RDA (Resource Description and Access) and its application to rare books, manuscripts, and their digital surrogates

Gordon Dunsire

Summer School in the Study of Rare Books, Zadar, Croatia, 28 September–2 October 2009; Topic III: New Conceptual Models for Information Organization

Pre-print

Published by Gordon Dunsire

Edinburgh 2012

RDA (Resource Description and Access) and its application to rare books, manuscripts, and their digital surrogates

Gordon Dunsire
Head, Centre for Digital Library Research
University of Strathclyde, Glasgow, Scotland

Abstract

The paper introduces the basic concepts of RDA - Resource Description and Access, a new set of guidelines for determining the content of bibliographic records, and Functional Requirements for Bibliographic Records (FRBR), a model for the structure of bibliographic records, and how they are aligned with each other. The paper discusses the application of RDA to describe bibliographic resources in general, and the specific guidelines given for rare books, manuscripts, and their digital surrogates.

Keywords

RDA; FRBR; rare books; manuscripts; cataloguing.

RDA

RDA: resource description and access is a new standard for creating bibliographic metadata [9]. It builds on the Anglo-American Cataloguing Rules (AACR) which have been in development since 1841 when Panizzi's rules for the British Museum catalogue were created. RDA is also based on more modern approaches, including the model developed in FRBR published in 1998 [8].

There are many features of RDA which have the use of bibliographic metadata as a focus. It is intended to meet the needs of all types of user, including those who need to find, identify, select, obtain and use information, and manage and organize information bibliographically. It covers all media, whether print-based or digital and irrespective of content, be it textual, visual, etc. Media are treated in an equal and even way which gives more control to the user in finding and choosing the most appropriate resources to suit the task in hand. RDA clearly distinguishes content from carrier, essential when almost any kind of information content can be carried on digital media, such as moving pictures on DVD and text on CD-ROM. The distinction is particularly helpful for users with special needs who may wish, for example, to exclude non-visual resources from a search. RDA provides controlled vocabularies for describing the characteristics of the content of a resource separately from the characteristics of the carrier of that content.

RDA is designed for multinational application. Anglo-American-centric features of AACR have been removed, along with cataloguer-eccentric practices arising from the legacy of over one hundred years of development and application. For example, abbreviations and acronyms are generally avoided, and Latinisms such as "s.n.", "s.l." and "et al." do not feature in RDA. The guidelines emphasise transcription over interpretation; the exact title of the resource, as displayed on the resource, should be transcribed as the content of the title element in the bibliographic record regardless of misprints or other errors, with any corrections given in notes. This ensures consistency of approach across cataloguing communities and cultures.

RDA is independent of technical metadata formats. It can therefore be used with MARC, DC (Dublin Core) and many other metadata storage and exchange formats. Application of RDA is intended to give the user familiar metadata regardless of what system is used to store, retrieve and display them. RDA is designed for the digital environment. It will be published as an online product, so there is potential for incorporating appropriate parts of RDA in user help facilities; for example, an explanation of how a "preferred title for the work" (equivalent to the uniform title of AACR) is derived. The online digital publication of RDA will allow it to be incorporated into online cataloguing software to provide contextualised workflows and local customization.

The core user-centred features of RDA, however, result from its close alignment with FRBR and the related model presented in Functional Requirements for Authority Data (FRAD) [4]. This allows RDA to be applied readily to catalogue records based on the FRBR and FRAD models and encourages the development of catalogue interfaces which better meet the requirements of users.

FRBR

The FRBR final report presents a common understanding of the aims and purpose of bibliographic metadata, with a framework and model to support that understanding. FRBR was developed for International Federation of Library Associations and Institutions (IFLA) between 1992 and 1997. FRBR takes a user-focussed approach, and identifies a set of user tasks associated with bibliographic metadata:

- Find an information resource matching the user's search criteria.
- Identify a resource confirming the user's criteria and distinguishing similar resources.
- Select a resource meeting the user's needs.
- Obtain a resource to access the information in the resource.

The FRBR model is based on an entity-relationship analysis of the data typically found in bibliographic records. An entity is a key object of interest to a user and acts as a focal point for a set of metadata statements or attributes. Entities interact with each other through one or more defined relationships.

