To: RDA Steering Committee

From: Gordon Dunsire, Chair, RSC

Subject: RDA Toolkit Glossary and RDA Reference

This document describes the background to significant changes made to RDA Elements, vocabulary encoding schemes, and instructions during recent development of the infrastructure used to manage the content of RDA in RDA Toolkit and the RDA Registry.

The changes made to RDA Toolkit do not affect the application of instructions or use of the glossary.

This document is primarily intended for users of RDA Toolkit who are interested in the context of the changes and the future development of RDA.

The document also gives a general explanation of how the "data elements" of the RDA "package of data elements, guidelines, and instructions"[1] are stored and distributed using the infrastructure. The infrastructure for managing the guidelines and instructions for RDA Toolkit will undergo further development in the near future.

RDA content serves two different audiences: the cataloguers and metadata editors who use RDA Toolkit, and the application developers and linked data communities who use the RDA Registry.

The "data elements" content is not only used in RDA Toolkit. Other components and products using this content may be unfamiliar to Toolkit users, although "RDA Reference", "Open Metadata Registry", and "RDA Registry" have been mentioned in recent RDA Steering Committee announcements and RDA documentation. The use of non-Toolkit terminology in describing the infrastructure is therefore unavoidable, but is only used where necessary, and in conjunction with a non-technical explanation.

RDA Reference data

RDA Reference data are the "data elements" in the RDA package.

---

[1] Making RDA Relevant to the Broader Community : a statement from the RDA Board. 24 June 2016. Available at: http://www.rda-rsc.org/Boardstatement2016june

RDA Reference comprises the RDA Elements, their definitions, and any related scope notes, and all vocabulary encoding scheme terms and definitions presented in the RDA instructions. RDA Reference includes translations from RDA Toolkit and external sources.

RDA Elements includes entities, elements, and relationship designators.

RDA Reference data are maintained for linked data communities in the Open Metadata Registry (OMR) and made available as the "RDA Vocabularies" in the RDA Registry and elsewhere. The data are machine-readable.

Essentially, the single source of RDA Reference data is re-used in RDA Toolkit and in the RDA Registry. Further information is given in the **Appendix: RDA Reference data maintenance and flow**.

The RDA Development Team scheduled the production of the **August 2016** release of RDA Toolkit for testing the infrastructure for generating RDA Toolkit data from the RDA Reference data held in the OMR.

RDA Toolkit Glossary

The RDA Toolkit Glossary was selected as the data for testing:

- It has a simple two column layout.
- The Glossary definition is a simple concatenation of RDA Reference definition and scope note.
- It is self-contained. The instructions link to the Glossary, but the Glossary does not link to anything except itself, in the form of cross-referencing.
- The cross-references are simple See and See also references.
- The Glossary entries have internal mark-up for linking and re-use within the RDA Toolkit content management system.

The RSC was asked to finish the development of the RDA vocabulary encoding schemes for publishing as RDF value vocabularies in the OMR. This was one of the activities agreed at the Data Model Meeting at the British Library in 2007[2]:

- disclosure of RDA Value Vocabularies using RDF/RDFS/SKOS.

---

[2] http://www.bl.uk/bibliographic/meeting.html

The RSC was also asked to consider adding the relationship designator labels and definitions (formerly only available in hierarchical order in appendices I-K) to the Glossary. Relationship designators are stored in the OMR along with entities and elements, and it is simpler to incorporate all of the data in the Glossary than to filter some of them out to be marked-up by a different process. This lengthens the Glossary significantly, but also provides the designators with an alphabetical listing to complement the hierarchical listing in the appendices, and a ready-reference source for the controlled terminology used in the RDA instructions.

The RSC used the opportunity to carry out actions to develop the vocabulary encoding schemes agreed during the November 2015 RSC meeting and in subsequent discussion. These include:

- Removal of plural forms of a label if the singular form is present.
- Retention of plural form only where necessary.
- Replacement of element names with vocabulary names in definitions of terms in vocabulary encoding schemes.
- Gathering of "meta-terms" used in the instructions to form a separate vocabulary. These are terms for the processes and products of cataloguing, such as "carrier" and "core", that are not specific RDA elements.
- Renaming the *Scale* element to *Scale designation*.
- Merger of the vocabularies for *Applied material*, *Base material*, and *Mount*.
- Consolidation of the vocabularies for *Extent …*, *Generation …*, *Layout …*, and *Production method …*
- Consolidation of definitions or separate labels for double-entry terms.

