

RDA metadata implementation scenarios

Gordon Dunsire, RSC Technical Team Liaison Officer, 4 October 2019

Abstract

This paper discusses enhancements to the original RDA database implementation scenarios that reflect the structure and content of the new Toolkit, makes recommendations on updating the scenarios, and proposes a draft text and layout for incorporating the revised scenarios in the new Toolkit.

Background

The original RDA Toolkit is associated with a set of database implementation scenarios.¹ The last revision was in 2009.

The Library of Congress response to the beta Toolkit recommended that implementation scenarios be incorporated in the new Toolkit.

Original scenarios

There are three database implementation scenarios associated with the original Toolkit:

- Scenario 1: Relational / object-oriented database structure
- Scenario 2: Linked bibliographic and authority records
- Scenario 3: 'Flat file' database structure (no links)

The document gives a brief overview and description of the scenarios, followed by a data structure diagram for each scenario.

Some of the content is inconsistent with the new Toolkit:

- References to FRBR and FRAD.
- Use of the term 'record'.
- Out-of-date labels for specific elements.

Recommendation 1: Update the content of the implementation scenarios to be consistent with the new Toolkit.

Enhancements

The bases of the original scenarios are still valid, although some detail needs to be updated.

Recommendation 2: Retain the original scenarios.

Scenario for linked open data

There is no scenario for implementing RDA metadata as linked open data for use in Semantic Web applications.

¹ RDA Database Implementation Scenarios. Available at: <http://www.rda-jsc.org/archivedsite/docs/5editor2rev.pdf>

The IFLA Library Reference Model “is developed very much with semantic web technologies in mind, and it is hoped that in the future, an update of this document will provide RDF examples as well.”²

The new Toolkit introduces the IRI recording method to support linked open data and Semantic Web applications.

Recommendation 3: Add a scenario for linked open data.

Scenarios and recording methods

There is an alignment between the four recording methods and four implementation scenarios.

The alignment is fuzzy, and is indicated by the kind of value used to link distinct datasets within the scenario. Many local applications will be a hybrid of two or more scenarios, with a mix of linking methods.

Dataset link	Recording method	Scenario	Example
name/title of entity	Unstructured description	Flat file	Printed bibliography
access point for entity	Structured description	Bibliographic/Authority	MARC 21 catalogue
Identifier for entity	Identifier	Relational/Object-oriented	MS Access database
IRI of entity	IRI	Linked open data	RDF graph

The alignment exposes the fundamental difference between ‘string’ and ‘thing’ scenarios, and the relative differences in degree of human and automated intermediation associated with the original ‘string’ scenarios.

The original flat file scenario uses access points as headings for the unlinked datasets for bibliographic and non-bibliographic entities, but this can be “dumbed-down” to preferred names or titles without loss of functionality.

The original relational/object-oriented scenario shows preferred and variant names and titles associated with entity datasets, but the actual linking method is primary and secondary keys (table/row number) or object identifier.

Recommendation 4: Clarify the alignment between scenarios, dataset linking methods, and recording methods.

Recommendation 5: Notate the scenarios as A, B, C, and D to avoid confusion with the original 1, 2, and 3:

Scenario A: Linked open data

Scenario B: Relational or object-oriented data

Scenario C: Bibliographic/authority data

Scenario D: Flat file data

² IFLA Library Reference Model. 4.2.4. Available at: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

Scenarios and application profiles

An application profile is a specification of one or more metadata description sets based on one or more entities. A profile determines several of the characteristics associated with implementation scenarios:

- The entities to be described.
- The elements to use in a description of an entity.
- The preferred recording method to use for an element.

For example, Scenario A (Linked open data) selects all entities, does not require any appellation elements for non-WEMI entities, and prefers the IRI recording method for all elements where applicable.

A scenario also determines the clustering of resource entities (WEMI) in basic metadata description sets. For example, Scenario C (Bibliographic/authority) treats WEM, with or without I, as a single description set.

In turn, a scenario may determine the selection of elements required for a Minimum description of a resource entity (https://beta.rdatoolkit.org/en-US_ala-f4fb013b-b64e-37b1-902c-561d27b1e832). For example, Scenario C (Bibliographic/authority) can use internal or latent identifiers as appellation elements for resource entities, and prefer the shortcut elements “work manifested”/”manifestation of work” to avoid describing an expression.

