

[Search](#) | [Back Issues](#) | [Author Index](#) | [Title Index](#) | [Contents](#)

D-Lib Magazine January/February 2007

Volume 13 Number 1/2

□ ISSN 1082-9873

Distinguishing Content from Carrier

The RDA/ONIX Framework for Resource Categorization

[Gordon Dunsire](#)

Centre for Digital Library Research, University of Strathclyde
<g.dunsire@strath.ac.uk>

Background

RDA: Resource Description and Access [1] is in development as a new standard for resource description and access designed for the digital world. It is being built on the foundation established for the Anglo-American Cataloguing Rules (AACR). Although it is being developed for use primarily in libraries, it aims to attain an effective level of alignment with the metadata standards used in related communities such as archives, museums and publishers, and to provide a better fit with emerging database technologies. A specific focus of RDA is the description of elements of the content and carrier of a resource that will help users to identify and select the resource to meet their needs with respect to the form of content, subject, volatility, etc., on the one hand, and the physical characteristics of the carrier, the formatting and encoding of the information, etc., on the other.

ONIX (Online Information Exchange) [2] is a standard for the use of publishers in distributing digital metadata about their products. The *ONIX Books Code Lists* [3] is a standard set of codes to be used in the metadata, including elements describing product content and carrier. The code lists are under constant review and development to meet the emerging needs of the publishing community.

Discussions in October 2005 between the Joint Steering Committee for Revision of AACR (JSC) and representatives of the publishing industry in the UK identified the categorization of resources by content and carrier as being of mutual interest with substantial potential benefit to be gained through cooperation. A joint initiative was subsequently funded by the organizations sponsoring the development of RDA and ONIX, with additional support from the British Library. The initiative aimed to develop a framework for categorizing resources in all media that could support the needs of libraries and the publishing industry, and would facilitate the transfer and use of resource description data across the two communities.

Methodology

The British Library hosted a two-day workshop in London in March 2006. The participants comprised the editor of RDA (Tom Delsey) and two consultants to EDItEUR (David Martin and Godfrey Rust), along with a facilitator (the author of

this article, who is also a member of the CILIP-BL Committee on AACR). The workshop produced an outline of the proposed framework that was subsequently developed to incorporate additional input and feedback, via email, from the participants and the communities they represent. The framework was tested against sample categories in RDA and ONIX, and some ad hoc examples from collection-level description and museum communities. Detailed discussion focussed on those aspects of categorization that are particularly appropriate for physical resources, and it is anticipated that the framework will undergo further stages of development.

The framework

Version 1.0 of the *RDA/ONIX framework for resource categorization* [4] was released in August 2006. The framework identifies and defines two distinct sets of attributes: one for the intellectual or artistic content of an information resource, and the other for the means and methods by which such content is carried. Closed sets of values are specified for some, but not all, of the attributes. The framework includes a methodology for constructing higher-level categories of resource content and carrier from the attribute and value sets, and recommendations on applying such categories to resource descriptions.

The attribute set for content includes:

- Character (language; music; image; other)
- SensoryMode (sight; hearing; touch; taste; smell; none)
- ImageDimensionality (two-dimensional; three-dimensional; not applicable)
- ImageMovement (still; moving; not applicable)
- Interactivity (interactive; non-interactive)
- CaptureMethod (*)
- ExtensionMode (succession; integration; not applicable)
- ExtensionTermination (determinate; indeterminate; not applicable)
- ExtensionRequirement (essential; inessential; not applicable)
- RevisionMode (correction; substitution; transformation; not applicable)
- RevisionTermination (determinate; indeterminate; not applicable)
- RevisionRequirement (essential; inessential; not applicable)
- Purpose (*)
- Subject (*)
- Form/Genre (*)

The parentheses contain the specified value sets for each attribute; * indicates that the value set is not specified and may be defined by user communities or by reference to a recognized namespace.

The attribute set for carrier includes:

- StorageMediumFormat (sheet; strip; roll; disc; sphere; cylinder; chip; file server)
- HousingFormat (binding; flipchart; reel; cartridge; cassette; not applicable)
- BaseMaterial (*)
- AppliedMaterial (*)
- FixationMethod (*)
- FixationTool (*)
- EncodingFormat (*)
- Generation (first; reproduction)
- IntermediationMethod (*)

- IntermediationTool (microform reader; microscope; projector; stereoscope; audio player; audiovisual player; computer; not required)

The framework allows user communities to define sub-values for any of the specified, so-called primary values. For example, the sub-values "regular" and "irregular" might be defined for the value "succession" in the ExtensionMode attribute. An obvious restriction is that each sub-value must belong to one, and only one, of the primary values.