The model identifies three groups of entities.

- Group 1 entities represent the different aspects of user interests in the products of intellectual or artistic endeavour.
- Group 2 entities represent what is responsible for the content, production and custodianship of Group 1 entities.
- Group 3 entities represent the subjects of one of the Group 1 entities.

There are four entities in Group 1:

- A work is a distinct intellectual or artistic creation.
- An expression is an intellectual or artistic realisation of a Work.
- A manifestation is a physical embodiment of an Expression.
- An item is a single exemplar of a Manifestation.

A work is realised through an expression which is embodied in a manifestation which is exemplified by an item. Conversely, a specific bibliographic resource in the hand of a cataloguer or sought by a user is an item which is an exemplar of a manifestation which is an embodiment of an expression which is a realisation of a work. All relationships between entities defined in the model come in reciprocal pairs; an expression is embodied in a manifestation, a manifestation is an embodiment of an expression.

Although a work can have multiple expressions, an expression multiple manifestations, and a manifestation multiple items, the reverse is restricted to single instances: an item can have only one manifestation, a manifestation one expression, and an expression one work.

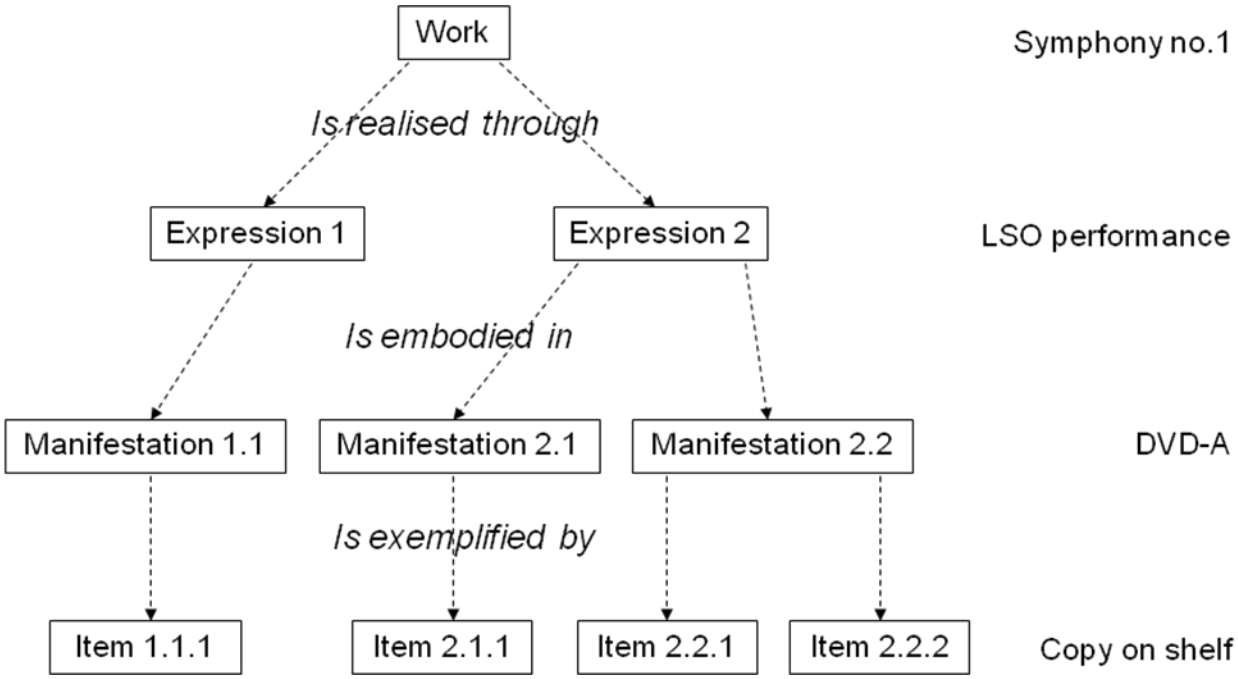

Diagram 1: Theoretical example of FRBR Group 1 entities and relationships

Diagram 1 shows the FRBR Group 1 entities and primary relationships applied to a theoretical musical work which is performed (expressed) on two separate occasions during which it is recorded. One recording is issued (manifested) on, say, audiocassette, while the other is issued on both audiocassette and DVD-A. One copy of each of the audiocassette manifestations is available, along with two copies of the DVD-A. The text beside the diagram refers to item 2.2.2: a copy of the DVD-A containing the LSO performance of Symphony no.1. The entity numbers are for referencing and are not part of the model.

