

To: Joint Steering Committee for Development of RDA
From: Gordon Dunsire, Chair, JSC Technical Working Group
Subject: RDA models for authority data

Abstract

This paper discusses the models used by RDA to accommodate authority. The paper makes some general recommendations for developing RDA to improve its accommodation of authority data.

Introduction

A task (#6) for the JSC Technical Working Group in 2015 is to investigate the development of RDA models for authority data and prepare a proposals/discussion paper.

Scope

The paper assumes that RDA will comply with the IFLA bibliographic models, consolidated as FRBR-LRM (Library Reference Model).

FRBR Models

RDA elements for authority data are based on the FRAD entities *Controlled Access Point* and *Name*.

RDA models for authority data should align with the consolidated FRBR-LRM model, which is expected to merge the FRAD entities *Controlled Access Point* and *Name* with the FRAD entity *Nomen*.ⁱ

ⁱ Introducing the Library Reference Model. Available at: <http://library.ifla.org/1084/>

The FRSAD model relates the FRSAD entities *Thema* and *Nomen* with the "has appellation" relationship:

- Thema – has appellation – Nomen

This is one half of the basic FRSAD model, which is focussed on subject authority data:

- Work – has as subject - Thema – has appellation – Nomen

FRBR-LRM generalizes the FRSAD model by re-labelling the *Thema* entity as "Res" and defining it as "Any entity in the universe of discourse".

This is conceptually equivalent to the OWL entity *Thing* in the context of RDA. This general "any thing" approach was taken when RDA was developed in 2015 to cover the subject relationship given in the other half of the basic FRSAD model. As a result, the entity *Thema* was not added to RDA; only the *subject relationship* element was added, with sub-types for specific RDA entities:

- Work – has subject relationship – [not specified]

The RDA treatment of the subject relationship remains compatible with FRBR-LRM.

The RDA Registry includes RDF properties for all RDA entities as subject of the Work, for example has subject (family), but RDA only provides designators for Work, Expression, Manifestation, and Item as subject.

Nomen entity

FRBR-LRM essentially says "Things have names":

- Res – has appellation – Nomen

The entity *Nomen* is defined as "Any sign or arrangement of signs by which an entity is known".

This covers sub-types such as identifiers, access points, and names (including titles and other labels):

- Res – has appellation – Authorized Access Point for the entity – sub-type of – Nomen
- Res – has appellation – Identifier for the entity – sub-type of – Nomen
- Res – has appellation – Name or title of the entity – sub-type of – Nomen

The high-level model can be refined for each entity in an application such as RDA.

For example:

- Work – has appellation – AAP for Work – sub-type of – Nomen
- Person – has appellation – AAP for Person – sub-type of – Nomen

Literal form

The *Nomen* entity requires an element for its literal form: the “sign or arrangement of signs”. This element is an attribute with a “literal” data type, intended to be mandatory and non-repeatable; that is, every *Nomen* has one and only one literal string. This approach is used in the SKOSXL namespace, where labels are treated as entities.²

For example:

- AAP for Person1 – has literal form – "Austen, Jane, 1775-1817"
- AAP for Work1 – has literal form – "Austen, Jane, 1775-1817. Emma"

This means that "Austen, Jane, 1775-1817", "Jane Austen (1775-1817)", "Austen, Jane, 1775-", etc. are all different *Nomens*.

Also, one *Nomen* may be the appellation of more than one entity. “Jane Austen” is likely to be a functioning *Nomen* for several different Persons, including the one with other appellations such as “Austen, Jane, 1775-1817”.

Provenance

The *Nomen* entity may have other attributes such as "uses rules", "has date of usage", "has scope of usage", etc. These data record the provenance of the *Nomen*, used in applications for the control of access points and normalization of identifiers.

Some of these attribute elements may also be relationships with other entities. *Rules* is a FRAD entity, and it can have "has version" as an attribute to distinguish, for example, different releases of RDA Toolkit. In the consolidated model, *Rules* is relegated to a type of *Work*, but this will not affect the basic relationship of "uses rules" between *Nomen* and *Work/ Expression* entities used to record versions.