The RSC Chair and RSC Secretary also proposed actions for clarifying the semantics and syntax of RDA Reference to get editorial benefit from the new infrastructure:

- Clear separation of definition and scope note.
- Removal of instructional information from definitions, scope notes, and cross-references.
- Removal of vocabulary encoding scheme terms from the definitions and scope notes of associated elements.
- Removal of other redundant or duplicate wording from definitions and scope notes of elements and values.
- Simplification of punctuation used in definitions and scope notes.
- Minimizing use of special characters, and applying character encoding where necessary.

Outcomes

Minor changes were made to the punctuation and wording of many RDA Reference definitions and scope notes. These are not listed.

Revisions to RDA Reference and related instructions included in the August 2016 release of RDA Toolkit are listed in RSC/Sec/4.

Declination, Right ascension, Latitude, Longitude

Separate properties for the pairs of sub-elements *Declination* and *Right ascension*, and *Latitude* and *Longitude*, were added to the *Work* element set.

Further action: Draft revisions of the instructions will be discussed by the RSC in due course.

Digital representation of cartographic content

There were two distinct sets of vocabulary terms given in the instructions at RDA 3.19.8.3.

For a) data type, the three terms listed were introduced as "i.e.", so the list was presumably exhaustive, and constituted a vocabulary encoding scheme.

For b) object type, the terms listed were introduced as "e.g.", and the associate examples used additional terms such as "network chains". This implied that the terms that can be used are open, and do not constitute a vocabulary encoding scheme.

For c) number of objects … there was no vocabulary encoding scheme because the expected value is a number.

A review of the element and its instructions showed that the number of objects component is a numerical phrase, not a simple number. It can take the form of 1 (x), 2 (x,y), or 3 (x, y, z) co-ordinate numbers, and can be specified as numeric data.

The existing OMR vocabulary was deprecated, and a new vocabulary was created for a new RDA element for *Cartographic data type* containing the three terms listed at a). The parent aggregating element is *Digital representation of cartographic content*.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Emulsion on microfilm and microfiche

The term "mixed" was not defined, and it is more appropriate to record it as *Details of emulsion on microfilm and microfiche*. The term was removed from the vocabulary at RDA 3.7.2.3.

[The same principle applied to the term "mixed materials" in the *Applied material* vocabulary. The term was moved from the consolidated vocabulary to an instruction at RDA 3.7.1.4.

The inclusion of "mixed" in the vocabulary encoding scheme for *Aspect ratio* was also reviewed.]

The consolidated *Material* vocabulary was able to accommodate this vocabulary; an emulsion is an applied material. The current labels of the vocabulary assume "emulsion"; this was made explicit by adding the term to the label, for example "silver halide emulsion".

The vocabulary for *Emulsion on microfilm and microfiche* was merged into the consolidated vocabulary for *Material*.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Encoding format element

The vocabulary for *Encoding format* was reviewed, taking into account long-standing issues such as overlap in category groupings, the ad hoc mix of terms based on file formats, abbreviations, and trademarks, the lack of definitions, and issues in the maintenance of currency.

The unpublished vocabulary for the *Encoding format* element was deprecated in the OMR.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Extent

The *Extent …* vocabularies were reviewed for consolidation.

The semantics of the specific categories of *Extent …* terms have been recorded as scope notes.

The instructions for Extent were to use the *Carrier type* vocabulary in general, with specific terms given in sub-vocabularies. The *Carrier type* vocabulary was therefore duplicated in the OMR vocabulary for *Extent* when it was first created. The duplicate terms needed to be removed; this left the Extent vocabulary with a large number of deprecations.

The *Extent of notated music* element uses the vocabulary for *Format of notated music*.

The separate sub-vocabularies for *Extent of cartographic images, Extent of still image, Extent of text,* and *Extent of three-dimensional form* were merged to form a new *Carrier extent unit* vocabulary. The existing *Extent* vocabulary was deprecated.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Further action: Proposals that develop the instructions and elements to further meet the goals of 6JSC/ALA/Discussion/5 will be prepared in due course.

Generation

The RDA element *Generation* has the definition "The relationship between an original carrier and the carrier of a reproduction made from the original". It appears to be a relationship element, but it has a vocabulary encoding scheme (proposed for consolidation) that suggests it is an attribute element. The trend noted by the FRBR-LRM is towards relationships and away from attributes. The element required consistent treatment as a relationship or as an attribute with an associated vocabulary encoding scheme.

The separate vocabularies were merged into a new general vocabulary for the *Generation* element, and the element sub-types for audio recordings, digital resources, microforms, motion picture films, and videotapes were deprecated to reduce the occurrence of instructions based on specific content or carrier categories.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Further action: The RDA Development Team and RSC Technical Working Group will investigate the replacement of the vocabulary encoding scheme for *Generation* with a set of relationship designators that refine a *Generation* relationship element.