This is quite different from the traditional approach, where the metadata for each manifestation and related expression and work are held in a single record (item metadata are typically held in a separate data file in modern automated systems). The example would consist of three separate records which duplicate some of the expression and all of the work attributes, and the user might have to examine all three records to be sure of fulfilling their task. The tree structure associated with

a catalogue based on the FRBR model is more intuitive and efficient; the user can carry out their tasks in stages by drilling-down from the top, and avoiding the display of duplicate metadata.

RDA and cataloguing

The cataloguer using RDA, on the other hand, will typically traverse the structure from the bottom up: determining the manifestation characteristics of the item in hand, then the expression, and finally the work.

RDA places more emphasis on cataloguer's judgment, and the instructions for determining metadata content are intended to be used as guidelines rather than "rules". Nevertheless, those instructions are likely to be referred to as cataloguing rules, as does this paper. RDA groups rules by bibliographic element or attribute such as title, rather than by format as in AACR. Those attributes are explicitly related to FRBR entities, and therefore ultimately to user tasks, to assist cataloguers in determining the utility of the metadata they create. They help the cataloguer to focus on why they are recording the information, rather than the specific infrastructure used to create and store it.

RDA includes guidance on authority control, now a much more important aspect of the bibliographic metadata environment than in the past as a result of the development of global information and communication networks. As the number of authors and subjects that can be retrieved in an online search increases, so does the need to determine whether different labels refer to the same author or subject, and similar labels refer to different authors or subjects.

Despite these new features, however, RDA is generally compatible with AACR, and will not be entirely unfamiliar to many cataloguers.

The objectives of RDA encompass comprehensiveness, consistency, clarity, rationality, currency, compatibility, adaptability, ease and efficiency of use, and format. The principles of RDA cover generalization, specificity, non-redundancy, terminology, and reference structure. The IFLA Statement of international cataloguing principles [7], published in 2009, is used to inform the cataloguing principles used throughout RDA.

As well as alignment with various IFLA models, RDA is influenced by, and in turn influences, approaches used by several other related communities. The controlled vocabularies for describing content and carrier types are based on the RDA/ONIX framework for resource categorization [10], a collaboration between RDA and the publishing community. The framework provides an ontology and methodology for creating "high-level" categories of content and carrier which are the basis for controlled terms. For example, "text" and "still image" are terms which can be used to describe the content type attribute of a FRBR expression, while "volume" and "online resource" are applicable to the carrier type attribute of a FRBR manifestation. The ontology is intended for general use, beyond the library and publishing communities [3].

Another initiative is the DCMI RDA Task Group [2], a collaboration between RDA and the Dublin Core Metadata Initiative. The aim of the group is to define RDA elements and controlled terms as an RDF [12] vocabulary, compatible with the Semantic Web, for use in developing a Dublin Core application profile [1], [5]. And, linking these various collaborations together, the Vocabulary Mapping Framework project [11] is developing an extension to the RDA/ONIX framework to improve

Semantic Web interoperability between elements of RDA, ONIX, DC, FRBR, FRAD and a number of other related metadata models and vocabularies.

The RDA elements are assigned to FRBR Group 1 entities; for example, the element for preferred title is part of the work entity. Although RDA does not mandate any element, some are designated as core elements of an entity. Core elements are those required to support the functional objectives of RDA entity and relationship elements. Special communities, such as those dealing with rare books or manuscripts, are expected to designate additional core elements. A full set of core elements is likely to form a minimum input standard for a community, along with guidance on the use of some of the few options provided in RDA, such as the preferred language of cataloguing agency.

