more powerful ones is that RDA clearly distinguishes content from carrier, such as moving pictures on DVD and text on CD-ROM: moving pictures and text are the content, while DVD and CD-ROM are computer disk carriers. This had become an increasing problem for AACR2 and other cataloguing codes, as computer disks and other storage media have been developed from carrying just audio content (the original CD) to text and still and moving images. RDA has special controlled vocabularies for content and carrier types which are based on an underlying set of attributes such as the human sense required to access the content (sight, hearing, etc.) and the type of intermediation device required to use the carrier (projector, computer, etc.). These can be very helpful for users with special needs so that, for example, a search can be restricted to resources that do not require sight or a computer.

RDA has many more controlled vocabularies than AACR2, ranging from the two terms for Groove pitch (“fine” and “standard”) to an extensive list of Encoding formats for digital manifestations. These reduce

transcription errors, improve consistency, and support user-interface functions like pull-down filters and context-sensitive help.

RDA has been designed for a multinational audience of users and cataloguers. Many areas of Anglo-centricity, as well as cataloguer-eccentricity, have been removed. Abbreviations and acronyms are generally avoided because they disadvantage users with a different educational or cultural background. Latinisms such as “s.l.” (sine locum) for “no place” and “et al.” for “and others” have been removed, despite objections that “everybody knows what they mean” and “Latin is perfect for an international environment because it is a dead language”. RDA is being translated into many other languages, including French and German.

RDA is independent of technical metadata formats, and it does not favour or promote one format over another. It can therefore be used with MARC, DC (Dublin Core), Encoded Archival Description, and the other formats that have proliferated in recent years. As a result, users will be able to see familiar metadata regardless of what system they are using.

RDA is designed for the digital environment, increasingly familiar and comfortable for today’s users. The main expression and manifestation of RDA is the RDA Toolkit, which is published as an online product. This has the potential to allow resource discovery systems to pull information from RDA to display to the user, for example the definitions of controlled vocabularies such as the content and carrier types.

There are several cataloguer-centred features of RDA. More emphasis is put on cataloguer’s judgment instead of providing lengthy prescriptive rules for every eventuality. The guidelines are grouped by bibliographic element rather than format and, as already noted, the bibliographic elements are related to the FRBR entities which in turn are related to user tasks. The cataloguer is encouraged to ask “Why am I recording this information?” and to appreciate how it supports resource discovery by a user.

Authority control of headings for people, corporate bodies, families and subjects is included in the guidelines, whereas it was excluded from AACR2. RDA is, however, generally compatible with AACR2 and will not be entirely unfamiliar to staff trained in the previous standard.

RDA and other metadata communities

RDA has been able to benefit from formal and informal partnerships with metadata communities other than IFLA. A working group drawn from RDA and the developers of ONIX (Online Information Exchange), the publishing industry metadata standard, resulted in the publication of the first version of the RDA/ONIX framework for resource categorization in 2006. The framework is what underpins the content and carrier type vocabularies in RDA, but is intended for general use.

RDA also enjoys a fruitful relationship with the Dublin Core Metadata Initiative (DCMI) and Semantic Web communities. A meeting with individuals working within these communities was held at the British Library, London, in 2007. This led to the formation of the DCMI RDA Task Group, with a charter “To define components of the draft standard ‘RDA - Resource Description and Access’ as an RDF vocabulary

for use in developing a Dublin Core application profile". As a result, the RDA elements and controlled vocabularies have been represented in Resource Description Framework (RDF), the basis of the Semantic Web, and added to the Open Metadata Registry (OMR) where they are freely available for use. These representations are in the process of being synchronized with the RDA Toolkit.

This activity stimulated the FRBR Review Group to initiate a project to also represent the Functional Requirements (FR) family of models, including FRBR, FRAD and Functional Requirements for Subject Authority Data (FRSAD), in RDF. The FR elements were also published using the OMR, in 2011. Development of a consolidated model incorporating the FR family is underway, and is being informed by the analysis required to develop the RDF representations. RDA and FRBR continue to liaise over the evolution of both standards.

The ISBD Review Group followed suit, publishing the element set and content and media type vocabularies of the new consolidated edition in the OMR, also in 2011. A joint meeting of the ISBD Review Group and the Joint Steering Committee for Development of RDA (JSC) in Glasgow in late 2011 agreed to update an alignment between the ISBD and RDA element sets and create appropriate mappings in RDF to improve the interoperability of linked data derived from each standard. The meeting also agreed to map the ISBD content and media type vocabularies to the RDA/ONIX framework for the same purpose.

RDA today

The three "national" libraries of the United States undertook a test of RDA, completed in 2011. The report made several recommendations for improving RDA, and further recommended that RDA be implemented by the Library of Congress, National Agricultural Library, and National Library of Medicine sometime after the beginning of 2013. As a result, the Library of Congress and its partner national libraries, including the British Library, have announced their intention to implement RDA by the end of March 2013.