For example:

- AAP for Person1 – uses rules – Rules (RDA April 2015)

Discussion

Refining the appellation relationship

There are use cases for refining the *appellation* relationship to accommodate specific information about the type of *Nomen* to which it relates.

As discussed above, RDA access points and identifiers are types of *Nomen*. Many RDA elements for specific types of entity are also types of *Nomen*. These include the RDA *Manifestation* attributes *title proper* and *identifier for the manifestation*, the RDA *Corporate Body* attributes *name of the corporate body* and *preferred name of the corporate body*, and so on.

² SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant. Available at: <http://www.w3.org/TR/skos-reference/skos-xl.html>

Such types can be accommodated in two ways: by creating element sub-types or entity sub-types:

1. Work1 – has preferred title of the work – Nomen1 – has literal form "Emma"
2. Work1 – has appellation – PreferredNomen (sub-type of Nomen) – has literal form "Emma"

In 1) the element *preferred title of the work* is treated as an element sub-type of the FRBR-LRM element *appellation*.

In linked data terms: <preferredTitleOfWork> rdfs:subPropertyOf <appellation> .

In 2) a new entity *PreferredNomen* is treated as a sub-type of the FRBR-LRM entity *Nomen*.

In linked data terms: <PreferredNomen> rdfs:subClassof <Nomen> .

BIBFRAME is discussing the second approach "in line with BIBFRAME philosophy to denote types by class rather than property wherever practical".³ The advantages stated are not significant for RDA, which intends RDA linked data to be usable in Semantic Web applications as well as applications which ignore semantic data

RDA generally takes the first approach, for example in its treatment of types of relationship as relationship designators. This approach also has the advantages of preserving the existing RDA "appellation" elements and avoiding the proliferation of entities.

Recommendation 1: RDA should represent sub-types of Nomen as element sub-types of the appellation element.

The existing RDA entity-specific "appellation" elements, for example *preferred title of the work*, can be treated as hierarchical relationship designators between an entity and types of *Nomen*.

For example:

- appellation > name or title > preferred name or title > preferred title > preferred title of the work
- appellation > name or title > alternate name or title > alternate title > alternate title of the manifestation > spine title
- appellation > identifier > identifier for the item
- appellation > AAP > AAP for the person

There does not seem to be any advantage in a specific hierarchical arrangement.

³ BIBFRAME title proposal June 12, 2015. Available at: <http://listserv.loc.gov/cgi-bin/wa?A2=ind1506&L=bibframe&T=0&P=5643>

This method of accommodating current RDA elements in the *appellation-Nomen* model essentially specifies a range for the element. Current elements do not have a range because they are treated as attributes, albeit with an expected literal value. A range of *Nomen*, or a sub-type of *Nomen*, must be specified if the element is to be an element sub-type of *appellation*, which itself has the range *Nomen*.

Current: <i>preferred title of the work</i>	
domain	Work
range	-
element subtype of	-
Proposed: <i>preferred title of the work</i>	
domain	Work
range	Nomen
element subtype of	<i>appellation</i>

Recommendation 2: Review and develop appropriate RDA elements for compatibility with the *appellation-Nomen* model by assigning element sub-types and ranges.

Derivation of Nomens

The example *Nomen* with literal form “Austen, Jane, 1775-1817” has internal syntactical structure. It is aggregated from three RDA elements: *preferred name for the person*, *date of birth*, and *date of death*. This is typical of the AAP type of *Nomen*, where RDA specifies the elements that may be aggregated into access points.

It is possible to consider a “family” of *Nomens* that are appellations of the same entity, say *Person*, derived from different aggregations of the same set of elements. For example, “Austen, Jane, 1775-1817”, “Austen, Jane, 1775-”, and “Austen, Jane (1775-1817)” are all based on the same set of three RDA elements. They are generated using the syntax encoding scheme templates:

- *preferred name for the person* + “, “ + *date of birth* + “-“ + *date of death*
- *preferred name for the person* + “, “ + *date of birth*
- *preferred name for the person* + “(“ + *date of birth* + “-“ + *date of death* + “)”

Note that the *Nomen* with literal form “LCCN n 79032879” is not a member of the family because it is based on the element *identifier for the person*.