Each of the attributes and primary values has a definition in Appendix A of the framework. It is therefore possible for a user community to assign their own terms for the values, including non-English translations and coded notations, but only if semantic equivalence is maintained.

Basic higher-level content and carrier categories are constructed by taking a single primary value from one or more attributes of the content and carrier attribute sets respectively. That is, a basic content category is defined by primary values exclusively from the content attribute set, and a basic carrier category is similarly defined by values from the carrier attribute set. For example, the basic content category defined by `<Character="image" + SensoryMode="sight" + ImageDimensionality="two-dimensional" + ImageMovement="moving">` is equivalent to code 06 of ONIX list 81, which has the descriptive label "moving images". The basic carrier category defined by `<StorageMediumFormat="file server" + HousingFormat="not applicable" + IntermediationTool="computer">` is equivalent to code DH of ONIX list 7, with the descriptive label "online resource".

Because the primary value sets are closed, the number of possible basic categories is finite, although some combinations of primary values are unlikely to be useful in practice. For example, it is difficult to think of any resource being described by a basic content category which includes `<ImageDimensionality="two-dimensional" + "ImageMovement"="not applicable">`. Some combinations are ruled out by definition; `<Character="music" + ImageDimensionality="two-dimensional">` is an invalid basic category because the value "not applicable" must be assigned to ImageDimensionality if the value for Character is not "image".

Basic categories can be qualified by adding values that are not specified by the framework. These include values for attributes for which no primary values are specified, as well as sub-values of primary values. As with basic categories, qualified categories must contain either content or carrier attribute values, but not both. Qualified categories allow the framework to be extended to suit the needs of specific user communities.

The framework allows user communities to choose their own labels for resource categories. A label may be descriptive or derived from codes assigned to the constituent values of the category. For example, the ONIX label for "online resource" is equivalent to "867" using the attribute order and primary value codes from Appendix D of the framework; for machine-to-machine (m2m) interoperability, the ONIX code 7DH translates to Framework code 867. It is important to note that the categories and codes given in appendices C (basic content categories) and D (basic carrier categories) of the framework are illustrative, and only include attributes targeted by the workshop group for the first release of the framework; they are not intended to be definitive and the Framework codes cited in this article (indicated by capital F) are informal examples.

Note that this approach of using attribute citation order rather than attribute codes requires that basic categories use a primary value from every available attribute to act as a place-holder. This is one reason why the framework provides primary values for "not applicable" or "not required". If these values are omitted from base categories, m2m interoperability requires the attribute type to be encoded as well as the primary values. Encoding attribute types allows the framework to be used by communities that choose to ignore a specific basic attribute or that cannot supply a primary value for one because of, say, the availability of legacy metadata.

More than one category may be applicable to a particular resource being described. Many user communities will want to assign at least one content category and one carrier category in an item-level description. If the item is an online text document, the base content category will be <Character="language" + SensoryMode="sight" + ImageDimensionality="not applicable" + ImageMovement="not applicable"> (category label "text" or Framework code 1133) and the base carrier category will be "online resource" or Framework code 867. However, some communities may not need both content and carrier to be described for every resource. For example, if the resource is a collection, either the item content may not be specified, such as in a collection of online resources, or the item carrier may not be specified, such as in a collection of moving images.

The framework allows multiple resource categories to be assigned to a resource, but it does not require this to be exhaustive. Communities may choose to assign only one content and one carrier category, irrespective of how many are applicable. The framework provides a set of "applicability" values that can be used to indicate the relative level or extent of the resource to which a particular category is applied.

These values are:

□

- All
- Predominant
- Substantial
- Some
- None

A community interested in a single type of content, say music, may wish to assign content categories which include <Character="music"> to every resource which contains music, even if it is a minor part of the resource, and can ignore all other types of content. The applicability values, if assigned, allow that community to share effective metadata with another community that allows users to identify resources containing significant (i.e., substantial, predominant, or all) content of their required character type.

The "none" value is included to cover situations where resources need to be assigned a fixed set of categories.

Beyond RDA and ONIX

The DOI community is actively engaged in ensuring compatibility of the framework with the standard currently under development by ISO, the International Organization for Standardization, as *ISO/CD 26324: Information and documentation -- Digital object identifier (DOI) system* [5].

Some ad hoc application of the framework to other metadata communities has been carried out. Note that these are the author's suggested mappings; the communities

themselves have not been involved.