Basic application profile characteristics of the implementation scenarios are given in Table 1:

Table 1: Application profile characteristics of implementation scenarios

Characteristic / Scenario	A	B	C	D
Related entity value				
name/title				Y
access point			Y	Y
identifier		Y	Y	
IRI	Y			
Resource entities				
WEMI cluster			Y	Y
WEMI separate	Y	Y		

Information on implementation scenarios is best presented in the Toolkit in conjunction with string encoding schemes and other resources for application profiles.

Recommendation 6: Add updated content on implementation scenarios to RDA Toolkit to complement Toolkit resources for application profiles.

Recommendations

Recommendation 1: Update the content of the implementation scenarios to be consistent with the new Toolkit.

Recommendation 2: Retain the original scenarios.

Recommendation 3: Add a scenario for linked open data.

Recommendation 4: Clarify the alignment between scenarios, dataset linking methods, and recording methods.

Recommendation 5: Notate the scenarios as A, B, C, and D to avoid confusion with the original 1, 2, and 3.

Recommendation 6: Add updated content on implementation scenarios to RDA Toolkit to complement Toolkit resources for application profiles.

Draft text

RDA implementation scenarios

RDA data are metadata created using RDA instructions and recorded with RDA entities and elements. All RDA data consist of one or more metadata statements recorded as metadata description sets. A metadata statement has an implicit three-part subject-predicate-object structure: the subject of the statement is the entity being described; the predicate of the statement is the characteristic (attribute or relationship) being recorded; the object of the statement is the value of the characteristic, using any applicable recording method.

The RDA entities and elements are conformant with the IFLA Library Reference Model, and constitute a comprehensive implementation of the model.

The RDA entities, elements, and controlled terminologies are published in the RDA Registry in Resource Description Framework. This provides a consistent and coherent machine-readable ontology for Semantic Web applications, and ensures that descriptions of fine granularity entities using fine granularity elements can be automatically re-used in broader applications. For example, any description of a Person entity is also a description of an Agent entity.

The utility of recording methods in automated applications of RDA data is clarified and extended to all RDA elements where applicable. The alignment of appellation elements (name/title, access point, identifier) with recording methods (unstructured description, structured description, identifier) allows RDA data to accommodate a wide range of implementation factors such as the efficiency of data creation and maintenance, the interoperability of data with other RDA and non-RDA data, and the ease and effectiveness with which users are able to apply the functional objectives that RDA is designed to fulfil.

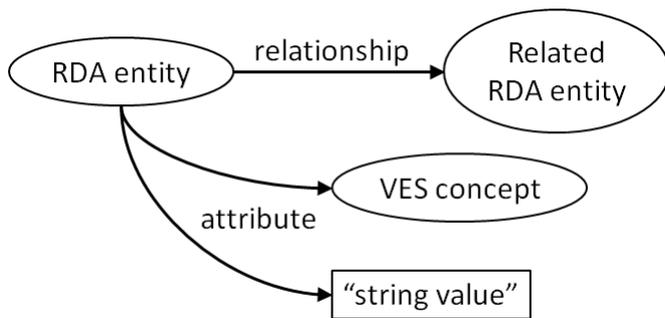
For example, automated transcription of an unstructured description is a very efficient method of creating data, but the resulting string is only effective for keyword searching. As another example, the use of separate descriptions for works and expressions in a relational or object-oriented database structure ensures access not only to all works and expressions associated with a particular person, etc., but to all related works (adaptations, etc.) as well, regardless of whether the name of that person is used to construct the authorized access points representing those works or not.

There are many database structures that are suitable for storing and supplying RDA data. The scenarios described below illustrate the range of potential configurations of RDA data and reflect the distinct structures that are commonly used for library and cultural heritage metadata.

The RDA ontology and guidance on recording methods allows RDA data to be moved or shared between implementation scenarios with a defined level of interoperability. In general, data for any scenario can be re-used, with loss of detail, in a scenario later in the enumerated sequence. For example, Scenario A (Linked open data) can be 'collapsed' into a Scenario D (Flat file data) implementation, and Scenario B (Relation and object-oriented data) can be coarsened into Scenario C (Bibliographic/authority data).

