

CENTRE FOR DIGITAL
LIBRARY RESEARCH



cdlr.strath.ac.uk

Collaborating communities: the RDA experience and its implications for common information environments

Gordon Dunsire

Presented at Archives, Libraries, Museums 11 (AKM11), Porec, Croatia, 2007

Published: Glasgow, Scotland : Centre for Digital Library Research, 2008

Centre for Digital Library Research

Livingstone Tower
26 Richmond Street
Glasgow
G1 1XH
Email: cdlr@strath.ac.uk
Website: <http://cdlr.strath.ac.uk/>



The CDLR at the University of Strathclyde brings together long-standing research interests in the digital information area. CDLR seeks to combine theory with practice in innovative ways with the aim of being a centre of excellence on digital library issues ranging from information policy and information retrieval to document storage technologies and standards.

Title

Collaborating communities: the RDA experience and its implications for common information environments.

Presented at Archives, Libraries, Museums 11 (AKM11), Porec, Croatia, 2007

Author

Gordon Dunsire

Introduction

This paper is presented as a case study which describes the interactions between several communities with a common interest in developing standards related to bibliographic information retrieval. Such interactions have mainly taken the form of a meeting followed by a programme of substantive work mutually agreed and carried out as a collaborative venture between technical representatives of those communities. The case study is therefore presented in the chronological order of those meetings

RDA

RDA stands for Resource Description and Access, a new standard designed for the digital world. It is built on foundations established by the Anglo-American Cataloguing Rules (AACR) during their development throughout the twentieth century. Although RDA is designed for use primarily in libraries, consultations are being carried out with other related communities such as archives, museums, and publishers in an effort to attain an effective level of alignment between RDA and the metadata standards used in those communities {1}. The development of RDA is being managed by the Joint Steering Committee for Development of RDA (JSC).

One of the key elements governing the way RDA is being designed is alignment with Functional requirements for bibliographic records (FRBR) {2} and Functional requirements for authority data (FRAD) {3}, the conceptual models for bibliographic and authority data developed by the International Federation of Library Associations and Institutions (IFLA) {4}. RDA also intends to provide references to other standards for describing and accessing information resources, such as those developed by the archive and museum communities, in order to supplement the more detailed instructions applicable to particular types of content, media, and modes of issuance, which are necessary to describe specific characteristics shown by a resource.

RDA is scheduled for publication in 2009, primarily as an online product with hypertext links for its sections, chapters and individual instructions. This will allow it to be customised and made context-sensitive to the cataloguer's task in hand.

ONIX

ONIX (Online Information Exchange) {5} is a standard for the distribution of metadata for the products of the publishing community. The standard is intended to be used by booksellers, other publishers, and anyone else involved in the sale of printed and digital bibliographic output. In particular, the ONIX for Books Product Information Message {6} is the international standard for representing and communicating book industry product information in electronic form. The ONIX community was aware of interest within the library community about re-using publisher metadata in a library community, and had encouraged the production of draft mappings from ONIX to the UniMARC and MARC21 metadata formats by the British Library and OCLC {7}.

Discussion between representatives of JSC and the publishing industry in the UK in October 2005 determined that significant mutual benefits might be realized through collaboration on the way resources are categorised on the basis of their content and how it is carried in physical and digital packaging. As a result the British Library hosted a meeting between the editor of RDA and consultants from EDItEUR, the organisation responsible for developing ONIX. The meeting took place in London between 27 and 28 March 2006, with a remit to discuss the adoption of a common high-level ontology of information carrier and content formats by the two communities. The meeting was very productive and participants continued to work together, via email, for the next four months. As a result, version 1.0 of the RDA/ONIX framework for resource categorization {8} was published on 3 August 2006.

RDA/ONIX framework

JSC stated that the framework would be used by the RDA Editor along with recommendations of the RDA GMD/SMD Working Group to prepare draft text for the RDA sections dealing with media category, type of carrier, and content category. GMD and SMD are respectively general material designation and specific material designation, terms used in AACR to denote content and carrier types. The categories given in the second edition of AACR (AACR2) were becoming increasingly difficult to apply in an environment where nearly all types of digital information content can be published on nearly all varieties of digital carrier such as compact disc and digital versatile disc. The framework ontology clearly separates content from carrier, and provides a method for creating high-level categories from the low-level attributes it defines. Further information is available from an article published in D-Lib magazine in January 2007 {9}.