Communities should take into account the purpose of the metadata, especially from the point of view of the user, and be aware of the availability of relevant metadata from other sources. For example, information about the nearest available item of a resource, helping a user to obtain it, can be typically obtained from a combination of bibliographic metadata with collection-level description and borrower transaction metadata. User needs often focus on manifestation and expression metadata rather than items. Expression metadata inform the user about editions and translations, while manifestation metadata indicate, for example, whether special equipment is required to access the resource, such as a computer with PDF reader software. There may little requirement for item metadata, other than, say, a barcode, if all items of a manifestation are essentially identical. On the other hand, specialized communities may need a different approach to how much metadata is required in the expression, manifestation and item areas. Rare books communities are likely to have a greater focus on item metadata than general libraries, for example recording the numbering of an item in a limited edition, or the provenance of an item. A manuscripts community may choose to ignore the item altogether if all useful metadata can be accommodated in the manifestation area.

The following examples of RDA rules for rare books and manuscripts illustrate some of the features of RDA. The examples are taken from the full draft of RDA produced for constituency review in 2008 [14]. The final version of RDA has not yet been published, but it is unlikely that there will be significant change from the draft.

RDA uses a hierarchical numbering system for delineating sections and rules. For example, RDA 2.8.4.1 states “For early printed resources, printers and booksellers are treated as publishers.” The number shows that this is a subsection of the rules for publisher's name (2.8.4), which is part of the publication statement (2.8) in the section dealing with identifying manifestations and items (2). This rule hierarchy aids the logical layout of RDA, but can be transparent in the digital environment in which RDA will be published, where hyperlinks can be used to group rules together and navigate through them.

The phrase “early printed resources” is controlled terminology; it is used throughout RDA to refer to the same type of resource, and is semantically equivalent to the controlled term "older monographic resources" used by International standard bibliographic description (ISBD) [6]. The use of controlled terminology improves the clarity and ease of use of the rules. Other relevant controlled terms used to refer to resource types in RDA are "manuscript" and "digital resource".

The RDA rule for recording the production method for manuscripts (RDA 3.9.2.3) says “If the resource is a manuscript, record one of the following terms as appropriate.” The terms are:

- holograph (written by the work creator)
- manuscript (written by other than the creator)
- printout
- typescript

This is another example of the use of controlled terminology, this time in the metadata created using RDA. It improves clarity and consistency for the user.

The RDA rule for title of the work (6.2) has subsections for early works and incunabula. For works created before 1501 (6.2.2.5) "... choose the title or form of title in the original language by which the work is identified in modern sources as the preferred title." For incunabula (6.2.2.8) "Choose as the preferred title for a work contained in an incunabulum the title found in standard reference sources for incunabula." Both of these rules are based on the more general RDA principle of representation (0.4.3.4) which states "... the preferred title for a work should be ... the title most frequently found in resources embodying the work in its original language, the title as found in reference sources, or the title most frequently found in resources embodying the work." This again improves consistency, both for the cataloguer and the user.

Digital surrogates

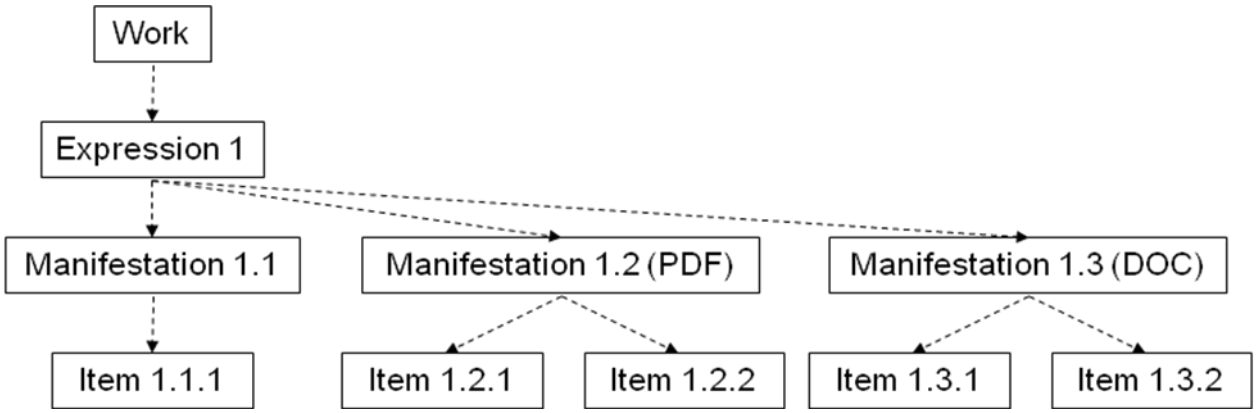

Diagram 2: FRBR model of a rare book and its digital surrogates.

