

VMF project announcement: June 15<sup>th</sup> 2009

## Major content metadata vocabularies to be mapped

Work is under way to create an extensive and authoritative mapping of vocabularies from major content metadata standards, creating a downloadable tool to support interoperability across communities.

The work is an expansion of the existing RDA/ONIX Framework into a comprehensive vocabulary of resource relators and categories, which will be a superset of those used in major standards from the publisher/producer, education and bibliographic/heritage communities (CIDOC CRM; DCMI; DDEX; DOI; FRBR; MARC21; LOM; ONIX; RDA – see reference section below for details).

The resulting tool will be known as the **Vocabulary Mapping Framework** (VMF).

The new vocabulary is not intended as a replacement for any existing standards, but as an aid to interoperability, whether automatic or human-mediated. The expanded Framework will include mappings of terms from code lists or allowed value sets in the existing standards to the RDA/ONIX vocabulary, enabling the computation of “best fit” mappings between any pairing of standards.

The results of the VMF project will be formally presented at an event at the British Library on the morning of November 9<sup>th</sup> this year, and made available on the Web.

The project, which is largely financed by a grant from the UK Joint Information Systems Committee (JISC), is being carried out by Godfrey Rust and Steffen Lindek of Rightscom and Gordon Dunsire, Depute Director of the Centre for Digital Library Research at Strathclyde University in Glasgow, Scotland, with input from other domain experts. A virtual Advisory Group drawn from interested parties is being convened.

The International DOI Foundation, which fully endorses this work, will provide the web hosting facility as part of its commitment to promoting the wider use of interoperable metadata, and will use the mapping vocabulary wherever possible to support the association of metadata with DOI names.

The existing RDA/ONIX Framework (which currently supports categorization of resource content and carriers) will be extended to support: *works; parties; relators between resources*, and *relators between parties and resources*.

ISO TC46/SC9 identifiers (ISBN, etc) are also among those standards which may be reviewed to support formal concept analysis, and some may be included in the Framework in future.

The intent is that the Framework will be maintained on an ongoing basis and that new vocabularies, standard or proprietary, may be added by interested parties.

The work will result in:

- a mapping of vocabularies from the source standards to support the building of crosswalks and transformations between any of them;
- a definitive reference set which editors can draw on when creating and developing standards;

- a downloadable RDF/OWL ontology to support the interchange of metadata content between these major standards, which will be useful to enable automated reuse of metadata from different sources and schemas, to improve the quality and access and reduce the cost of metadata;
- a governance scheme to oversee further development.

The vocabulary and mappings will be captured and made available in human-readable tabular form as well as in RDF/OWL.

It is hoped that further phases of the project will deliver:

- support for the automatic generation of term-to-term mappings between any pair of mapped schemes;
- an operational framework maintained on a web host, with mechanisms for additions to existing vocabularies and the addition of new vocabularies.

The RDA/ONIX framework categorizes resources in all media that could support the needs of libraries and the publishing industry, and builds on successful standards from the book publishing community (ONIX) and the library world (RDA: Resource Description and Access for the digital world, built on the foundation established for the Anglo-American Cataloguing Rules). It builds on the principles of interoperability established in the indecs Content Model.

Mark Bide, the Executive Director of EDItEUR (the global trade standards organisation supporting electronic commerce in the book and serials sectors, and responsible for the ONIX family of standards) said of the VMF: "We have every reason to be grateful to both the JISC and the IDF for providing the resources to make this project possible. There has been an explosion of metadata standards in the media over the last 15 or 20 years. The requirement for metadata and identifier interoperability across the boundaries between different media sectors and between schemes developed within individual sectors to meet different demands was recognised in the <indecs> project very nearly a decade ago. Now, as we approach the end of the first decade of the 21st Century, the growth in trading in content online, and the reality of convergence between the media, has created a sense of real urgency in the search for sophisticated interoperability solutions. The Vocabulary Mapping Framework is an essential foundation stone in that development."

The Committee of Principals of RDA welcomed the VMF project with this statement: "Semantic interoperability is a precondition to the efficient reuse of metadata across sectors. The *RDA/ONIX Framework for Resource Description* created a mechanism to facilitate transfer of resource description metadata between two communities. This initiative will not only extend the Framework to encompass other communities, but also broaden its scope by mapping additional vocabularies. JISC and IDF are to be commended for their support of this important initiative."