Layout

A consolidation of the *Layout …* terms was created as a vocabulary encoding scheme.

The information about the specific categories was recorded as scope notes. This does not lose detail because the categories themselves are not identified as entities or sub-entities in RDA.

The separate sub-vocabularies for *Layout of cartographic images* and *Layout of tactile musical notation* were merged into the general vocabulary for the *Layout* element.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Medium of performance element

The sub-vocabularies *RDA Choruses*; *RDA Instrumental Music for Orchestra, String Orchestra, or Band*; *RDA Solo Voices*; and *RDA Standard Combinations of Instruments*, were all in unpublished status with no definitions, and contained terms that are found in external general vocabularies for *Medium of performance*.

The vocabulary for *Medium of performance* and related sub-vocabularies were deprecated in the OMR.

Further action: RSC/MusicWG/3, to be discussed at the November 2016 RSC meeting, proposes the removal of the lists of terms from the instructions for the *Medium of performance* element.

Place entity

The RDA elements *Name of the place, Preferred name for the place*, and *Variant name for the place* did not have representations in the OMR or RDA Registry because they lacked a Place entity. Place is anticipated to become an RDA entity as an outcome of the FRBR-LRM and the work of the RSC Places Working Group

It is more efficient if the general development of the RDA elements and instructions for Place and LRM Nomen accommodates these "name of place" elements in a separate RDA Registry element set consistent with the current RDA entities.

A placeholder entity was added for *Place* in the RDA Class element set, and a new RDA element set for *Place* was created with attribute properties representing *Name of the place, Preferred name for the place*, and *Variant name for the place*.

There was no impact on the RDA instructions.

Production method

A consolidation of the *Production method …* terms was created as a vocabulary encoding scheme.

The separate sub-vocabularies for *Production method for manuscript* and *Production method for tactile resource* were merged into the general vocabulary for the *Production method* element, and the element sub-types *Production method for manuscript* and *Production method for tactile resource*, and their corresponding *Details of production method for manuscript* and *Details of production method for tactile resource* were deprecated to reduce the occurrence of specific elements and instructions based on content or carrier categories.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

Scale and ratio elements

A review of the impact on the instructions of renaming the *Scale* element to *Scale designation* identified the need to retain a separate *Scale* element to record the scale as a ratio. An element for *Scale designation* remained a requirement for the terms in the vocabulary encoding scheme.

The vocabulary was renamed as "Scale designation", and the definitions amended accordingly.

RDA specifies numerical values for various forms of ratio for the elements:

- Aspect ratio
- Reduction ratio
- Scale

The treatment of these elements in RDA was inconsistent and incompatible with the requirements of linked data.

Associated revisions were made to the instructions for all these elements, as detailed in RSC/Sec/4.

Aspect ratio

The instructions at RDA 7.19 referred to two distinct sets of values for the *Aspect ratio* element.

Aspect ratio is defined as "The ratio of the width to the height of a moving image", but the main instruction at RDA 7.19.1.3 was associated with a vocabulary encoding scheme. The terms in the vocabulary have definitions covering a range of ratios.

The instructions then said "In addition, record the numerical ratio in standard format with a denominator of 1".

There was no other element for this data, and the examples showed it being added with standard punctuation to the terms from the vocabulary encoding scheme.

There was an incompatibility: the *Aspect ratio* element was associated with a vocabulary encoding scheme, but the terms could be qualified by a specific ratio.

The *Aspect ratio* element was restricted to a value of a numerical ratio. A new element was added for *Aspect ratio designation* that has the range of the vocabulary encoding scheme.

The instructions were split so that the "In addition …" instruction became the main instruction for the *Aspect ratio* element, with the former main instructions reassigned to the new *Aspect ratio designation* element with references from the instructions for *Aspect ratio*.

Reduction ratio

The instructions at RDA 3.15 referred to two distinct sets of values for the *Reduction ratio* element.

Reduction ratio is defined as "the size of a micro-image in relation to the original from which it was produced", but the main instruction at RDA 3.15.1.3 was associated with a vocabulary encoding scheme. The terms in the vocabulary encoding scheme have definitions covering a range of ratios.

The instructions then said "If the reduction ratio is ultra high (i.e., greater than 90×), specify the ratio, in parentheses, following ultra high reduction". The ratio is a numerical value .

There was no other element for this data, and the examples showed it being added with the prescribed punctuation to the term from the vocabulary encoding scheme.