Diagram 2 shows the FRBR model of a rare book which is the single extant item of a work. Item 1.1.1 is the sole exemplar of manifestation 1.1 which is the sole embodiment of expression 1 which is the sole realization of the work. Digitizing the content of the item is equivalent to producing another manifestation (1.2) of the expression. The new manifestation is a digital carrier with its own characteristics, including the kind of technical encoding it uses to store and represent the content of the expression; in this example, the encoding is portable document format (PDF). One of the objectives of digitizing a rare item is to improve access, which can be achieved not only by publishing its content online but also by providing multiple copies. Here, two copies have been made available (items 1.2.1 and 1.2.2). If the digitized expression is encoded in a different format, such as Microsoft

Word document format (doc), a different manifestation (1.3) is created, which can have its own multiple copies.

Digitization, then, typically results in one or more new manifestations for which metadata must be created. If those manifestations use a physical carrier such as CD-ROM or DVD, then item metadata are also required to distinguish copies. If a manifestation is made available as an online resource in several places (that is, with more than one electronic address or Uniform Resource Locator (URL)), then explicit item metadata are required to describe those addresses. However, if the manifestation is an online resource at only one location, item metadata may be embedded in the manifestation record rather than form a discrete item record.

Digitization does not usually result in the need for new work or expression metadata. In a cataloguing system using discrete FRBR records a new manifestation record is merely linked to the existing expression record, as in Diagram 2. In a more traditional system, the new manifestation metadata is combined with a copy of the existing work and expression metadata to form a new catalogue record for the resource.

Relating manifestations

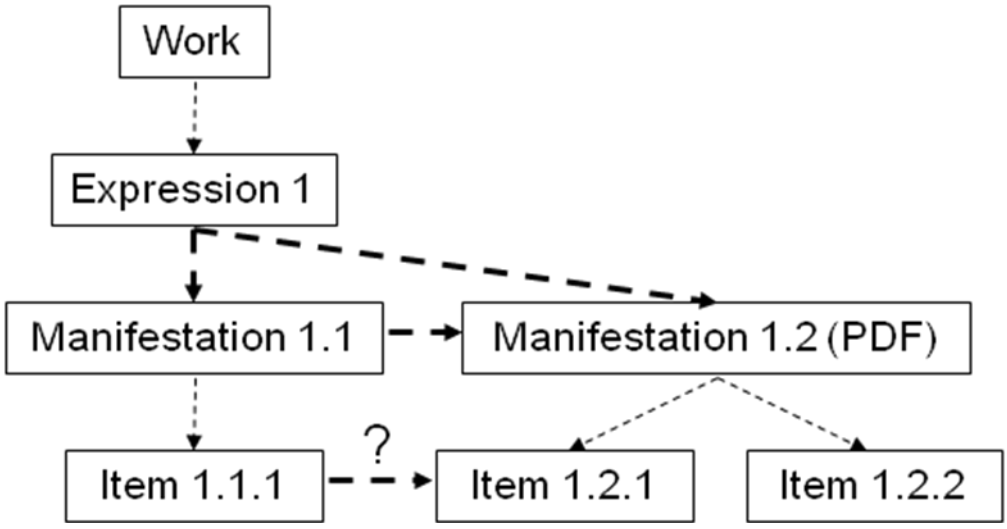

Diagram 3: Implicit and explicit relationships in the FRBR model.

Diagram 3 shows the FRBR model for a manuscript and a digital surrogate. As with the example in Diagram 2 there is only one manifestation and item for the original resource, as is typical for a manuscript. The manuscript has been digitized in PDF format and made available online at two different URLs.