Further information:

Godfrey Rust, Rightscom (godfrey.rust@rightscom.com)  
 Gordon Dunsire, University of Strathclyde (g.dunsire@strath.ac.uk)  
 Norman Paskin, IDF (n.paskin@tertius.ltd.uk)  
 Alastair Dunning, JISC (a.dunning@jisc.ac.uk)

## References

**CIDOC**<sup>1</sup> is the international focus for the documentation interests of museums and similar organisations. The CIDOC Conceptual Reference Model (CRM) is an ISO standard which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

The Dublin Core Metadata Initiative (**DCMI**)<sup>2</sup> is an open organization engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models.

The Digital Data Exchange (**DDEX**)<sup>3</sup> was established to develop a single set of standard XML messages for the business-to-business communication of information between organisations operating in the digital content value chain, principally for recorded music.

The Digital Object Identifier (**DOI**®) System<sup>4</sup> is a managed system for persistent identification of content on digital networks. It can be used to identify physical, digital, or abstract entities.

Functional Requirements for Bibliographic Records (**FRBR**)<sup>5</sup> is a conceptual entity-relationship model developed by the International Federation of Library Associations and Institutions (IFLA) that relates user tasks of retrieval and access in online library catalogues and bibliographic databases from a user's perspective.

The <**indecs**> project<sup>6</sup> was created to address the need, in the digital environment, to put different creation identifiers and their supporting metadata into a framework where they could operate side by side, especially to support the management of intellectual property rights.

The **JISC**<sup>7</sup> is a joint body of the UK Higher and Further Education Funding Councils. It was established to support the further and higher education sectors in the use of information communications technology (ICT). It does this by providing a network service and by funding ICT development and services that support learning, teaching and research.

The IEEE's Learning Object Metadata (**LOM**)<sup>8</sup> is an international standard way of describing reusable learning resources, sometimes known as "learning objects".

**MARC 21**<sup>9</sup> formats are standards for the representation and communication of bibliographic and related information in machine-readable form. The MARC 21 formats are maintained by the Library of Congress in consultation with various user communities.

The **ONIX** for Books Product Information Message<sup>10</sup> is the international standard for representing and communicating book industry product information in electronic form, and is the most established of a group of ONIX standards which also cover serial publications and publishing licensing terms.

Resource Description and Access (**RDA**)<sup>11</sup> is a standard for bibliographic resource description and access designed for the digital world providing a comprehensive set of guidelines and instructions on resource description and access covering all types of content and media.

The **RDA/ONIX** Framework is a vocabulary matrix developed by a joint working group in late 2006 as a tool for creating well-formed and interoperable vocabularies for the library and content-provider communities.

**RDF and OWL**<sup>12</sup> are Semantic Web standards that provide a framework for asset management, enterprise integration and the sharing and reuse of data on the Web.

**Rightscom**<sup>13</sup> is a consultancy that specialises in the provision of solutions for the management, trading and protection of digital content and intellectual property rights in the network environment.

<sup>1</sup> <http://cidoc.mediahost.org>

<sup>2</sup> <http://dublincore.org>

<sup>3</sup> <http://www.ddex.net>

<sup>4</sup> [http://www.doi.org/about\\_the\\_doi.html](http://www.doi.org/about_the_doi.html)

<sup>5</sup> <http://www.ifla.org/en/frbr-rg>

<sup>6</sup> [http://www.doi.org/topics/indecs/indecs\\_framework\\_2000.pdf](http://www.doi.org/topics/indecs/indecs_framework_2000.pdf)

<sup>7</sup> <http://www.jisc.ac.uk>

<sup>8</sup> <http://ltsc.ieee.org/wg12/par1484-12-1.html>

<sup>9</sup> <http://www.loc.gov/marc>

<sup>10</sup> <http://www.editeur.org/onix.html>

<sup>11</sup> <http://www.rdaonline.org>

<sup>12</sup> <http://www.w3.org/2004/01/sws-pressrelease>

<sup>13</sup> [www.rightscom.com](http://www.rightscom.com)