UNIMARC, RDA and the Semantic Web

Gordon Dunsire
Depute Director, Centre for Digital Library Research
University of Strathclyde
Glasgow, Scotland

Meeting: **135. UNIMARC**

Abstract:

The paper will discuss the application of Resource Description and Access (RDA), the emerging successor to the Anglo-American Cataloguing Rules, as a content standard for metadata encoded in UNIMARC. RDA is designed for international application in a digital environment, and is not aligned with any specific bibliographic record encoding format, although work is ongoing to develop its application to MARC21 and Dublin Core formats. The paper will also discuss the implications of making components of RDA and associated models such as Functional Requirements for Bibliographic Records (FRBR) and Functional Requirements for Authority Data (FRAD) compatible with the Semantic Web.

UNIMARC, RDA and the Semantic Web

UNIMARC[1] is developed and maintained by the International Federation of Library Associations and Institutions (IFLA)[2]. It is a carrier format intended for the exchange of bibliographic metadata between the systems used by national agencies. As such, it does not specify the metadata structure and content to be used within individual systems.

The goals of UNIMARC's current strategic plan[3] are:

1. Ensure the maintenance and development of UNIMARC, in alignment with other MARC formats and new bibliographic standards.
2. Enhance the portability of UNIMARC data to the Web environment and the interoperability of UNIMARC with other data standards.
3. Improve the updating and availability of UNIMARC documentation.

4. Advance knowledge of UNIMARC and its usage and promote mechanisms and actions towards sharing of expertise and improvement in user support.

An important element of UNIMARC is its alignment with International Standard Bibliographic Description (ISBD)[4], also developed by IFLA. The primary goal of ISBD is to offer consistency in sharing bibliographic information at international level by specifying the data elements to be used as the basis of metadata records, and a mechanism for identifying and displaying such elements independent of the language of the record. The ISBD elements have been mapped to the entities and relationships[5] defined in Functional Requirements for Bibliographic Records (FRBR)[6], a model for bibliographic data developed by IFLA.

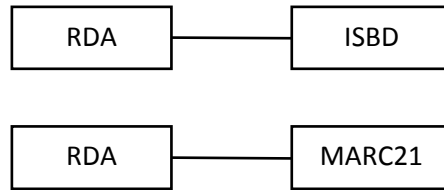These alignments and mappings are shown in this simple diagram:

| UNIMARC | — | ISBD | — | FRBR |

The entity-relationship model of FRBR has been extended as an object-oriented model, FRBRoo[7], which is compatible with the CIDOC Conceptual reference model (CRM)[8], initially developed for the museum community.

RDA - resource description and access is a new metadata standard for describing the content of information resources to improve identification of, and access to, such content. The standard is designed for the digital environment, but is built on over one hundred years of experience in developing the Anglo-American Cataloguing Rules (AACR). It is intended for international use, and is not constrained by conventions used in English-speaking countries. While RDA is primarily focussed on its application to resources in library collections, it aims to achieve an effective degree of compatibility with metadata approaches used in related communities such as archives, museums, and publishers.[9]

An important element of RDA is its alignment with FRBR and the associated Functional Requirements for Authority Date (FRAD)[10], which is a model for authority data developed by IFLA.

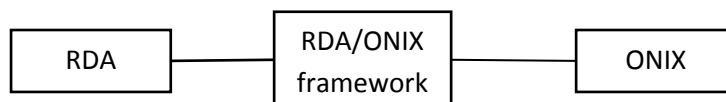| RDA | — | FRBR | — | FRAD |
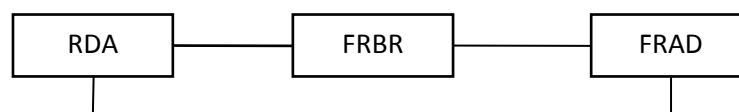
Another significant element of RDA is its independence from any specific structure or format for storing or displaying metadata. It is designed, however, to maximise the integration of RDA-based data with existing data, and specifically that produced using AACR and related standards. Accordingly, Appendix D of the RDA draft contains mappings between elements of RDA and ISBD and RDA and MARC21.
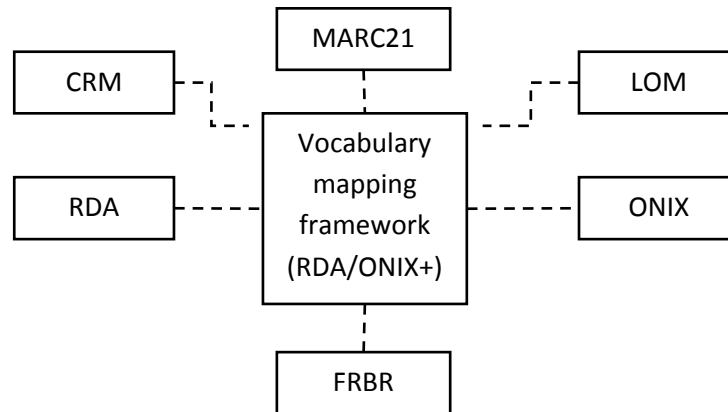
```
┌──────────┐          ┌──────────┐
│   RDA    │──────────│   ISBD   │
└──────────┘          └──────────┘

┌──────────┐          ┌──────────┐
│   RDA    │──────────│  MARC21  │
└──────────┘          └──────────┘
```