Scenario A: Linked open data

Diagram



Characteristics

Metadata description sets are expressed in Resource Description Framework (RDF) using IRIs taken from the RDA Registry.

Descriptions of the resource entities that comprise a single information resource are recorded in a separate metadata description set for each entity.

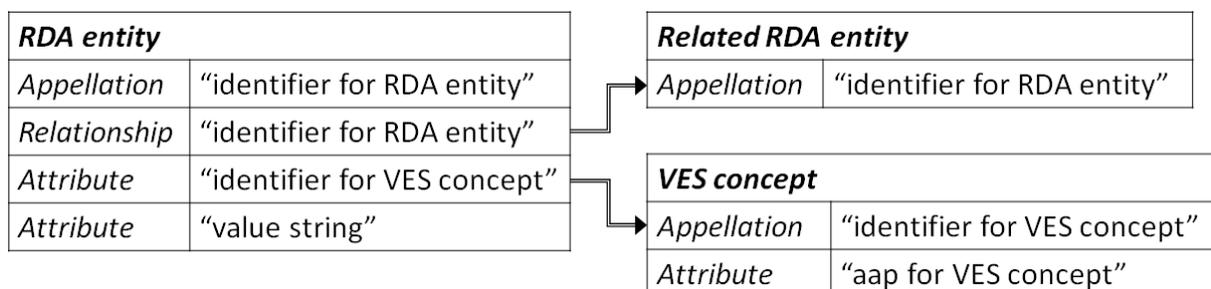
Descriptions of other entities that are associated with an information resource are recorded in a separate metadata description set for each entity.

The IRI recording method is preferred for values taken from a vocabulary encoding scheme.

A metadata description set for an entity is linked to a metadata description set of a related entity using an IRI of the related entity.

Scenario B: Relational or object-oriented data

Diagram



Characteristics

Metadata description sets are expressed in a set of structured data tables and columns that correspond directly to entities and elements taken from the RDA Registry.

Descriptions of the resource elements that comprise a single information resource are recorded in a separate metadata description set for each entity.

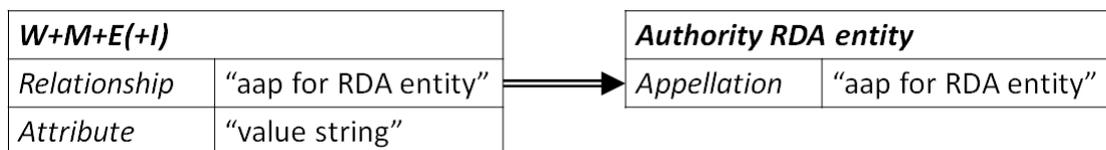
Descriptions of other entities that are associated with an information resource are recorded in a separate metadata description set for each entity.

The identifier recording method is preferred for values taken from a vocabulary encoding scheme.

A metadata description set for an entity is linked to a metadata description set of a related entity using an identifier for the related entity based on primary keys taken from a relational or object-oriented database.

Scenario C: Bibliographic/authority data

Diagram



Characteristics

Metadata description sets are expressed in an encoding schema that aligns with entities and elements taken from the RDA Registry.

Descriptions of the resource elements that comprise a single information resource are recorded in a single integrated metadata description set. The component resource entities are not explicitly identified.

Descriptions of other entities that are associated with an information resource are recorded in a separate metadata description set for each entity.

The structured description recording method is preferred for values taken from a vocabulary encoding scheme.

A metadata description set for an entity is linked to a metadata description set of a related entity using an access point for the related entity.

Scenario D: Flat file data

Diagram

<i>Combination of RDA entities</i>	
<i>Relationship</i>	"value string"
<i>Attribute</i>	"value string"

Characteristics

Metadata description sets are expressed in a layout that uses a set of string encoding schemes to specify entities and elements taken from the RDA Registry.

Descriptions of the resource elements that comprise a single information resource are recorded in a single integrated metadata description set.

Descriptions of other entities that are associated with an information resource are recorded in a separate metadata description set for each entity.

The unstructured description and structured description recording methods are preferred for values taken from a vocabulary encoding scheme.

A metadata description set for an entity is not linked to a metadata description set of a related entity.