This approach can be generalized effectively by introducing finer granularity into the *preferred name for the person* element by splitting it into name of the *family name* [of the person] and *given name of the person*. Note that the RDA elements *name of the family* and *preferred name for the family* are unsuitable because they are Family attributes. The values for specific instances will overlap, however.

- Person1 – has family name – Nomen1 – has literal form – "Austen"
- Person1 – has given name – Nomen2 – has literal form – "Jane"

This would allow *preferred name for the person* and some instances of *name of the person* to be aggregated from the same sub-elements and re-use their data:

- *given name of the person* + " " + *family name* (for example, "Jane Austen")
- *family name* + ", " + *given name of the person* (for example, "Austen, Jane")

This provides greater flexibility for the recording and use of data for the *Nomen* of a *Person*, and is considered the norm in databases outside the library domain.

The given/family name substructure is not universally applicable at international level, and is one of the reasons why RDA did not, for example, restore the approach which had been taken by UKMARC. Nonetheless, there are clear benefits to the user tasks of find and navigate (new in LRM) if RDA defines these refinements and seeks advice on the internal structure of names of persons in non-Anglophone communities. There are strong cases for other sub-elements of name, for example the "personal" name used by the DNB. This development encroaches on the application of RDA data, and may be accommodated as an extension/refinement to the RDA Registry.

Recommendation 3: **a** Consider adding the RDA elements *family name* and *given name* as sub-elements of *name of the person*. **b** Develop these for the RDA Registry in any case, to improve interoperability of RDA linked data.

Proliferation of Nomens

Every distinct text string used to name an entity is a different *Nomen*, so many Nomens be associated with an entity.

This does not represent a significant overhead for data-processing. All labels stored as data must have a local identifier such as a database row number, MARC21 Authority tag, MADS node etc. This is equivalent to the *Nomen* entity. For RDA linked data applications, a local identifier (or blank node) is replaced with a URI global identifier for the *Nomen* linked to its literal form, the label.

An instance of an RDA AAP type of *Nomen* can be generated on-the-fly from component values taken from the entity of which it is the *appellation*, so an application may not need to store it for re-use, and therefore not need assign a URI for global re-use.

The main use of recorded *Nomen* data in many applications will be as provenance information, and much of this will be machine-generated from data interfaces such as catalogue input forms.

Relationships between Nomens

The *Nomens* “Jane Austen” and “Austen, Jane” may be explicitly related by a *browse form* or *normalized form* element:

- Person1 – has appellation - Nomen1 – has literal form "Jane Austen"
- Person1 – has appellation – Nomen2 – has literal form "Austen, Jane"
- Nomen1 – has browse form – Nomen2

Similarly, there might be an *acronym form* element

- CorporateBody1 – has preferred name for the corporate body – Nomen1 – has literal form “Jane Austen Society of North America”
- CorporateBody1 – has name of the corporate body – Nomen2 – has literal form “JASNA”
- Nomen1 – has acronym – Nomen2

This will become complicated in situations where traditionally a change in *Nomen* results in a change in the characteristics of the entity, for example if a *Corporate Body* or continuing *Work* changes its preferred *Nomen* in certain ways, it is to be treated as a new instance of the entity. For example:

- CorporateBody2 – has preferred name for the corporate body – Nomen2 – has literal form “JASNA”
- CorporateBody1 – has successor – CorporateBody2

The same *Nomen*, with literal form “JASNA”, is a name of a corporate body that is an acronym of the preferred name of the corporate body. It is also the preferred name of another corporate body that succeeds the corporate body.

Relationships between *Nomen* cover the relationships found in traditional authority control systems between “headings” or access points in authorized and variant forms. These are expected to vary between traditions and applications, and are better specified in application profiles than the RDA instructions. There may be some common relationships that could be specified in RDA Toolkit and the RDA Registry, and these are expected to be identified in the final version of FRBR-LRM.