The ISBD Material Designation Study Group[11] has prepared the proposed new area 0 of ISBD[12] in alignment with the near-final draft of RDA. ISBD area 0 covers content form and media type, and is intended to replace the current expression of general material designation (GMD) while continuing to have utility as an "early warning device" for catalogue users, placed at the beginning of a record. This indicates that a resource requires a particular human sense or intermediary device to access the resource being described. Content form refers to the fundamental way in which the content of a resource is expressed; for example "image". Media type refers to the carrier which conveys the content of the resource; for example "audio". This area thus clearly distinguishes the content of the resource from its carrier, which the current GMD approach fails to achieve. The proposal is aligned with RDA.

```
┌──────────┐          ┌──────────┐
│  ISBD 0  │──────────│   RDA    │
└──────────┘          └──────────┘
```
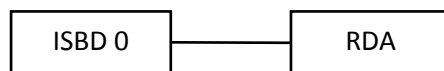
The RDA media and carrier type categories (carrier type is an extension of media type) and content type categories are themselves based on the RDA/ONIX framework for resource categorisation.[13] This is an ontology for determining high-level content and carrier categories for information resources, and is aligned with the ONIX metadata schema for the publishing community. The framework is intended to meet the needs of any community requiring metadata categories for resource content or carrier[14], although it has so far only been applied in practice to RDA[15].

```
┌──────────┐     ┌──────────┐     ┌──────────┐
│   RDA    │─────│ RDA/ONIX │─────│   ONIX   │
│          │     │framework │     │          │
└──────────┘     └──────────┘     └──────────┘
```
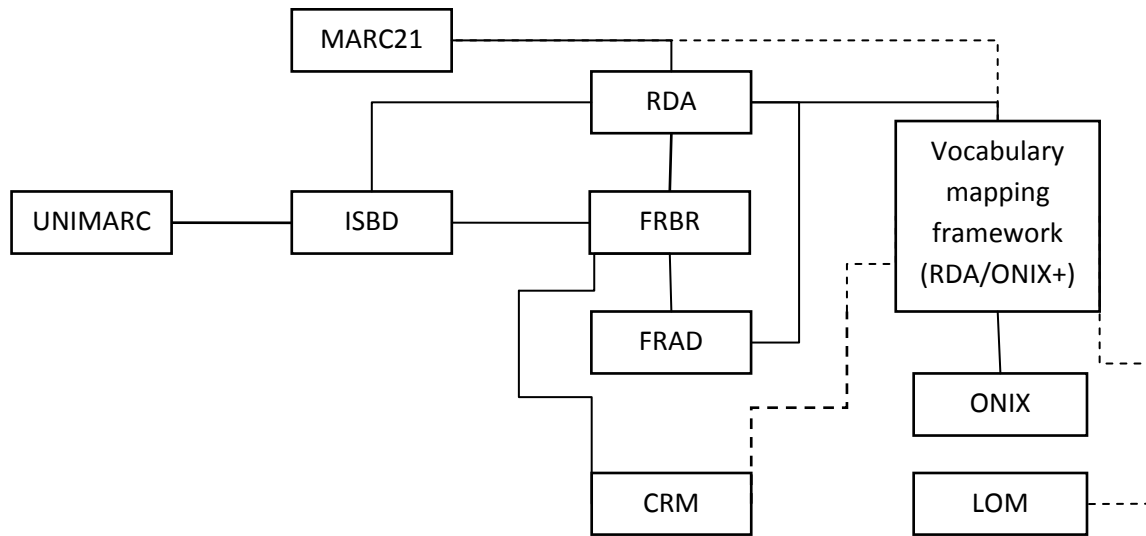
A project to extend the RDA/ONIX framework to cover bibliographic roles and relationships will run from June to November 2009, with the provisional title Vocabulary mapping framework (VMF). This will require the addition of agent categories to the framework, in order to express, for example, relationships between FRBR group 1 and group 2 entities. Group 1 entities represent the products of intellectual or artistic endeavour, such as an expression of a work, while group 2 entities represent the agents responsible for various aspects of those products, such as the creator of a work. The extended framework will also go beyond RDA and ONIX, taking into account appropriate areas of CRM, Dublin Core Metadata Initiative (DCMI), FRBR, IEEE Learning object metadata (LOM), and MARC21. It will thus provide alignments between metadata used in the museum, Web, and teaching communities in addition to library and publishing communities.

```
                    ┌─────────┐
                    │ MARC21  │
                    └─────────┘
┌──────┐               │
│ CRM  │- -┐    ┌───────────────┐    ┌──────┐
└──────┘   └- - │   Vocabulary  │ ┌- -│ LOM  │
                │    mapping    │-┘   └──────┘
┌──────┐        │   framework   │
│ RDA  │- - - - │  (RDA/ONIX+)  │- - - ┌──────┐
└──────┘        │               │      │ ONIX │
                └───────────────┘      └──────┘
                        │
                   ┌─────────┐
                   │  FRBR   │
                   └─────────┘
```

