Translating RDA: a Discussion of the Opportunities and Challenges of an International Cataloging Standard

James Hennelly, ALA Publishing and Media Rania Osman, Bibliotheca Alexandrina Daniel Paradis, National Library and Archives of Québec Marja-Liisa Seppälä, National Library of Finland



Translating RDA Reference

Daniel Paradis

National Library and Archives of Québec



What is RDA Reference?

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

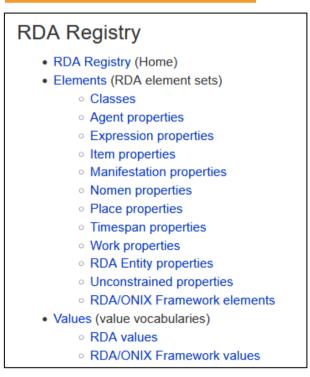
RDA Elements include entities, attributes, and relationships described in RDF. The value vocabularies include the controlled terminologies for RDA data given in the Toolkit instructions, and a controlled terminology used with the instructions in general (RDA Terms).

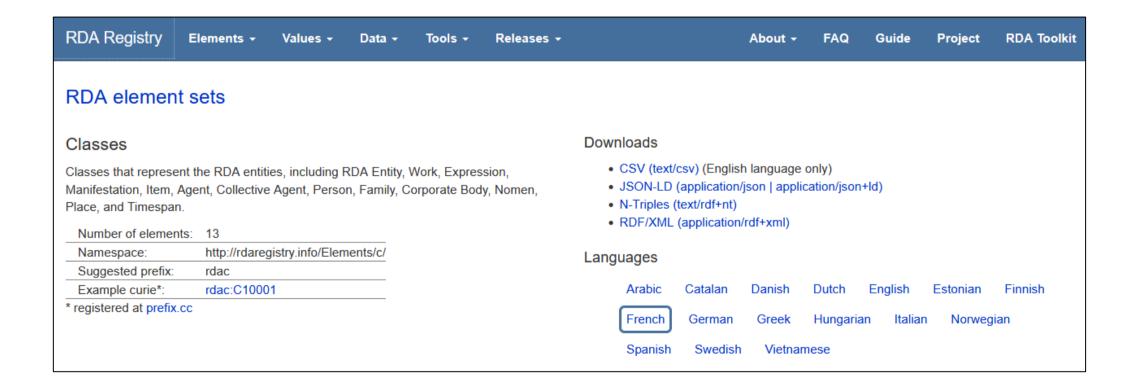
Source: RDA Registry

Partial translations vs full translations

- Translating RDA Reference is required as the initial step to translate the RDA Toolkit (full translation)
- RDA Reference and its translations are published as linked data in the

RDA Registry





Partial translations vs full translations

Less-than-full translations of RDA that cover all or part of RDA
 Reference are also possible, if the translation is maintained by an RDA
 user community or individual authorized by ALA Digital Reference

- Translations are uploaded into the Staff Registry using CSV files called translation templates
- UTF8 encoding supports all scripts
- There used to be one template per element set and vocabulary but the current templates combines them all in one file
- Contains one row for each property related to an element or a vocabulary value
- More than 18300 rows (including deprecated elements and vocabulary values) representing:
 - 3013 elements (deprecated elements not included)
 - 637 vocabulary values (deprecated values not included)
 - Set of unconstrained elements and RDA/ONIX Framework value vocabularies whose translation is optional

Tra

• There are always 4 properties for each element or vocabulary value, i.e. 2 types of labels and 2 types of definitions ("Registry" and Toolkit)

Elements	Vocabulary values
Labels: • rdfs:label • rdakit:toolkitLabel	Labels: • skos:prefLabel • rdakit:toolkitLabel
Definitions:skos:definitionrdakit:toolkitDefinition	Definitions:skos:definitionrdakit:toolkitDefinition

 This means a minimum of 14 600 rows for which a translation must be provided

7

- For value vocabularies, the two types of labels and the two types of definitions are identical
- For elements, the label and definition appearing in the RDA Registry are verbalized versions of the RDA Toolkit label and definition
 - The label is a verbal phrase using the pattern "is ..." or "has ..." + toolkitLabel
 - The definition includes the type of entity to which the element applies (the domain), using the template pattern "Relates a [domain entity] to " + toolkitDefinition

"<mark>has</mark> singer agent"

"Relates an expression to an agent who is a performer who contributes to an expression by using their voice, with or without instrumental accompaniment, to produce music."