Although the framework was developed by direct collaboration between the RDA and publishing communities, it was always intended to be of more general use. Relevant standards documentation from other communities was taken into account by the developers of the framework, and a preliminary attempt to apply the framework to the DCMI type vocabulary {10}, Multipurpose Internet mail extensions (MIME) media types {11}, and CIDOC conceptual reference model (CRM) {12} was discussed in the D-Lib magazine article. This demonstrates that the framework could be of significant use in reducing ambiguity in the content and carrier categories used by other communities, as

well as improving interoperability in the metadata produced and shared between such communities.

Dublin Core and related communities

The Dublin Core (DC) Scholarly works application profile (SWAP) was originally developed as the EPrints application profile {13} in 2006 by a working group of the UK's Joint Information Systems Committee. It is an application model developed for describing scholarly works such as papers, theses and conference presentations, and is based on the FRBR model. This was just one of several initiatives and informal discussions where the benefits of alignment between RDA and DC and related communities were raised, culminating in the organisation of a meeting "to examine the fit between RDA and models used in other metadata communities" held in London on 30 April and 1 May 2007. The so-called Data model meeting was again hosted by the British Library, with representation from RDA (including the Editor), the Dublin Core Metadata Initiative, Simple knowledge organisation system (SKOS), and Institute of Electrical and Electronics Engineers-Learning Object Metadata (IEEE-LOM). This meeting, too, proved fruitful, with participants agreeing that RDA and DCMI should collaborate to build on the existing work of both communities, and to specifically undertake a number of activities including the development of an RDA element vocabulary, development of an RDA DC application profile based on FRBR and FRAD, and disclosure of RDA value vocabularies using RDF/RDFS/SKOS. RDF(S) is Resource description framework (schema), a syntax standard supported by W3C, while SKOS is a W3C standard for modelling vocabularies.

DCMI/RDA Task Group

The agreed activities are being undertaken by a DCMI/RDA Task Group which is co-chaired by two of the participants at the Data models meeting. Neither DCMI nor JSC is able to fund the consultancy work that is required, so funding from external bodies is being sought. The Task Group has been able to set up a public wiki {14} and private shared workspace, and communicates via email, tele-conference, and ad hoc and planned meetings at conferences.

One of the first activities was the organisation of a Special session on RDA at the annual DC meeting, DC2007, in Singapore on 30 August 2007. The meeting was full, indicating the high level of interest in the DC community regarding RDA. The Task Group has also decided to use the National Science Digital Library metadata registry sandbox {15} to test vocabulary disclosure in RDF/SKOS. The sandbox provides easy-to-use interfaces for adding and amending controlled vocabulary terms and outputting them in correct RDF/SKOS syntax. The results are publicly visible, but cannot be confused with published vocabularies. Some preliminary work has been done on registering the carrier vocabulary from RDA chapter 3, which will therefore also be of interest to the ONIX community.

FRBR

FRBR was a significant influence on the development of AACR2 to what was intended to be the third edition (AACR3) and subsequently abandoned in favour of the more radical approach used in RDA. FRBR remains a key underpinning of RDA, with RDA metadata elements or attributes related as appropriate to FRBR user tasks, entities and relationships. Such links were reaffirmed and reinforced by the JSC at its meeting of 15 to 20 October 2007 {16} when a new organisation of the RDA chapters was agreed, with one of the aims being better alignment with FRBR and FRAD.

Although not made public until the minutes were published in November 2007, the FRBR Review Group meeting of 21 August 2007 {17} held at the World Library and Information Congress in Durban proposed a project to "To define appropriate namespaces for FRBR (entity-relationship) in RDF and other appropriate syntaxes." The proposal was stimulated by the outcomes of the Data modelling meeting, which might require references to FRBR elements in the definitions of RDA elements in RDF/SKOS. The project is intending to complete work early in 2008. This FRBR project fits well with the schedule of the DCMI/RDA Task Group, which is intending to start work around the same time. Close links are maintained with the Task Group, and space on the wiki has been made available to the FRBR project {18}. The NSDL sandbox is also being used to test the registration of vocabularies for FRBR entities, FRBR relationships, and FRBR user tasks.

Common information environments

"A common information environment offers easy, convenient access to the widest range of information resources catering for the widest range of needs of the widest range of users. It requires the collaboration of archives, libraries, museums and governments" {19}. It is clear that the Internet and World-Wide Web ("Web1.0") will be a significant and necessary component of any large-scale common information environment.

"Web2.0" is often used to refer to the evolution of the World Wide Web from a digital communication, publishing and retrieval platform used by separate persons and groups to an environment which encourages shared approaches to these activities. Without Web2.0, it seems unlikely that the communities involved in this case study could or would have been able to collaborate so effectively in such a short period of time. Wikis, teleconference management and the NSDL registry are all examples of Web2.0 services which have proved essential to reaching agreement, scheduling activities, and carrying out tasks between participants who are located in many different countries around the world.