It should be noted that many of the alignments and mappings described above are not exact, although all improve the interoperability of metadata from different communities and in different formats. Closer alignment may be possible, particularly when specific metadata schemas are undergoing development and refinement. This is, indeed, the case with the alignment of ISBD area 0 and RDA. The development of GMD in ISBD was initiated by the work of the IFLA Meeting of Experts on an International Cataloguing Code (IME ICC)[16], and subsequently took into account the RDA/ONIX framework and RDA itself [17].

The development of RDA also suggested potential changes to MARC21 to maintain or improve alignment between them.[18] These include changes to the way that GMD and specific material designation (SMD) are treated in MARC21, including the coded leader fields in MARC21, plus adjustments in specific fields. These range from the fairly trivial, for example the addition of "RDA" as a value in fields recording which cataloguing rules have been used to create the metadata, to more complex issues such as the subfields used to record publication and distribution metadata. As a result, the RDA/MARC Working Group was set up under the auspices of the British Library, the Library and Archives Canada, and the Library of Congress in 2008, to "identify what changes are required to MARC 21 to support compatibility with RDA and ensure effective data exchange into the future."[19] An early output of the group was a discussion paper[20] for the Machine-Readable Bibliographic Information (MARBI) Committee[21]. Recommendations and decisions arising from this activity are not yet complete (as of June 2009). For example, RDA distinguishes two modes of issuance for monographs: "single unit" and "multipart monograph". MARC21 has encoding for "monograph" and "multipart resource", and has a choice of either splitting the "monograph" value into two, or combining both values (which are encoded in different places), to improve the alignment with RDA; the issue remains under discussion. However, significant changes to MARC21 have already been approved with respect to content and carrier categories. There are three new tags for media type, content type, and carrier type, "intended as replacements for the general material designation (GMD) defined in AACR2 1.1C, currently recorded in field 245 (Title statement), subfield $h (Medium)."[22] These echo the changes proposed for ISBD.

