RDA data capture and storage

Gordon Dunsire
Chair, RDA Steering Committee

Presented to Committee on Cataloging: Description and Access II (CC:DA) - ALCTS CaMMS

Overview

• RDA for data management: a continuous process of development
• From here to the future, and what is on the way
RDA data

RDA is a package of data elements, guidelines, and instructions for creating library and cultural heritage resource metadata that are well-formed according to international models for user-focussed linked data applications.

RDA Toolkit provides the user-focussed elements, guidelines, and instructions.

RDA Registry provides the infrastructure for well-formed, linked, RDA data applications.
Recording relationships in RDA

RDA offers a choice of techniques for recording relationships between entities. The number of options varies depending on the type of entity:

- 4 techniques for relationships between works, expressions, manifestations, and items
- 3 techniques for primary relationships between works, expressions, manifestations, and items
- 2 techniques for relationships between persons, families, and corporate bodies.
The 4-fold path

Recording Relationships between Works, Expressions, Manifestations, and Items

Record the relationship between a work, expression, manifestation, or item and a related work, expression, manifestation, or item by using one or more of these techniques, as applicable:

a) identifier for the related work, expression, manifestation, or item (see 24.4.1 RDA)

b) authorized access point representing the related work or expression (see 24.4.2 RDA)

and/or

c) description of the related work, expression, manifestation, or item (see 24.4.3 RDA).

Record an appropriate relationship designating the relationship (see 24.5 RDA).

Description of the Related Work, Expression, Manifestation, or Item

Provide a description of the related work, expression, manifestation, or item by using either a structured or an unstructured description, as appropriate:
The 3-fold path

Techniques Used to Record Primary Relationships

Record primary relationships by using one or more of these techniques, as applicable:

a) identifier for the work, expression, manifestation, or item (see 17.4.2.1 RDA)
b) authorized access point representing the work or expression (see 17.4.2.2 RDA)
c) composite description (see 17.4.2.3 RDA).

Composite Description

Provide a composite description that combines one or more elements identifying the work and/or expression embodied in a manifestation with the description of that manifestation.
The 2-fold path

Recording Relationships between Persons, Families, and Corporate Bodies

Record the relationship between a person, family, or corporate body, and a related person, family, or corporate body by using one or both of these techniques:

a) identifier (see 29.4.1 RDA)

and/or

b) authorized access point (see 29.4.2 RDA).

Record an appropriate relationship designator to specify the nature of the relationship (see 29.5 RDA).
New FRBR-LRM entities

Place

Collective agent

Timespan

Nomen

Encompasses:
- Identifier
- AAP
- VAP
- Structured description
- Transcribed title, etc.

What techniques will apply to new RDA entities?

a: Identifier
b: AAP
c1: Structured?
c2: Unstructured?
Structured description

A full or partial description of the related resource using the same data that would be recorded in RDA elements for a description of that related resource presented in an order specified by a recognized display standard.

[Example: ISBD display pattern]
Title proper: other title information / statement of responsibility

How full?
A complete ISBD record with all of the data?

RDA: an introduction / by J. Smith

Title proper: “RDA”
Other title information: “an introduction”
Statement of responsibility: “by J. Smith”
Database implementation scenarios

0: Linked data

1: Relational or object database

2: Bibliographic and authority records

3: Flat-file

Fully linked (global)

Fully linked (local)

AAP/Identifier linked

Not linked
Techniques for obtaining data

Categorization of elements?

<table>
<thead>
<tr>
<th>Recorded elements</th>
<th>Sources</th>
<th>any (authoritative, recognized, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tasks</td>
<td>all (Find, Identify, Select, Obtain, Explore)</td>
</tr>
<tr>
<td></td>
<td>Entities</td>
<td>all</td>
</tr>
</tbody>
</table>

**is form of?**

<table>
<thead>
<tr>
<th>Transcribed elements</th>
<th>Sources</th>
<th>Manifestation (Item in hand)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tasks</td>
<td>Identify</td>
<td></td>
</tr>
<tr>
<td>Entities</td>
<td>Manifestation</td>
<td></td>
</tr>
</tbody>
</table>
What you see is what you get?

EDINBURGH:
Printed for the Author,
And sold at his Music-shop at the Harp and Hautboy,
M D C C L X I I.