There was an incompatibility: the *Reduction ratio* element was associated with a vocabulary encoding scheme, but one of the terms could be qualified by a ratio.

The *Reduction ratio* element was restricted to a numerical ratio, and not confined it to values greater than 150x1. A new element for *Reduction ratio designation* was added for the vocabulary encoding scheme.

The instructions were split so that the conditional instructions became the unconditional instructions for the *Reduction ratio* element, with the main instructions reassigned to the new *Reduction ratio designation* element with references from the instructions for *Reduction ratio*.

Scale

The instructions at RDA 7.25 referred to two distinct sets of values for the *Scale* element.

Scale is defined as "the ratio of the dimensions of an image or three-dimensional form contained or embodied in a resource to the dimensions of the thing it represents". The main instruction at RDA 7.25.1.3 was compatible with this definition and the value of the element was to be recorded "as a representative fraction expressed as a ratio".

However, the main instruction was followed by a number of alternative and conditional instructions that allowed terms that were incompatible with the element definition, such as "Not drawn to scale". The terms were prescribed, and were used as the basis of a vocabulary encoding scheme for the element. The RSC agreed to change name of the vocabulary to "Scale designation" to improve consistency and coherency in the term definitions.

There was a fundamental incompatibility: the *Scale* element could be associated with either a ratio or a vocabulary encoding scheme, but not both.

The *Scale* element has element sub-types *Horizontal scale of cartographic content* and *Vertical scale of cartographic content*. Both are associated with a ratio, not a vocabulary encoding scheme.

A new element was added for *Scale designation*, associated with the vocabulary encoding scheme.

The instructions were split so that the *Scale* element remains largely unchanged. The alternative and conditional instructions became the instructions for *Scale designation*, with references from the instructions for *Scale*.

Sound content

The vocabulary for *Sound content* in the OMR did not appear as a vocabulary encoding scheme in the Toolkit instructions.

The vocabulary definition and term definitions did not completely match the meaning of the instructions.

The *Sound content* vocabulary was treated consistently with other vocabulary encoding schemes in the instructions, and specifically with those associated with the *Colour content* and *Illustrative content* elements.

Associated revisions were made to the instructions, as detailed in RSC/Sec/4.

The definitions refer to "primary" content, but this is not defined, and there is no specific guidance in the RDA instructions. A similar issue occurs with the use of "main or predominant work or content" in RDA 2.1.2.2 and RDA 2.1.2.3.

Further action: The RSC RDA/ONIX Framework Working Group, the RSC Chair and Secretary, and/or one or more RSC communities will review the instructions and consider adding guidance on the concepts of primary, main, and predominant content, consistent with the RDA/ONIX Framework.

Surname and Given name

The RDA Development Team agreed to add properties for sub-elements for *Surname* and *Given name* to the OMR, associated with the *Person* entity.

The RSC Technical Working Group has identified appropriate places in the instructions to add the sub-elements, and submitted as RSC/TechnicalWG/2 for discussion at the November 2016 meeting of the RSC.

The existing Glossary entry for "surname" was updated as a sub-element.

RDA Terms

A separate vocabulary encoding scheme for "RDA Terms" was set-up in the OMR.[3]

The vocabulary contains terms from the pre-August 2016 Glossary that are not the labels of elements or designators, and not terms from value vocabularies for specific elements. It also includes additional terms that are used in the instructions to refer to things not otherwise specified by a description, context, etc.

The vocabulary can be viewed in the OMR as a self-contained vocabulary with internal consistency and scope. Its terms are used to describe the context and environment of RDA data: the "meta-terms" discussed at the 2015 RSC meeting.

Formalizing the vocabulary as a separate vocabulary encoding scheme benefits translation teams and international communities, creators of derivative products and training materials, and linked data communities.

Related resource

The term "related resource" and definition were added to the RDA Terms vocabulary, with "related entity" added as a Use for/See from reference.

Duplicate terms

It is inevitable that the same term is used in different vocabulary encoding schemes with different definitions. The RSC policy is to disambiguate terms within the same vocabulary to allow them to be represented as linked data. There is no requirement to disambiguate terms in different vocabularies.

The full new Glossary reveals several sets of duplicate terms. These are under investigation because they appear to arise in differing circumstances, and a single resolution may not be appropriate.

The RSC Translations Working Group has requested that English synonyms be treated as separate terms, as they are not necessarily synonyms in other languages.

Further action: Proposals will be made in due course.