The various alignments described above can be linked into a chain or network (this is not a complete picture!):

```
  MARC21 ─────────────┐ - - - - - - - - - - - - - - - - ┐
                      │                                 │
                ┌─── RDA ──────────────────────┐        │
                │     │                  Vocabulary     │
  UNIMARC ──── ISBD ── FRBR              mapping        │
                │       │                framework      │
                │     FRAD ──────────── (RDA/ONIX+)     │
                │       │                   │           │
                │       │                 ONIX - - - ┐  │
                └──── CRM                  LOM - - - -┘
```

This connects UNIMARC to RDA, primarily via ISBD.

```
  UNIMARC ──── ISBD ──── RDA
```
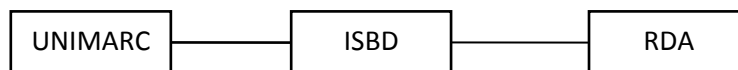
UNIMARC is also aligned directly with other components linked to RDA. There is a mapping between UNIMARC and MARC21.[23] This has not been updated since 2001, and requires review because of the changes to MARC21 arising from RDA, let alone other developments.

```
  UNIMARC ──── MARC21
```
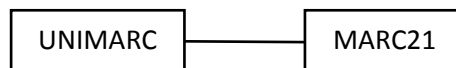
This introduces a separate connection between UNIMARC and RDA, via MARC21.

```
  UNIMARC ──── MARC21 ──── RDA
```
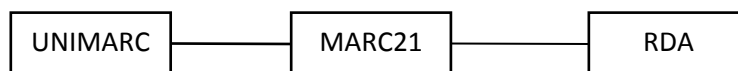
This suggests that a general review of the alignments, albeit indirect, between UNIMARC and RDA is required if the first and second goals of the current UNIMARC strategic plan are to be met. Trivial examples are the representation of mode of issuance for single part and multipart monographs, and "RDA" as a new value for UNIMARC appendix H (cataloguing rules and formats codes), discussed above in relation to MARC21. More serious issues may be hidden in the complexity of the network of alignments; minor misalignments in each part of the chain can result in major misalignments in indirect links.

If national cataloguing rules wish to evolve and benefit from RDA, FRBR, etc., then alignment with UNIMARC can be very important. In fact, there may be many nodes in the network of alignments to which any other cataloguing rules might be linked, even if they are not based on RDA. For example, if the new Italian cataloguing rules (which have not been published at the time of writing) are aligned with FRBR, does alignment of FRBR with RDA and alignment of the current Italian rules with UNIMARC have any significance for the new cataloguing rules? Another example is the French cataloguing rules, which use ISBD directly and specify which options are to be taken. Reviewing the network of alignments with respect to any national cataloguing rules would be justified because of the potential benefit.

The network of alignments described so far is incomplete with respect to the second goal of the UNIMARC strategic plan. There is a mapping from UNIMARC to Dublin Core (DC).[24] Although the published version dates from 1997 (and there have been some updates to it since), it remains essentially current because DC has been stable and the mapping is very lossy; that is, changes to UNIMARC are unlikely to affect the alignment.

```
┌─────────────┐         ┌─────────────┐
│   UNIMARC   ├─────────┤     DC      │
└─────────────┘         └─────────────┘
```
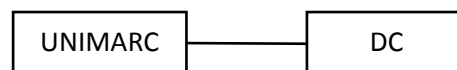
There is also a "formal representation of UNIMARC rules and associated vocabularies" in extensible markup language (XML)[25] and current activity to introduce UNISlim XML Schema as a refinement of UNIMARCXML. These go some way to meet the strategic plan. MARC21 also has a representation in XML[26], and ISBD XML is under consideration. XML is essentially a mechanism for data interchange.

However, there is a great deal of activity underway to make various nodes of the network of alignments compatible with the Semantic Web. This requires expression of elements of the various standards using Resource description framework (RDF).[27] More specifically, elements relating to metadata structure, such as tags, fields, and attributes, need to be expressed as classes and properties in RDF schema (RDFS)[28], while elements relating to metadata content, such as codes and controlled vocabularies, need to be expressed in Simple knowledge organization system (SKOS).[29] Complicated semantic relationships in vocabularies can be expressed in Web ontology language (OWL).[30]

The DCMI RDA Task Group[31] is engaged in expressing RDA controlled vocabularies, including the content and carrier types, in SKOS, and the RDA metadata structure elements in RDFS. The FRBR Review Group[32] is engaged in expressing FRBR entities and relationships in RDFS. The VMF project is likely to express the extended RDA/ONIX ontology, and the high level vocabularies for categories and relationships based on it, in RDFS and SKOS. The Library of Congress intends to express elements of MARC21 in formats compatible with the Semantic Web: "LC has embarked on initiatives to provide SKOS representations for vocabularies and data elements used in and across standards, such as RDA, MARC, PREMIS and METS."[33] The Library of Congress Authorities and Vocabularies service[34] makes Library of Congress Subject Headings (LCSH) available in SKOS, and intends to add MARC geographic area, language, and relator codes, and the Thesaurus of graphic materials. There are numerous benefits to be gained from this work.[35] [36]