The *Dublin Core Collection Description Application Profile* [6] uses the `clد:itemType` property to describe item content, and specifies the *DCMI Type Vocabulary* [7] as its namespace. The term name "MovingImage" from this namespace is equivalent to the basic content category with Framework code 3112 and ONIX code 8106. However, not all values in the namespace can be mapped completely to the four targeted content attributes in the framework and some require qualified categories; the minimum framework content categories for DCMI types are:

- Collection = <ExtensionMode="succession">
- Dataset = <Character="other">
- Event = <Subject="*event">
- Image = <Character="image">
- InteractiveResource = <Interactivity="interactive">
- MovingImage = <Character="image" + ImageMovement="moving">
- PhysicalObject = <Character="image" + ImageDimensionality="three-dimensional">
- Service = <Subject="*service">
- Software = <Character="other">
- Sound = <SensoryMode="hearing">
- StillImage = <Character="image" + ImageMovement="still">
- Text = <Character="language" + SensoryMode="sight">

Event and Service require qualified categories. Note that item carriers in a collection are described using the property `clد:ItemFormat` and no namespace is specified.

The MIME top-level media types [8] can be mapped to basic content categories:

- text = <Character="language" + SensoryMode="sight">
- image = <Character="image">
- audio = <SensoryMode="hearing">
- video = <Character="image" + ImageMovement="moving">
- application = <Character="other">

MIME subtypes can be mapped to qualified carrier categories, specifically <EncodingFormat>, and might form the namespace defined by some communities for this attribute.

Examples of mappings of classes from the *CIDOC Conceptual Reference Model* [9] for facilitating the integration, mediation and interchange of heterogeneous cultural heritage information are:

- E36 Visual Item = <SensoryMode="sight">
- E37 Mark = <Character="language" + SensoryMode="sight">
- E38 Image = <Character="image" + SensoryMode="sight">

These mappings expose some semantic inconsistency in the namespaces of some of the communities, at least with respect to the framework.

Conclusion

The RDA/ONIX framework successfully attains its aims of supporting the needs of libraries and the publishing industry for categorising resources by their content and

carrier, and of facilitating interoperability between the metadata produced by those communities. It also improves resource discovery for users by clearly distinguishing those aspects of a resource which pertain to its content and those associated with the way the content is conveyed. This in turn meets the needs of users who are interested in a specific type of content, such as musical composition and performance, or a specific physical or digital format of a resource, such as PDF files, or both, such as PDF files of musical compositions or DVDs of musical performances.

Although the first version of the framework has paid particular attention to the description of physical resources, its applicability to digital resources is evident. The ability to assign multiple categories and indicate relative levels of applicability is likely to be of use in describing complex digital information objects. The framework is also extendable to other communities who wish to improve metadata interoperability with libraries and publishers, and may be a useful tool for identifying and resolving inconsistencies in the semantics of namespaces for content and carrier description.

It is worth noting that the ontological approach to facilitating metadata interoperability taken by the framework is similar to that used by communities with mutual and overlapping interests in areas such as licensing and rights management, such as *DDEX (Digital Data Exchange)* [10] and *ONIX for Licensing Terms* [11].

References

- [1] *RDA: Resource Description and Access : Prospectus*
<<http://www.collectionscanada.ca/jsc/rdaprospectus.html>>.
- [2] ONIX (Online Information Exchange)
<<http://www.bisg.org/onix/index.html>>.
- [3] *ONIX Books Code Lists*, Issue 6, July 2006
<http://www.editeur.org/codelists/ONIX_Code_Lists_Issue_6.PDF>.
- [4] *RDA/ONIX framework for resource categorization*, version 1.0, August 2006
<<http://www.collectionscanada.ca/jsc/docs/5chair10.pdf>>.
- [5] *ISO/CD 26324: Information and documentation -- Digital object identifier (DOI) system*
<<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=43506&scopelist=PROGRAMME>>.
- [6] *Dublin Core Collection Description Application Profile*
<<http://dublincore.org/groups/collections/collection-application-profile/>>.
- [7] *DCMI Type Vocabulary*
<<http://dublincore.org/documents/dcmi-type-vocabulary/>>.
- [8] *Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types*
<<http://www.isi.edu/in-notes/rfc2046.txt>>.
- [9] Definition of the *CIDOC Conceptual Reference Model*
<http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.pdf>.
- [10] *DDEX: Digital Data Exchange*

<<http://www.ddex.net/index.htm>>.

[11] *ONIX for Licensing Terms*

<http://www.editeur.org/onix_licensing.html>.

Copyright © 2007 Gordon Dunsire



[Top](#) | [Contents](#)



[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)



[Previous Article](#) | [First Opinion](#)



[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)



doi:10.1045/january2007-dunsire