There are general modelling issues with “preferred” or authoritative forms of name. The SKOS element *prefLabel* accommodates the requirement in KOSs for a single “authority” label for a concept, so the SKOS ontology expects only one preferred label (per language). All other labels are “alternate”. Problems occur if an alternate label becomes preferred, and vice-versa; the label is the same literal form of, and therefore the same, *Nomen*, but it becomes the value of a different element. The problems are aggravated if the “preferred”

semantic is represented by an entity sub-type: a *Nomen* of the hypothetical entity sub-type *PrefLabel* becomes a *Nomen* of the hypothetical entity sub-type *AltLabel*. This can be difficult to manage while maintaining semantic coherency.

The RDA guidelines and instructions may require a different set of assumptions, and the opportunity should be taken to consider what benefits this might bring.

Recommendation 4: **a** Investigate the functionality and utility of “preferred” forms of appellation element sub-types in relation to RDA and application profiles in the context of the appellation-Nomen model. **b** Investigate the utility of relationships between Nomen and how RDA should accommodate them.

Nomens and Linked data

AAPs are usually constructed from two or more elements in a given order of citation. The aim is to produce unambiguous identifiers for the entity.

In a linked data context, this functionality is automatically met by the mandatory use of URIs. The functionality of AAPs, variant access points, and any other type of access point, should therefore be confined to the context of applications, and particularly applications requiring human-readable strings to identify specific instances of an entity.

Authority records and data

RDA does not address “records” as sets of data because it does not mandate any specific set of elements. At best it designates certain elements as “core”, indicating the utility of the attribute and relationship in supporting FRBR, FRAD, and FRSAD user tasks.

What we have called bibliographic records might better be considered as descriptions of manifestations (with bits of work and expression data thrown in to make things messy); what we have called authority records might better be considered descriptions of persons, families, corporate bodies, works, and expressions. This suggests that the nature of the “authority record” needs to change, from a collection of metadata needed to support the control of “headings” — the text strings (or “literals”) that we have traditionally used to represent the entity — to a richer set of attributes that support the identification of the entity.

This blurs the distinctions between the “primary” RDA entities *Work*, *Expression*, *Manifestation*, and *Item*, and the “secondary” associated entities *Person*, *Family*, and *Corporate Body*. The addition of other FRBR-LRM entities such as *Place*, and *Nomen* itself, is likely to generalize the utility of an attribute from component of an access point to entity identification.

Names or *Nomen* are likely to remain the method preferred by humans for identifying entities. Identity management is a requirement of linked data infrastructure, but this is only in the context of machine-actionability. In RDF, all entities, properties, concepts, and

instances must be associated with at least one URI, and that is sufficient for machine linking. URIs are often designed to be, and frequently become, opaque to humans – but there is no universal agreement on what human-readable identifiers should be used. Perhaps the term "authority" should mean reliable and complete enough, in terms of basic facts, to generate flexible sets of access points that can be better tailored to the context and client preferences.

Another requirement of names, titles, and other entity labels is that they should support the LRM user task explore, and the data should be processable for searching, listing, and browsing.

There is still a need for data to support the "name-control" function. Data for an entity should contain all the factual identifying information about the entity; this information should be universally applicable and valid. The information can also be put under the control of the identified entity (if it is an *Agent*, *Person*, *Family*, or *Corporate Body*), but this raises issues of bias, privacy, and trust, as exemplified by recent JSC discussions on the *gender* element.

However, viewed from a broad perspective, the text-strings that we use to represent the entity are not universally applicable; they differ based on rules, agency policies, etc. So, while a single description of an entity might be possible and desirable, a single representation of the entity's "official" or "preferred" name(s) is neither.

One approach is similar to that of VIAF, where separate forms of name are clustered around a specific entity. Each form of name is linked with its own URI to the URI of the cluster using the foaf:foaf element, defined as "the underlying or 'focal' entity associated with some SKOS-described concept".⁴

This is compatible with the FRBR-LRM model, where multiple *Nomen*, each with its own rules, etc., can be related to a single entity, of any type, with the *appellation* element. The major difference is that the entity is directly associated with all of its descriptive, factual data, including the component values to be used in access points, whereas in VIAF the entity focus links to the related "authority heading" data which in turn may link to descriptive data. The FRBR-LRM model makes a clear distinction between the data describing the entity and the data describing its *Nomens*.

