

Internationalization of RDA Toolkit during the 3R Project

Gordon Dunsire, Technical Team Liaison Officer, 21 Sep 2020

Introduction

This is an informational report on progress made during the RDA Toolkit Restructure and Redesign (3R) Project to support the internationalization strategy of the RDA Board.

The 3R Project started at the end of 2016 and is now scheduled to finish at the end of 2020. Internationalization has been one of the main drivers of the project, along with the need for a more modern and responsive design, and implementation of the IFLA Library Reference Model (LRM) that consolidates the international conceptual models on which RDA was originally based.

Principles of internationalization

The RDA Board has recently codified the principles on which the continuing internationalization of RDA is based.

Principle 1

The design and development of the standard should support international adoption.

Principle 2

Developing and maintaining the standard should provide opportunities to involve a wide range of individuals and communities.

Principle 3

The standard should recognise, respect and be informed by local knowledge, values, and practices.

Principle 4

Including global perspectives in the standard should benefit all parties involved.

Principle 5

The practice of integrating global viewpoints into the standards should be sustainable.

Achievements of the 3R Project

The following table lists general and specific changes to RDA Toolkit that were applied during the 3R Project. Each entry gives an indication of which internationalization principles are supported by the new feature. This is somewhat arbitrary; most features support all the principles to some degree.

Feature	1	2	3	4	5
The DITA standard for markup in the CMS is widely used for structured and codified applications and is supported by documentation aimed at a wide range of users.	✓	✓		✓	✓
<i>The CMS uses a specified sub-set of DITA elements to support consistency, problem-solving, the development of scripts for automated processes, and oversight by the publishers.</i>		✓		✓	✓
<i>Instructions and guidance use a high degree of reusable content driven by a file of marked-up boilerplate text. A single piece of boilerplate may be reused in thousands of files; over a thousand element files contain only the basic structure and boilerplate. The boilerplate file is translated separately from the bespoke content and populates a significant portion of the content of a translation. Policy statement agencies are encouraged to take a similar approach and use boilerplate for consistency and ease of maintenance.</i>	✓	✓		✓	✓
<i>The CMS makes extensive use of keys to represent pathways to link to specific content. The keys are independent of language, and preserve navigation links between sections of content across all translations.</i>	✓			✓	✓
<i>All logical components of RDA Toolkit content are identified separately for flexibility in linking and re-use.</i>	✓	✓		✓	✓
The modular design of the CMS supports flexible maintenance processes for translation teams and policy statement agencies.	✓	✓		✓	✓
<i>Instructions for each element are in separate files, grouped within a separate folder for each entity. Contents of guidance chapters are in separate files, grouped within a single folder.</i>	✓	✓			✓
<i>The instructions for each element are laid out in a basic common structure with standard labels. This allows machine-generation of the basic structure for translations and policy statements.</i>	✓	✓		✓	✓

The RDA element sets and value vocabularies are made available as 'RDA Reference' under an open licence that allows commercial reuse.	✓	✓	✓	✓	✓
<i>The underlying RDA data model is an implementation of the IFLA Library Reference Model.</i>	✓			✓	✓
<i>Translations of RDA Reference and mappings to related standards are included in the open license.</i>	✓	✓	✓	✓	✓
<i>RDA Reference content is encoded as linked data in several RDF serializations, and in comma-separated variable spreadsheets.</i>	✓	✓		✓	✓
<i>The RDA data model is extensible because it is based on established principles and good practice recommendations for the Semantic Web.</i>	✓	✓	✓		✓
The new Toolkit has a neutral focus on entities and elements. The same basic layout is applied to guidance and instructions for each entity and for each element.	✓	✓	✓	✓	✓
<i>Element hierarchies are encapsulated in the semantics of the element set and are displayed in appropriate contexts. Elements are otherwise listed in alphabetical order of label.</i>	✓		✓		
<i>The instructions support four distinct methods of recording metadata values, covering a wide range of applications and implementation scenarios.</i>	✓		✓		✓
The requirements for conformance with well-formed RDA metadata are as minimal as possible.	✓	✓	✓	✓	✓
<i>The mandatory requirements for recording an entity are minimized to the provision of at least one appellation (name, title, access point, or identifier). Resource entities (Work, Expression, Manifestation, and Item) additionally require at least one or two primary relationships to maintain the integrity of the underlying data model (LRM).</i>	✓		✓	✓	✓
<i>Where possible, specific instructions are optional and cover a range of application requirements.</i>	✓	✓	✓		
<i>Guidance on well-formed RDA metadata and conformance to RDA is included in the Toolkit.</i>	✓			✓	✓