There is an implicit relationship between the two manifestations because they have the same expression. But this relationship cannot describe, for example, whether the digital manifestation is a scanned image of the original manifestation or was derived from another, unrecorded digital manifestation using a different encoding scheme (as might be the case with the PDF and DOC format

manifestations in Diagram 2). Although this information may be included in each manifestation record in the form of a note or some other attribute, it may not be in a form from which a direct manifestation to manifestation relationship can be inferred by a human, let alone a machine.

It is desirable, then, that relationships between manifestations are explicitly stated in the metadata. FRBR provides a set of relationships between Group 1 entities; the appropriate relationship pair for this example is "has a reproduction" and "is a reproduction". RDA further refines this by adding two sub-relationships:

- digital transfer: A manifestation resulting from the transfer of another manifestation from an analog format to a digital one, or from one digital format to another.
- electronic reproduction: A digital manifestation resulting from the reproduction of an analog manifestation.

Each of these has a reciprocal. So a scanned image manifestation is an electronic reproduction of the original manifestation, while a re-encoded digital manifestation is a digital transfer of the manifestation with the original encoding.

There is a question about whether the relationship in diagram 3 should be created more appropriately between the items rather than the manifestations. RDA allows for this and provides the relationships "digital transfer of (item)" and "electronic reproduction of (item)" if the cataloguer decides that this is a better approach, but it should only be used if it is important to distinguish the particular item used as the source of the digital surrogate. This might be the case, for example, if the only item of a rare book available for digitization is damaged or incomplete. Conversely it might not be the case, for example, for a manuscript if there has only ever been one item, and the manifestation and item can be effectively treated as a single entity.

An extensive set of FRBR Group 1 relationships is given in Appendix J of RDA.

Organizing metadata

The "traditional" approach used by modern library catalogues is based on the concept of a catalogue record for each resource containing all the metadata relevant to the resource except authority headings which are linked to the record. Metadata created using RDA, and therefore FRBR, can be accommodated by this approach; indeed, this is likely to be the scenario when RDA is adopted by communities already using AACR. RDA is independent of metadata formats, and there is no specific reason why the order of metadata elements in a record should follow the sequence given by RDA in order to group them as FRBR entities. Some of those elements will be used, variously, to create indexes, short and full record displays, etc. following current practice.

But this approach does not take full advantage of the FRBR model. Work and expression metadata has to be duplicated for each manifestation to form the resource record, whereas FRBR allows a single instance of a work or expression record to be linked to multiple manifestation records. Avoidance of duplication is more efficient as it takes less cataloguer time, and is more effective for the user as it improves hierarchical browsing. It may not always be possible to process resource records to create a virtual, duplicate-free FRBR view of the catalogue for the user, particularly if records are taken from multiple cataloguing sources with variations in adherence to RDA, even though RDA strives to maximize the potential for such processing by, for example, giving clear and

unambiguous instructions for determining titles and other match-points. A "best value" application of RDA and FRBR instead suggests that future catalogues may consist of discrete work, expression, manifestation, and (possibly) item records linked together using global machine-processable identifiers to allow aggregations from multiple catalogues. This link rather than copy approach is characteristic of the Web today, and underpins the development of the Semantic Web [13].