"Web3.0" is now being discussed as the next step in evolving from Web2.0. It cannot be just a refinement of the existing environment, as that would be labelled Web2.1 or some such. Instead, Web3.0 must represent a significant "quantum leap" in the utility and functionality of our open, global information and communications technology infrastructure. Web3.0 must also support and enhance the functionality of the common information environment, for at a global level there is only one environment. It is likely to

achieve this by improving machine assistance for information retrieval across multiple heterogeneous collections described by varying metadata structures with varying syntax, labels and content. RDF, SKOS and "namespace" declarations are core technologies for making machine-mediated retrieval more effective, by providing a solid basis for machine interoperability between different metadata formats by using metadata element and value vocabularies.

But these technologies are also the basis of the development of the semantic web, so the proposition that Web3.0 is the semantic web is a reasonable one to make.

RDA and Semantic Web

Incorporating RDA and its associated standards (RDA+) such as FRBR into the semantic web should help to provide users with answers to some common, fundamental questions about the environment within which they formulate and carry out their information-related tasks; questions such as:

- Can I restrict my search to non-visual resources? I don't care whether they are from archives, libraries and museums or any other type of organisation.
- Does your collection define audio content the same way as I do?
- Is your concept of "title" the same as mine? Does the computer "know" that Archive.title = Library.title = Museum.title = Bibliotheque.titre = MARC.245?

If we professionals can't provide the answers, how can we programme a computer to enhance the effectiveness of our services by indexing the Web in a structured way? The semantic web infrastructure allows archive, library, museum and related communities to provide consistent information about their metadata standards and practices and begin to give clear and definite answers to all of these questions. And widespread adoption of RDA+ by those communities will ensure that those answers are generally positive.

References

- {1} RDA - Resource description and access: a prospectus. Revised: 17 December 2007. Available at: <http://www.collectionscanada.gc.ca/jsc/rdapropectus.html>
- {2} Functional requirements for bibliographic records: final report. 1998. Available at: <http://www.ifla.org/VII/s13/frbr/frbr.pdf>
- {3} Functional requirements for authority data: a conceptual model. Draft 2007-04-01. Available at: <http://www.ifla.org/VII/d4/FANAR-ConceptualModel-2ndReview.pdf>
- {4} International Federation of Library Associations and Institutions: IFLAnet. Available at: <http://www.ifla.org/>
- {5} ONIX (Online Information Exchange). Available at: <http://www.bisg.org/onix/index.html>

- {6} ONIX for books. Release 2.1 (revision 02). 2004. Available at:
<http://www.editeur.org/onix.html>
- {7} ONIX mappings to MARC. Available at: <http://www.editeur.org/onixmarc.html>
- {8} RDA/ONIX framework for resource categorization. Version 1.0. 2006. Available at:
<http://www.collectionscanada.gc.ca/jsc/docs/5chair10.pdf>
- {9} Distinguishing content from carrier: the RDA/ONIX framework for resource categorization. January/February 2007. Available at:
<http://www.dlib.org/dlib/january07/dunsire/01dunsire.html>
- {10} DCMI type vocabulary. Available at:
<http://dublincore.org/documents/dcmi-type-vocabulary/>
- {11} Multipurpose Internet mail extensions (MIME) part two: Media types. Available at:
<http://www.isi.edu/in-notes/rfc2046.txt>
- {12} Definition of the CIDOC conceptual reference model. Available at:
http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.pdf
- {13} Eprints application profile. Available at:
http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile
- {14} DCMI/RDA Task Group wiki. Available at: <http://dublincore.org/dcmirdataskgroup/>
- {15} NSDL Registry sandbox. Available at: <http://sandbox.metadatatregistry.org/>
- {16} Outcomes of the meeting of the Joint Steering Committee held in Chicago, USA, 15-20 October 2007. Available at:
<http://www.collectionscanada.gc.ca/jsc/0710out.html>
- {17} IFLA Cataloguing Section, FRBR Review Group, Meeting report, Durban, August 21, 2007. Available at: http://www.ifla.org/VII/s13/wqfrbr/FRBR_RG_Mtg2007.pdf
- {18} Namespace for FRBR entities/elements in RDF. Available at:
http://dublincore.org/dcmirdataskgroup/namespace_20for_20FRBR_20entities_2f_elements_20in_20RDF
- {19} The Common Information Environment: a newly emerged concept. Presented at Archives, Libraries, Museums 9 (AKM9), Porec, Croatia, 2005. Available at:
<http://cdlr.strath.ac.uk/pubs/dunsireg/akm2005cie.pdf>