The Toolkit provides contextual mappings and alignments between RDA elements and related standards. Related standards currently comprise Dublin Core, IFLA-LRM, MARC 21 Authority, and MARC 21 Bibliographic. The information is presented in the standard section for 'Element reference' of each element. A legacy mapping to ISBD is available in the RDA Registry and is awaiting update following the current revision of ISBD. The relationship between RDA carrier, content, and media types with the RDA/ONIX Framework for Resource Categorization is explained in guidance chapters. Plans are in development for mappings to BIBFRAME 2.0, the ISSN Manual, and UNIMARC.	✓		✓	✓	✓
The Toolkit provides value vocabularies for some elements as part of RDA Reference; these are translated as appropriate, but retain language-neutral codes, identifiers, and IRIs.	✓	✓	✓		✓
<i>The Toolkit provides accommodation for local value vocabularies for all relevant elements.</i>	✓	✓	✓	✓	
Guidance and instructions for manipulating text strings during transcription and for the creation of access points and other structured descriptions are moved to an area of the Toolkit that can be managed and developed by local RDA communities.	✓	✓	✓		✓
<i>The Anglophone focus on the manipulation of text strings is de-emphasized; guidance and instructions for strings in English are presented equally with other languages.</i>	✓		✓		✓
<i>The arrangement of local guidance and instructions is based on language, and brings together content such as abbreviations and capitalization rules for over 60 languages that was previously scattered.</i>	✓		✓		✓
<i>All non-English content is identified by an XML language code from the IANA/ISO standard. This content is not intended for translation. A file-wide language code is applied to English content that is intended for translation, and is amended during translation to the code for the translation language.</i>	✓			✓	✓
<i>Language codes and other mark-up follow standards for supporting access to content via screen-reader software.</i>		✓			✓

<i>The prominence of guidance and instructions in the Anglo-American cataloguing tradition for legal, musical, official, and religious resources is reduced.</i>	✓	✓	✓		
Glossary terms, definitions, scope notes, and cross-references are more consistent and cover a wider range of specialist and professional terminology used in the guidance and instructions.	✓			✓	✓
<i>Translations of the Glossary can use include synonyms that are appropriate for a specific language community. Translations of definitions and scope notes preserve the meaning, but are not required to be literal translations of the original English.</i>	✓		✓		✓
<i>The Translations Working Group has processes for proposing amendments to Glossary entries to improve clarity, consistency, and ease of translation.</i>	✓	✓	✓	✓	
The infrastructure of the CMS for maintaining examples is designed to improve context and scalability, and to allow translations to use local and appropriate examples for different language communities.	✓	✓		✓	✓
<i>Local examples can be added to policy statements.</i>	✓	✓	✓	✓	
Policy statements are maintained using the same infrastructure as other Toolkit content. This allows the sharing of documentation, special processes such as scripts for automatic updating of content, exchange of expertise, and transfer of skills between content editors.	✓	✓	✓	✓	✓
<i>The infrastructure is scalable and extensible to accommodate a larger number and wider range of policy agencies, including agencies for special kinds of collections and end-users, and in different sectors of library and other cultural heritage domains.</i>	✓	✓	✓	✓	✓
The Toolkit provides tools and accommodation for user-contributed content such as workflows and documentation localized to specific organizations and individuals. Tools include a text editor and documentation templates.	✓	✓	✓		✓
<i>User-contributed content can be shared with all Toolkit users as well as 'local' users.</i>	✓	✓		✓	

The structure and markup of RDA content is optimized for use with application profiles. This allows a profile to specify what elements to use and their mandatory and repeatability status. The recording method and specific options, vocabulary or string encoding schemes can also be specified.	✓	✓	✓		✓
The Toolkit offers a user-friendly online form for submitting feedback. Feedback is monitored at regular intervals and appropriate action is taken to improve the Toolkit.		✓	✓		✓

Utility

The internationalization of RDA Toolkit during the 3R Project has been comprehensive, although it is by no means complete. The latent Anglo-American focus of the guidance and instructions has been de-emphasized but coverage beyond this focus remains relatively shallow or does not yet exist.

For example, the treatment of religious materials associated with Christianity is more detailed than that of other 'major' religions, and there are no specific instructions for indigenous religions.

The restructuring of the Toolkit has created a burden for users who expect focused guidance and instructions relevant to local cataloguing traditions, processes, and applications. It will take time and effort to 'redesign and restructure' local interfaces to Toolkit content, and the long-term benefit of doing so may not be apparent to individuals who create and maintain library metadata at a local level.

The information in this paper may be useful for making the case for internationalization and to help Toolkit users understand the bigger picture.

Question: Is this information complete?

The 3R Project has been ambitious and complex, and has involved reviewing and reformatting every piece of content of the original Toolkit. It is inevitable that some features associated with internationalization have been inadvertently omitted from this paper.

The paper does not cover internationalization of RSC governance, such as regional representation and international membership of working groups.