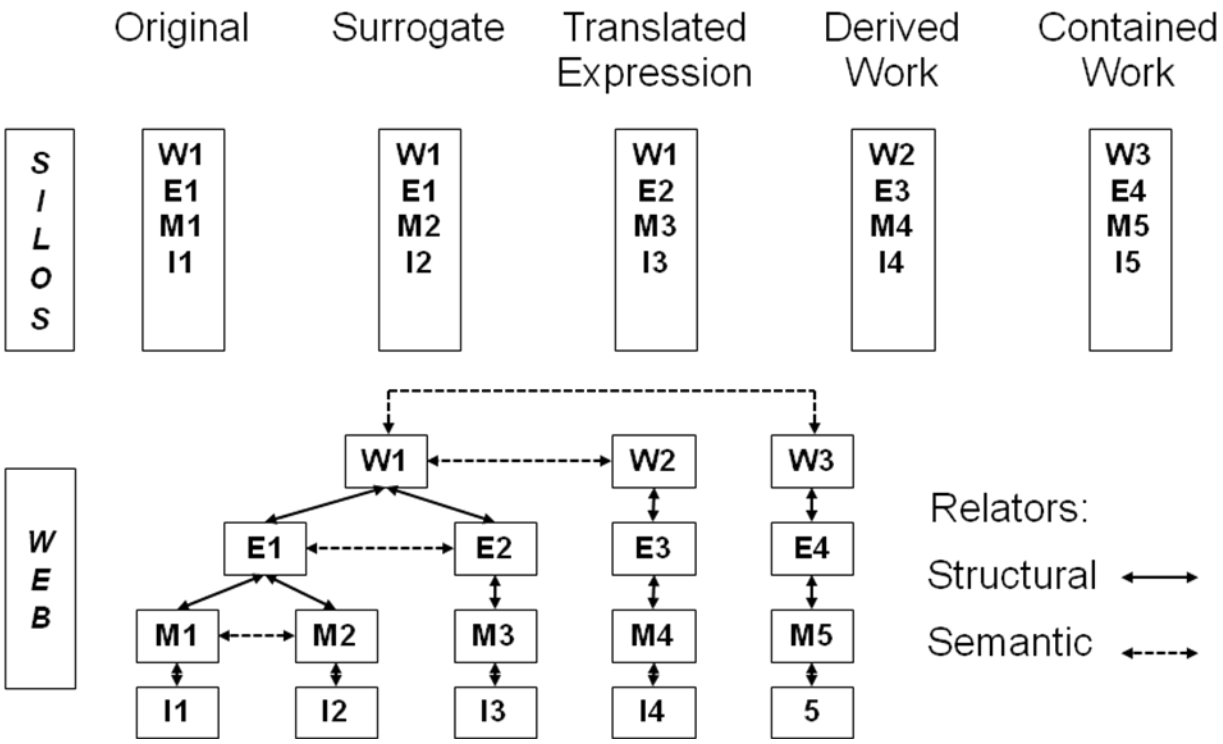

Diagram 4: Metadata silos vs web of links

The current, traditional approach and potential linked-record approach are contrasted in Diagram 4. Five separate resources are represented, consisting of the manuscript and digital surrogate given in Diagram 3 together with a translation of the content of the manuscript, a summary derived from the content, and a part of the content (for example an illustration) which the cataloguer has decided is self-contained and important enough to warrant treatment as a separate work with its own metadata. The translation is a new expression because it changes the content without affecting the abstract creation behind it; the work remains the same. The summary is treated as a new work because it has a different creator and conception.

The top part of Diagram 4 shows the five records and their FRBR component entities that result from the traditional approach. (In practice, the cataloguer may choose to reduce the number of records in the same catalogue by just adding, say, a note about the translation to the record for the original manuscript, but this makes it more difficult for a user to find and identify the translation, and in any case consistency cannot be guaranteed in an aggregation of records from different catalogues.) The duplication of entities is indicated by their numbering. The metadata associated with the entities may not, in fact, be exactly duplicated, especially if, as has already been noted, the records are created by different cataloguers at different times; this makes it much more difficult for links

between these records to be generated by machine, or for the content to be de-duplicated for display. So the user must browse the content of all five records, and usually has to carry out additional searches to find associated works rather than navigating seamlessly from one to another.

The bottom part of Diagram 4 shows the result of treating the same metadata as discrete FRBR records linked by explicit relationships. Relationships can be broadly categorized as structural, reflecting the logical structure of FRBR, and semantic, indicating the differences in meaning represented in specific relationships. The FRBR relator "is embodied in" is structural, while the RDA relator "electronic reproduction" between M1 and M2 is semantic. In the diagram, the RDA relator between W1 and W2 is "summary (work)", that between W1 and W3 is "contained in (work)", and that between E1 and E2 is "translation of (expression)"; all are semantic. Although this results in a larger number of discrete records, 12 instead of 5 (excluding item records), each is significantly smaller and overall the metadata require fewer personnel and time resourcing to create, maintain, store, display and browse.