EDINBURGH:
Printed for the Author,
And fold at his Music-shop at the Harp and Hautboy,
M D C C L X I J.
EDINBURGH: PRINTED FOR THE AUTHOR, And fold at his Music-shop at the Harp and Hautboy. MDCCLXII.
What you see is what you get?

EDINBURGH:
Printed for the Author,
And sold at his Music-shop at the Harp and Hautboy.

Recording Publication Statements [2014/04]
Record a publication statement or statements for a published resource.
Transcribe places of publication and publishers’ names as they appear on the source of information (see 1.7 RDA).

Record dates of publication as they appear on the source of information. Apply the general guidelines on transcription for words that are not numbers (see 1.7 RDA). Apply the general guidelines on numbers expressed as numerals or as words (see 1.8 RDA).

Edinburgh: Printed for the author, and sold at his music-shop at the Harp and Hautboy, 1762
Transcription for “Identify” task

Digital image is:
• Quickest and cheapest
• Easiest for user with item/image in hand

App + Camera + Touch-screen + Image matching software service

21st century! Web of machines!

Transcription string for item citation:
• User must know transcription rules
• Is OCR good enough?
  • Feedback capture = Crowdsourcing
Recording for user tasks

If data is not transcribed, it is recorded

Recording excludes (more or less):
• Typos
• Deliberate errors
• Fictitious entities

Some of the recorded data support the Find, Identify, Select, Obtain, or Explore user tasks

How can the data best be accommodated in RDA?
N-fold path

1. Unstructured string.
   1. Exact transcription (OCR or born digital).
   2. Transcription using the RDA guidelines.
   3. Data recorded from another source.

2. Structured string of delimited sub-values.
   1. Access point.
   2. Structured description.

3. Structured string.
   1. Identifier

4. URI of entity, including Nomen.
   1. URI/URL of digital image.
The path starts here

Paths are available for describing related entities

The same paths describe the entity in focus

ID: xox-oxox

URI

Xox oxox oxo xoxo x oxo

Xox oxox: oxo xoxo.
/ xoxo. - x oxo xo xoxo. -
Xo xox oxox; oxo xo oxo.
Developing Toolkit guidance and instructions

Methods of recording RDA data

- General guidance on techniques (4-fold path)
- General instruction sets for specific entities and element categories (attribute, relationship)
- Specific instructions for specific elements
Developing RDA Registry for applications

Elements for storage of RDA (linked) data

Element domain = parent Entity (constrained)

Element range = type of path (not currently specified)

Sub-properties (sub-types) of each element have 2 types of range to accommodate 4-fold path: literal and object

<table>
<thead>
<tr>
<th>Element range</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literal</td>
<td>Unstructured</td>
</tr>
<tr>
<td>Literal (associated with construction encoding scheme)</td>
<td>Structured/AP/Identifier</td>
</tr>
<tr>
<td>Object</td>
<td>URI</td>
</tr>
</tbody>
</table>
New entities

Place

Timespan

Agent

Collective agent

Nomen

New high-level relationship elements

New relationship designators (cross-entity)
Toolkit Entity views

Proposed development to provide a focus for each RDA entity and its elements

Replaces out-of-date Element set views

Acts as a ready-reference to all elements and instructions associated with the entity
Entity view: a dictionary/reference for RDA

Possible layout

- Entity definition, etc.
- Entity elements
  - Common elements
  - Specific elements

With n-fold path:
- literal range + associated structure
- object (Entity) range

Guidance and instructions
Re-organizing the Toolkit

• Appendices and tabs
• Vocabulary Encoding Schemes
  • Sharing, extending, linking (RDA and other communities)
• RDA Reference (entities, elements, terms)
• Glossary
  • How far beyond entities, elements, and vocabulary terms?
• Translations
• Policy statements and application profiles
• Entity views, Relationship designators, etc.
Some issues

• Needs of international, cultural heritage, and linked data communities

• Primary (WEMI) vs Secondary (PFC ...) entities
  • Reciprocal relationships/links/designators
  • Elements other than those for access points?

• Structure in descriptions
  • How much specification?
    • International communities use different structures

• Nomen control (a kind of authority control?)

• Relationship designators
  • Cross-entity, and many more (labels, definitions?)
Thank you!

- rscchair@rdatoolkit.org
- http://access.rdatoolkit.org/
- http://www.rdaregistry.info/
- http://www.rda-rsc.org/