Another important recommendation from the RDA test was that the U.S. national libraries should demonstrate credible progress towards a replacement for MARC. The Library of Congress subsequently announced the Bibliographic Framework Transition Initiative, although there has been little progress to date.

RDA itself continues to develop. Placeholders for authority control guidelines are now being filled by JSC following the publication of FRAD and FRSAD. Suggestions for improving RDA are welcomed, and there are a number of specialist groups working on various aspects of the standard. For example, the Committee on Cataloging: Description and Access (CC:DA) of the American Library Association is investigating how the Extent element can be structured to improve machine-processing; this would allow, for example, a library to better calculate shelving requirements for audio-visual carriers or the number of pages to be scanned in digitization projects.

The RDA Toolkit introduces new features at regular intervals to make it easier to use RDA. The JSC itself is expanding to include a wider representation beyond the traditional Anglophone communities, and recently added a representative from the Deutsche Nationalbibliothek.

RDA impact on acquisitions

RDA is unlikely to have any immediate significant impact on library acquisitions processes. A replacement for MARC needs to be identified and implemented, and library management systems re-engineered, before the main benefits of RDA are realized. This is some way off; perhaps 3-5 years or longer.

The longer term impact is likely to be profound, affecting every aspect of how metadata is used for acquisitions. RDA should be seen as part of a much larger transition into the global online environment of linked data and the Semantic Web, promising a revolution in resource discovery that exceeds the impact of the Web itself on publishing and user-driven accessibility and sharing of data. Acquisition processes already benefit from social networking infrastructure and the availability of publisher and bookseller metadata along with user reviews and recommendations. RDA should improve the discovery of digitized and e-book versions of print resources; they are additional manifestations of existing works and expressions. The interoperability of RDA metadata with metadata from other library standards and formats in global assemblages of linked data will improve support for the trend in distributed storage of and access to resources by providing information about nearest available copies, digitization on demand services, and other alternatives to purchase for the local collection. It will also be much easier to compare prices for various formats, print and digital, offered by a wider range of suppliers.

How long do we have to wait for this linked data utopia? There is an enormous amount of work to be done, particularly in creating high-level applications from the foundations of RDF. Only a tiny percentage of existing library catalogue records has been converted to RDA. Only a tiny percentage of existing library metadata has been published as linked data. But there are many technical and professional groups working in these areas, and the potential rewards are very high, so we may see new applications (rather than implementations of existing applications) emerge in 3-10 years' time. On the other hand, the revolution may be just around the corner from an unexpected direction; think Google, Facebook, Twitter, YouTube, Pinterest ...

Further online reading

Bibliographic Framework Transition Initiative: <http://www.loc.gov/marc/transition/>

Cataloguing standards [at the British Library]: <http://www.bl.uk/bibliographic/catstandards.html>

CILIP-BL Committee on RDA: <http://www.slainte.org.uk/aacr/>

Distinguishing content from carrier: the RDA/ONIX framework for resource categorization: <http://www.dlib.org/dlib/january07/dunsire/01dunsire.html>

European RDA Interest Group: <http://www.slainte.org.uk/eurig/>

FR family in the Open Metadata Registry: <http://iflstandards.info/ns/fr/>

Functional Requirements for Bibliographic Records: <http://archive.ifla.org/VII/s13/frbr/frbr.pdf>

Joint Steering Committee for Development of RDA: <http://www.rda-jsc.org/>

RDA/ONIX framework for resource categorization: <http://www.loc.gov/marc/marbi/2007/5chair10.pdf>

RDA Toolkit: <http://www.rdatoolkit.org/>

RDA in the Open Metadata Registry: <http://metadataregistry.org/rdabrowse.htm>

Resource Description and Access (RDA) [at the Library of Congress]: <http://www.loc.gov/aba/rda/>

Statement of International Cataloguing Principles: http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf

Testing Resource Description and Access (RDA): <http://www.loc.gov/bibliographic-future/rda/>

Acronyms

AACR: Anglo-American cataloguing rules

AACR2: Anglo-American cataloguing rules, 2nd edition.

CC:DA: Committee on Cataloging: Description and Access

CD: Compact disc

CD-ROM: Compact disc read-only memory

DC: Dublin core

DCMI: Dublin Core Metadata Initiative

DVD-A: Digital versatile disc audio

FR: Functional requirements [family]

FRAD: Functional requirements for authority data

FRBR: Functional requirements for bibliographic records

FRSAD: Functional requirements for subject authority data

ICP: International cataloguing principles

IFLA: International Federation of Library Associations and Institutions

ISBD: International standard bibliographic description

MARC: Machine-readable cataloguing

OMR: Open metadata registry

ONIX: Online information exchange

RDA: Resource description and access

RDF: Resource description framework