Conclusion

This paper has given an overview of RDA and FRBR and discussed some aspects of their application to rare books and manuscripts metadata. Communities of practice for those types of resource will need to consider a number of issues before being able to exploit the features of RDA and obtain best value from them.

Some have been mentioned in this paper, including the designation of additional core metadata elements and development of a minimum input standard, and the need, or not, to distinguish item and manifestation metadata. Others include the circumstances in which components of a work might be treated as separate works in their own right. For rare books, there may well be a case for describing owner bookplates and stamps, and even bindings, as works embedded in the book. The use of controlled terms for relators between FRBR Group 1 and Group 2, given in Appendix I of RDA, is another consideration; there is an extensive list of relators of interest to the rare books and manuscripts communities, including “book designer”, “engraver”, “platemaker”, “printer”, “current owner”, “previous owner”, and “autographer”.

RDA appears to offer a contradictory mix of guidelines which on the one hand encourage greater uniformity of metadata, for example through the use of controlled vocabularies, and on the other allow different interpretations of what metadata should be recorded, for example through community-specified core element sets. But this is a reflection of the proliferation of metadata content and structure models and standards which have emerged in the past few years; a one-size-fits-all does not seem to meet the needs of the various communities in the bibliographic environment. RDA allows a community to select whatever subset of RDA's rich metadata element set suits the community best, and ensures that the bibliographic descriptions produced using that subset are as consistent as possible.

Application of RDA by specialised communities of practice such as rare books and manuscripts will improve the potential for their metadata to be better integrated with that produced by other communities, to accommodate the outputs of digitization initiatives, to contribute to the emerging Semantic Web, and, most importantly, to provide an information retrieval environment that better serves the needs of its users.

References

[1] Coyle, Karen; Tom Baker. Guidelines for Dublin Core application profiles. 2009. Available at: http://dublincore.org/documents/profile-guidelines/

[2] DCMI/RDA Task Group wiki. [Cited: 2010-01-22]. Available at: http://dublincore.org/dcmirdataskgroup/

[3] Dunsire, Gordon. Distinguishing content from carrier: the RDA/ONIX framework for resource categorization.
D-Lib Magazine 13, 2 (January/February 2007). [Cited: 2010-01-22]. Available at: http://www.dlib.org/dlib/january07/dunsire/01dunsire.html

[4] Functional requirements for authority data: a conceptual model / edited by Glenn E. Patton. München : K.G. Saur, 2009.

[5] Hillmann, Diane; and others. RDA vocabularies: process, outcome, use.
D-Lib Magazine 16 1/2 (January/February 2010). [Cited 2010-01-22]. Available at: http://www.dlib.org/dlib/january10/hillmann/01hillmann.html

[6] IFLA. ISBD Review Group. International standard bibliographic description (ISBD). Preliminary consolidated edition. 2007. [Cited: 2010-01-22]. Available at: http://www.ifla.org/files/cataloguing/isbd/isbd-cons_2007-en.pdf

[7] IFLA. Statement of international cataloguing principles. 2009. [Cited: 2010-01-22]. Available at: http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf

[8] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records: final report. 1998, amended Feb. 2009. [Cited: 2010-01-22]. Available at: http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

[9] RDA - resource description and access: a prospectus. July 2009. [Cited: 2010-01-22]. Available at: http://www.rda-jsc.org/docs/5rda-prospectusrev7.pdf

[10] RDA/ONIX framework for resource categorization. 2006. [Cited: 2010-01-22]. Available at: http://www.rda-jsc.org/docs/5chair10.pdf

[11] Vocabulary Mapping Framework. [Cited: 2010-01-22]. Available at: http://cdlr.strath.ac.uk/VMF/

[12] W3C. RDF Working Group. Resource description framework (RDF). [Cited: 2010-01-22]. Available at: http://www.w3.org/RDF/

[13] W3C Semantic Web activity. [Cited 2010-01-22]. Available at: http://www.w3.org/2001/sw/

[14] RDA Resource description & access: constituency review. 2008. [Cited 2010-01-22]. Available at: http://www.rdaonline.org/constituencyreview/