The "Registry label" for elements is given as an alternate label in the Element Reference section of element pages

- Some elements and vocabulary values have one or two additional elements:
 - skos:scopeNote, for a scope note
 - skos:altLabel, for synonyms (including labels from original Toolkit when they have been modified during the 3R Projetc)

Example of properties for a vocabulary value (with synonym)

skos:prefLabel	multiple unit	unités multiples	
rdakit:toolkitLabel	multiple unit	unités multiples	
skos:definition	A mode of issuance of a manifestation that is issued as	Mode de parution d'une manifestation qui est	
	a multipart physical unit or intangible multipart logical	publiée en multiples unités physiques ou en	
	unit.	multiples unités logiques immatérielles.	
rdakit:toolkitDefinition	A mode of issuance of a manifestation that is issued as	Mode de parution d'une manifestation qui est	
	a multipart physical unit or intangible multipart logical	publiée en multiples unités physiques ou en	
	unit.	multiples unités logiques immatérielles.	
skos:altLabel	multipart monograph		

Note: Not all columns in the template are included

Example of properties for an element (with scope note)

rdfs:label	has singer agent	a pour agent chanteur			
skos:definition	Relates an expression to an agent who is a performer	Met une expression en relation avec un agent			
	who contributes to an expression by using their voice,	interprète qui contribue à une expression en			
	with or without instrumental accompaniment, to	utilisant sa voix, avec ou sans accompagnement			
	produce music.	instrumental, pour produire de la musique.			
skos:scopeNote	A performance of a singer may or may not include	Une prestation d'un chanteur peut inclure ou non			
	actual words.	des paroles.			
rdakit:toolkitLabel	singer agent	agent chanteur			
rdakit:toolkitLabel rdakit:toolkitDefinition	singer agent	-			
	singer agent	agent chanteur			
	singer agent An agent who is a performer who contributes to an	agent chanteur Agent interprète qui contribue à une expression en			

Note: Not all columns in the template are included

Adding a translation

Column3	Column4	Column5	Column6	▼	Column7	7	¥
property	original_en	modified_en	translation_original_fr		translation_new_fr		
skos:prefLabel	dataset	dataset			1		
	A category of linked data work that is an	A category of linked data work that					
	aggregation of metadata description sets	is an aggregation of metadata					
	that describe two or more instances of	description sets that describe two					
	one or more entities.	or more instances of one or more					
rdakit:toolkitDefinition		entities.					
rdakit:toolkitLabel	dataset	dataset					
rdakit:toolkitLabel	value vocabulary	value vocabulary					
	A category of linked data work that	A category of linked data work that		Mary on nor	vised translations		
	contains concepts that are used as values	contains concepts that are used as		new or rev	viseu	sed translations	
	for properties in a dataset.	values for properties in a dataset.		are added in this column		nis column	
rdakit:toolkitDefinition				are added	111 (1		
skos:prefLabel	value vocabulary	value vocabulary					
skos:altLabel	metadata element set	metadata element set					
rdakit:toolkitLabel	element set	element set					
	A category of linked data work that	A category of linked data work that					
	defines classes and properties that are	defines classes and properties that					
	used in a dataset.	are used in a dataset.					
rdakit:toolkitDefinition							

Challenges of the Reference translation process

- Every label must be unique
 - In French, series and collection are both translated by collection. How do we differentiate the labels? Keeping in mind that the term will also appear in element labels
- Some vocabularies use specialized technical terminology, e.g. for material and production method
- Some terms simply don't exist in French, e.g. some terms for Braille music, and French equivalents had to be devised
- The 3R Project introduced many neologisms, e.g. aggregating work, amalgamation, cardinality restriction, entity boundary
- In some cases, gender agreement in French was not obvious, e.g.: author person = personne auteure ou personne autrice? production designer family = famille chef décoratrice ou famille cheffe décoratrice?

Translating RDA

Challenges of the Reference translation process

- Consistency within an element set
 - Elements with the entity Agent as the range or the domain generally have 4
 narrower elements with similar labels and definitions, e.g. singer agent, singer
 collective agent, singer corporate body, singer family and singer person
 - Once a pattern is established, translating the narrower elements is easy but corrections can get complicated
- Consistency across different element sets
 - Elements and their inverses often have similar labels and definitions
 - Having a single template as is currently the case will help achieve consistency
- Consistency might be a bigger challenge if more than one translator is involved

Issues encountered during translation

- Using parentheses to qualify similar labels by element range was found inconsistent and confusing in the English version; parentheses were removed
 - augmentation of (work) → augmentation of work
 - catalogue (work) → catalogue work
- Parentheses cannot always be easily removed in French
 - augmentation of work → augmentation d'une œuvre
 - sequel work → œuvre ...?
 - catalogue work → œuvre catalographique?
- Parentheses retained in some instances
 - suite (œuvre)
 - catalogue (œuvre)

16

Issues encountered during translation

- Typos
- Inconsistencies between similar definitions
- Ambiguities in wording
- Problematic definitions. For example :
 - Definitions not adjusted for elements that had become relationship elements.

Example:

- coordinates of cartographic content (range = Place)
- Original definition in beta Toolkit: A mathematical system for identifying a place covered by the cartographic content of a work.
- Revised definition: A place that is the area of coverage of a cartographic work that is described using a mathematical system to identify its boundaries or location.

Problem with "issue"

- Does issuing cover only publication? What is the distinction between publication, release and issuance?
 - Scope note for name of publisher: Publication includes release and issuance.
 Google translation of note (July 2019): La publication inclut la publication et la publication (!!).
- issuing body agent, issuing body collective agent, issuing body corporate body, etc.
 - Issuing body used to refer only to a corporate body under AACR2. How to translate issuing body when the term applies to all types of agents? Could body be omitted?

To learn more about the original French translation project:

Clément Arsenault, Daniel Paradis & Pat Riva (2014) Translating RDA into French, Cataloging & Classification Quarterly, 52:6-7, 704-722, DOI:

10.1080/01639374.2014.889059