

Creating Metadata Records

Introduction

For the purposes of bibliographic control, the representations of information objects are derived from statements that describe the prominent features of those objects; these descriptions are referred to as metadata. Metadata elements are usually collected and arranged according to a specified format, or schema, to create a metadata record, which serves as a useful aid in the organisation and retrieval of information, especially networked information resources. Ultimately, then, a metadata record is a set of assertions about an information object structured in prescribed way that can be stored and accessed effectively within an information organisation system.

This report will address various issues involved in the process of creating metadata records to describe each of a group of five different resources. It will also describe the differences between the two schemata used to create these records—the Electronic Resource Citation (ERC) schema and the Dublin Core (DC) Metadata Elements—as well as the particular challenges presented by each.

The five resources described by the metadata records created under each schema are all electronic resources available on the Web:

- Introduction to Metadata: Pathways to Digital Information. Murtha Baca. 1998.
(http://www.getty.edu/research/conducting_research/standards/intrometadata/)
- The Office for Library and Information Networking. Metadata. (<http://www.ukoln.ac.uk/metadata/>)
- ANSI/NISO Z39.85—2001 Dublin Core Metadata Element Set. 2001.
(<http://www.niso.org/standards/resources/Z39-85.pdf>)
- XML Organic Bibliographic Information Schema. [Website.] 2003.
(<http://laneweb.stanford.edu:2380/wiki/medlane/schema>)
- Metadata Harvesting and the Open Archives Initiative. Clifford A. Lynch. 2001.
<http://www.arl.org/newsltr/217/mhp.html>

Electronic Resource Citation Records

A metadata record was created for each of the resources using the ERC schema; these records are included below:

Introduction to Metadata: Pathways to Digital Information. Murtha Baca. 1998.

erc:

who/edited: Baca, Murtha

what: Introduction to Metadata: Pathways to Digital Information

when: 2000

where: http://www.getty.edu/research/conducting_research/standards/intrometadata/

erc-contributor:

who: Gilliland-Swetland, Anne

what: Setting the Stage

where: http://www.getty.edu/research/conducting_research/standards/intrometadata/2_articles/index.html

erc-contributor:

who: Gill, Tony

what: Metadata and the World Wide Web

where:

http://www.getty.edu/research/conducting_research/standards/intrometadata/2_articles/gill/index.html

erc-contributor:

who: Woodley, Mary

what: Crosswalks: The Path to Universal Access?

where:

http://www.getty.edu/research/conducting_research/standards/intrometadata/2_articles/woodley/index.html

erc-support:

who: J. Paul Getty Trust | Getty Standards and Digital Resource Management Program

where: http://www.getty.edu/research/conducting_research/standards/index.html

erc-rights:

who: J. Paul Getty Trust

what: Copyright

when/copyrighted: 2000

where: <http://www.getty.edu/legal/copyright.html>

The Office for Library and Information Networking. Metadata.

erc:

who: Day, Michael | Powell, Andy

what: UKOLN Metadata

when/last modified: 2005 01 19

where: <http://www.ukoln.ac.uk/metadata/> | <http://purl.org/net/ukoln/metadata> |

<http://purl.eu.org/net/ukoln/metadata>

erc-about:

what: Metadata | UKOLN (UK Office for Library Networking)

erc-support:

who: UKOLN, University of Bath

where: <http://www.ukoln.ac.uk/ukoln/>

ANSI/NISO Z39.85—2001 Dublin Core Metadata Element Set. 2001

erc:

who: National Information Standards Organization

what: The Dublin Core Metadata Element Set : An American National Standard

when: 2001 09 10

where: <http://www.niso.org/standards/resources/Z39-85.pdf>

erc-about:

what: Dublin Core Metadata Scheme | Metadata standards | Information organization

XML Organic Bibliographic Information Schema. [Website.] 2003.

erc:

who: Medlane Project - Lane Medical Library, Stanford University Medical Center

what: The Medlane Project: schema | XOBIS: The XML Organic Bibliographic Information Schema

when/created: 2002

where: <http://laneweb.stanford.edu:2380/wiki/medlane/schema>

erc-about:

what: XML | Extensible Markup Language | XML schemas for modelling MARC data | MARC records | Machine Readable Cataloging

erc-contributor:

who/documentated: Miller, Dick R. | Clarke, Kevin S.