While XML representation of UNIMARC and related standards is essential for compatibility with the Semantic Web, because it allows machine-to-machine interoperability with RDF XML, in which SKOS, RDFS and OWL are implemented, a further essential requirement is that human-readable values can be referenced with a Uniform resource identifier (URI)[37] to allow effective and efficient machine-processing. These developments are likely to have a significant impact on cataloguing concepts and workflows as a result of evolving from the scenario 2 to scenario 1 implementation of RDA[38] following the "FRBRization" of the catalogue record and impact of the Semantic Web[39].

As an example, the RDA content type value "spoken word" will be available in a SKOS representation[40] soon after the publication of RDA. The SKOS representation assigns a URI to this term. The Deutsche Nationalbibliothek (DNB) has added the German translation "gesprochene Worte" to the representation. Any alignment which uses the URI to link to this term will automatically provide human-readable alignments in both English and German, depending only on software and not on content. Any number of translations of this term can be added to the representation, and all that will be required to provide multilingual interoperability of catalogues based on RDA will be minor adjustments to the processing software. This very much meets the primary aim of UNIMARC.

References

[1] IFLA Universal Bibliographic Control and International MARC Core Programme (UBCIM). UNIMARC manual : bibliographic format 1994. Introduction. Available at: http://archive.ifla.org/VI/3/p1996-1/sectn1.htm

[2] International Federation of Library Associations and Institutions (IFLA). Available at: http://www.ifla.org/

[3] IFLA UNIMARC Core Activity: strategic plan 2007-2009. Jul 2007. Available at: http://archive.ifla.org/VI/8/annual/unimarc-sp-2009.pdf

[4] IFLA. International standard bibliographic description (ISBD). Preliminary consolidated edition. April 2007. Available at: http://www.ifla.org/files/cataloguing/isbd/isbd-cons_2007-en.pdf

[5] Mapping ISBD elements to FRBR entity attributes and relationships. 28 Jul 2004. Available at: http://archive.ifla.org/VII/s13/pubs/ISBD-FRBR-mappingFinal.pdf

[6] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records. Available at: http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

[7] International Council of Museums. FRBRoo introduction. Available at: http://cidoc.ics.forth.gr/frbr_inro.html

[8] International Council of Museums. The CIDOC conceptual reference model. Available at: http://cidoc.ics.forth.gr/index.html

[9] Joint Steering Committee for Development of RDA. RDA — resource description and access: a prospectus. Version of 28 Oct 2008. Available at: http://www.rda-jsc.org/docs/5rda-prospectusrev6.pdf

[10] IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Functional requirements for authority data. Available at: http://archive.ifla.org/VII/d4/wg-franar.htm

[11] IFLA. Material Designations Study Group. Available at: http://www.ifla.org/en/node/938

[12] IFLA. ISBD Review Group. Proposed Area 0 for ISBD. Available at: http://archive.ifla.org/VII/s13/isbdrg/ISBD_Area_0_WWR.htm

[13] Joint Steering Committee for Development of RDA. RDA/ONIX framework for resource categorization. Version 1.0. 1 Aug 2006. Available at: http://www.rda-jsc.org/docs/5chair10.pdf

[14] Dunsire, G. Distinguishing content from carrier: the RDA/ONIX framework for resource categorization. In: D-Lib magazine, volume 13 number 1/2 (Jan/Feb 2007), ISSN 1082-9873. Available at: http://www.dlib.org/dlib/january07/dunsire/01dunsire.html

[15] Delsey, T. Categorization of content and carrier. 8 Aug 2006. Available at: http://www.rda-jsc.org/docs/5rda-parta-categorization.pdf

[16] IFLA Cataloguing Section. IME-ICC: IFLA Meetings of Experts on an International Cataloguing Code. Available at: http://www.ifla.org/node/576