This applies to all entities except *Nomen* itself. The *Nomen* of a *Nomen* leads to the *Nomen* of a *Nomen* of a *Nomen*, and so on, with little utility or benefit. This is similar to the exclusion of a URI as a *Nomen* in the context of linked data: if a URI is a *Nomen*, then the URI of that *Nomen* is effectively a URI of a URI.

⁴ FOAF focus. Available at: http://xmlns.com/foaf/spec/#term_focus

An instance of any entity other than *Nomen* will usually require at least one human-readable *Nomen* to complement its URI; only the URI is necessary for an instance of *Nomen*. The attributes and relationships of a *Nomen* support the control of names as literal text strings to be used as access points to the related entity; the attributes and relationships of every other entity support its description and identification. This removes the traditional distinction between “authority” and “bibliographic” data: there is *Nomen* data, and entity data.

Internationalization

The failure of the top-down approach to Universal Bibliographic Control indicates that RDA will not be able to meet the needs of international communities by specifying too rigidly the elements and their citation order in an RDA AAP.

Conformance FRBR-LRM is likely to introduce new entities into RDA, for example Place. These may have different AAP characteristics.

The Working Group strongly supports the assumption is that the construction of an AAP is a matter for local specification at the application profile layer. RDA itself should accommodate a wide range of elements that can be used to disambiguate AAPs, and might suggest a high-level organizing principle for determining citation order.

Such a principle could be made machine-actionable, for example for generating access points from data elements on the fly.

Recommendation 5: The RDA instructions for constructing AAPs should be replaced with general guidelines for assigning *Nomens* for applications supporting the user task explore, as part of the development of guidelines and instructions for creating *Nomen* data.

Cultural heritage communities

RDA is required to develop to meet the needs of cultural heritage communities in general, including archives, galleries, and museums.

There is strong evidence that RDA will have to develop its treatment of title-type *Nomens*, that is, the titles of WEMI entities, because most cultural resources do not describe themselves in text. This may have some impact on associated AAPs. Titles supplied by cataloguers and users for multiple levels of resource aggregation will need to be accommodated. Development of authority control in RDA must occur with the cooperation of the cultural heritage communities, to ensure that library cultural assumptions are tested during the process.

There need to be many possible access points that can be considered valid in different contexts. Even with RDA, there are going to be some entities on which users of RDA internationally will not be able to agree upon a single AAP. For example, place names may be beyond the boundaries of agreement: English-speakers will never accept “Deutschland”

and German-speakers will never accept “Germany”. RDA will have to develop mechanisms that will support documenting all this different usage, as well as context-sensitive use of the data.

Recommendation 6: The development of RDA guidelines, instructions, and elements with respect to entity labels, identifiers, and access points should be carried out in consultation with international cultural heritage communities.

Recommendations

Recommendation 1: RDA should represent sub-types of Nomen as element sub-types of the appellation element.

Recommendation 2: Review and develop appropriate RDA elements for compatibility with the appellation-Nomen model by assigning element sub-types and ranges.

Recommendation 3: **a** Consider adding the RDA elements *family name* and *given name* as sub-elements of *name of the person*. **b** Develop these for the RDA Registry in any case, to improve interoperability of RDA linked data.

Recommendation 4: **a** Investigate the functionality and utility of “preferred” forms of appellation element sub-types in relation to RDA and application profiles in the context of the appellation-Nomen model. **b** Investigate the utility of relationships between Nomen and how RDA should accommodate them.

Recommendation 5: The RDA instructions for constructing AAPs should be replaced with general guidelines for assigning *Nomens* for applications supporting the user task explore, as part of the development of guidelines and instructions for creating *Nomen* data.

Recommendation 6: The development of RDA guidelines, instructions, and elements with respect to entity labels, identifiers, and access points should be carried out in consultation with other cultural heritage communities.