---

[3] http://metadataregistry.org/vocabulary/show/id/420.html

Benefits

The infrastructure developments create a number of benefits:

- The data do not require further human mediation outside of the OMR. Updated data are automatically incorporated in RDA Toolkit, the RDA Registry, and elsewhere without the need for additional editing.
- The articulation between the OMR and RDA Toolkit and the RDA Registry allows greater flexibility in the management of releases. RDA Reference can be continuously updated without constraining the schedule of RDA Toolkit releases.
- The same data can be re-used for multiple purposes in RDA Toolkit, training materials, and other RDA applications.
- The data are available under an open license.
- The data can be easily downloaded in a common format for spreadsheet processing.
- The data are well-formed for the Semantic Web and linked data applications.

Appendix: RDA Reference data maintenance and flow

RDA Reference data

RDA Reference includes all RDA Elements, their definitions and any related scope notes, and all vocabulary encoding scheme terms and definitions.

RDA Elements include entities, attributes, relationships, and relationship designators described in RDF.

For linked data, RDA Elements are represented as "element sets" and RDA vocabulary encodings schemes are represented as "value vocabularies".

Open Metadata Registry

RDA Reference data are stored in the Open Metadata Registry[4] (OMR) in Resource Description Framework (RDF) linked data format.

The data are exported in multiple RDF formats, or serializations, to a GitHub open file repository with version control named RDA Vocabularies[5]. Each export is a cumulative snapshot of the OMR data. Specific snapshots are designated as releases and given a semantic version number.

RDA Vocabularies (GitHub)

GitHub makes all releases and snapshots available for download, so applications can freeze or roll-back the data.

GitHub also makes all release and change notes available, along with automatic change comparison documentation.

GitHub offers an open active issue management system, used by RDA developers to raise queries, notify errors, and make suggestions for improvement.

---

[4] http://metadataregistry.org/

[5] https://github.com/RDARegistry/RDA-Vocabularies

RDA Registry

The RDA Registry[6] provides links to download the individual element sets and value vocabularies in the current release of RDA Vocabularies.

The Registry offers additional documentation and tools to RDA developers. These include a Technical Guide, examples of RDA linked data, and machine-actionable maps for transforming RDA data for use by non-RDA applications.

The RDA Registry also provides updated data to the RDA Vocabulary Server whenever a new release is published.

RDA Vocabulary Server

The Vocabulary Server provides a de-referencing service for RDA URIs used in linked data, serving RDF serializations of the vocabularies in response to requests from applications.

The Server runs several services that provide language-specific extractions from the 'all-languages' vocabulary serializations available from the RDA Registry, based on specific requests for one or more languages.

Services also extract individual Elements and Concepts/Values from the broader element sets and value vocabularies in order to provide RDF representations of specific URIs.

RDA Toolkit

The Glossary in RDA Toolkit[7] is based entirely on the RDA Reference data.

The RDF/XML serialization available in the current release of RDA Vocabularies is processed with a transform script to extract the relevant data and format it in XML for the Toolkit Content Management System.

---

[6] http://www.rdaregistry.info/

[7] http://access.rdatoolkit.org/

RIMMF

The RDA data editor RIMMF3[8] (RDA in Many Metadata Formats) uses the RDA Vocabularies data to keep up-to-date with RDA Toolkit.

RIMMF3 provides active links from its RDA data input forms to the instructions in RDA Toolkit.

Data maintenance

The OMR allows data to be updated under password control through:

1. a standard online administrator interface
2. a batch file import facility within the administrator interface

The OMR provides a vocabulary administrator interface which allows an authorized editor to update individual Elements and Concepts/Values, including additions, amendments, and deletions.

The administrator interface includes a data export facility for each element set and value vocabulary. The facility is open and does not require a password. The data are exported as a comma-separated variable (csv) file, a general spreadsheet format.

The file is amended by authorized members of the RSC using a spreadsheet application.

The file is prepared for re-import to the OMR prior to a scheduled release of RDA Toolkit.

The administrator interface has a data import facility for each element set and value vocabulary. The facility requires both a password and authorization. The data are imported as a csv file.

The OMR rejects malformed data; otherwise processing takes a few seconds.

The batch update procedure is also used for translations of RDA Reference.

Data flow

All updates take place in the OMR. The OMR data are openly available.

---

[8] http://www.marcofquality.com/wiki/rimmf3/

From there, data flow to the RDA Vocabularies file repository as snapshots. The RDA Vocabularies files are openly available.

The three primary users of the RDA Vocabularies files are:

- the Glossary in RDA Toolkit
- the RDA Registry
- the RDA Vocabulary Server
- RIMMF3