[17] Tillett, B. The future of cataloguing codes and systems: IME ICC, FRBR, and RDA. In: UNIMARC & friends: charting the new landscape of library standards. Proceedings of the international conference held in Lisbon, 20-21 March 2006. Edited by Plassard, Marie-France. Berlin, New York (Walter de Gruyter – K. G. Saur) 2007. Pages 27–40. eBook ISBN: 978-3-598-44034-2. Print ISBN: 978-3-598-24279-3

[18] Joint Steering Committee for Development of RDA. RDA and MARC21. 17 Nov 2006. Available at: http://www.rda-jsc.org/docs/5chair12.pdf

[19] Joint Steering Committee for Development of RDA. RDA/MARC Working Group update. Available at: http://www.rda-jsc.org/rdamarcwg.html

[20] Library of Congress. Network Development and MARC Standards Office. MARC discussion paper no. 2008-DP04: Encoding RDA, Resource Description and Access data in MARC 21. 14 Dec 2007. Available at: http://www.loc.gov/marc/marbi/2008/2008-dp04.html

[21] American Library Association. Machine-Readable Bibliographic Information (MARBI) Committee. Available at: http://www.ala.org/ala/mgrps/divs/alcts/mgrps/divgroups/marbi/marbi.cfm

[22] Library of Congress. Network Development and MARC Standards Office. MARC proposal no. 2009-01/2: New content designation for RDA elements: Content type, Media Type, Carrier Type. 9 Jan 2009. Available at: http://www.loc.gov/marc/marbi/2009/2009-01-2.html

[23] Library of Congress. Network Development and MARC Standards Office. UNIMARC to MARC 21 conversion specifications. Version 3.0 (August 2001). Available at: http://www.loc.gov/marc/unimarctomarc21.html

[24] Day. M. Mapping Dublin Core to UNIMARC. July 1997. Available at: http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html

[25] BookMARC. The UNIMARC and MARC21 manual in XML: automating validation, explanation and help systems for bibliographic records. 2004. Available at: http://www.bookmarc.pt/unimarc/

[26] Library of Congress. Network Development and MARC Standards Office. MARCXML: MARC21 XML schema. Available at: http://www.loc.gov/standards/marcxml/

[27] W3C. Resource description framework (RDF). Available at: http://www.w3.org/RDF/

[28] W3C. RDF vocabulary description language 1.0: RDF schema. 10 Feb 2004. Available at: http://www.w3.org/TR/rdf-schema/

[29] W3C. SKOS: Simple knowledge organization system - home page. Available at: http://www.w3.org/2004/02/skos/

[30] W3C. Web ontology language (OWL). Available at: http://www.w3.org/2004/OWL/

[31] Dublin Core Metadata Initiative. DCMI/RDA Task Group Wiki. Available at: http://dublincore.org/dcmirdataskgroup/

[32] IFLA. FRBR Review Group. Available at: http://www.ifla.org/en/frbr-rg

[33] Marcum, D.B. Response to On the record: report of the Library of Congress Working Group on the Future of Bibliographic Control. I Jun 2008. Available at: http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf

[34] Library of Congress. Authorities & vocabularies. Available at: http://id.loc.gov/authorities/

[35] Styles, R., Ayers, D., & Shabir, N. Semantic MARC, MARC21 and the Semantic Web. In: Proceedings of the Linked Data on the Web Workshop, Beijing, China, April 22, 2008, CEUR workshop

proceedings, ISSN 1613-0073. Available at: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-369/paper02.pdf

[36] Kruk, S.R., Synak, M., & Zimmermann, K. MarcOnt – integration ontology for bibliographic description formats. Presented at DC2005, Madrid. Available at: http://dc2005.uc3m.es/program/presentations/Thursday%2015.%2015.30%20h%20-%20s.kruk.pdf

[37] IETF. Uniform resource identifier: generic syntax. Available at: http://tools.ietf.org/html/rfc3986

[38] Delsey, T. RDA implementation scenarios. 14 Jan 2007. Available at: http://www.rda-jsc.org/docs/5editor2.pdf

[39] Dunsire, G. The Semantic Web and expert metadata : pull apart then bring together: Presented at Archives, Libraries, Museums 12 (AKM12), Poreč, Croatia, 2008. Available at: http://cdlr.strath.ac.uk/pubs/dunsireg/akm2008semanticweb.pdf

[40] Available at: http://metadataregistry.org/concept/show/id/522.html