Metadata Harvesting and the Open Archives Initiative. Clifford A. Lynch. 2001

erc:

who: Lynch, Clifford A.

what: Metadata Harvesting and the Open Archives Initiative

when: 2001 08

where: <http://www.arl.org/newsltr/217/mhp.html>

in: ARL Bimonthly Report 217 | 2001 08 | p1-9

erc-about:

what: Metadata | Metadata Harvesting Protocol | Networked information services | Open Archives
Metadata Harvesting Protocol | Information Organization**Dublin Core Records**

A metadata record was created for each of the resources using the DC Metadata elements; these records are included below:

Introduction to Metadata: Pathways to Digital Information. Murtha Baca. 1998.

```
<META NAME="DC.Title" LANG="en" CONTENT="Introduction to Metadata: Pathways to Digital
Information (Getty Research Institute)">
<META NAME="DC.Creator" LANG="en" CONTENT="Baca, Murtha (editor).">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata.">
<META NAME="DC.Subject" LANG="en" CONTENT="Database management.">
<META NAME="DC.Description" LANG="en" CONTENT="Contains articles that provide an introduction
to and overview of metadata. Includes a listing of metadata standards crosswalks, a glossary, a list of
acronyms and URLs about metadata, and details about the contributors">
<META NAME="DC.Description.tableOfContents" LANG="en" CONTENT="Introduction; Articles;
Crosswalks; Glossary; Acronyms & URLs; Contributors">
<META NAME="DC.Publisher" LANG="en" CONTENT="Getty Research Institute for the History of Art
and the Humanities.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Gill, Tony.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Gilliland-Swetland, Anne.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Woodland, Mary.">
<META NAME="DC.Date" LANG="en" CONTENT="2000">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Type" LANG="en" CONTENT="interactive">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en" CONTENT="J. Paul Getty Trust
http://www.getty.edu/legal/copyright.html">
```

The Office for Library and Information Networking. Metadata.

```
<META NAME="DC.Title" LANG="en" CONTENT="UKOLN Metadata">
<META NAME="DC.Title.alternative" LANG="en" CONTENT="Metadata">
<META NAME="DC.Creator" LANG="en" CONTENT="Day, Michael">
<META NAME="DC.Creator" LANG="en" CONTENT="Powell, Andy">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="UKOLN (UK Office for Library Networking)">
<META NAME="DC.Description" LANG="en" CONTENT="Includes lists of projects, resources, tutorials,
initiatives, and registries related to the creation and application of metadata. Provides links to URLs and
definitions of metadata">
<META NAME="DC.Publisher" LANG="en" CONTENT="UKOLN, University of Bath">
<META NAME="DC.Date.Modified" LANG="en" CONTENT="2005-01-19">
```

```
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.ukoln.ac.uk/metadata/">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://purl.org/net/ukoln/metadata">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://purl.eu.org/net/ukoln/metadata">
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

ANSI/NISO Z39.85—2001 Dublin Core Metadata Element Set. 2001

```
<META NAME="DC.Title" LANG="en" CONTENT="The Dublin Core Metadata Element Set : An
American National Standard">
<META NAME="DC.Creator" LANG="en" CONTENT="National Information Standards Organization">
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core Metadata Scheme">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata standards">
<META NAME="DC.Subject" LANG="en" CONTENT="Information organization">
<META NAME="DC.Description.Abstract" LANG="en" CONTENT="Defines fifteen metadata elements for
resource description in a crossdisciplinary information environment">
<META NAME="DC.Description.tableOfContents" LANG="en" CONTENT="Foreword; 1. Scope and
Purpose; 2. Referenced Standards; 3. Definitions; 4. The Element Set; 5. The Elements; Appendix A.
Further Reading; Appendix B. Maintenance Agency.">
<META NAME="DC.Publisher" LANG="en" CONTENT="NISO Press">
<META NAME="DC.Contributor" LANG="en" CONTENT="American National Standards Institute">
<META NAME="DC.Date" LANG="en" CONTENT="2001-09-10">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.niso.org/standards/resources/Z39-
85.pdf">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsPartOf National information standards series">
```

XML Organic Bibliographic Information Schema. [Website.] 2003.

```
<META NAME="DC.Title" LANG="en" CONTENT="The Medlane Project: schema">
<META NAME="DC.Title.alternative" LANG="en" CONTENT="XOBIS: The XML Organic Bibliographic
Information Schema">
<META NAME="DC.Creator" LANG="en" CONTENT="Medlane Project - Lane Medical Library, Stanford
University Medical Center">
<META NAME="DC.Subject" LANG="en" CONTENT="XML">
<META NAME="DC.Subject" LANG="en" CONTENT="Extensible Markup Language">
<META NAME="DC.Subject" LANG="en" CONTENT="XML schemas for modelling MARC data">
<META NAME="DC.Subject" LANG="en" CONTENT="MARC records">
<META NAME="DC.Subject" LANG="en" CONTENT="Machine Readable Cataloging">
<META NAME="DC.Description" LANG="en" CONTENT="Documents XOBIS an XML schema for
modelling MARC data developed by members of the Medlane Project at Lane Medical Library, Stanford
University. Includes contents in .pdf and .html formats">
<META NAME="DC.Description.tableOfContents" LANG="en" CONTENT="Introduction; Related Efforts
in the Community; There's No Business, Like XOBIS: Background Material; XOBIS Root Element and
General Organization; Principal Elements and Core Structure; Generic Elements; Attributes; Individual
Principal Elements; Relationships; Indexing Implications; Conclusion; References; Acknowledgements;
Disclaimer">
<META NAME="DC.Publisher" LANG="en" CONTENT="Lane Medical Library, Stanford University">
<META NAME="DC.Contributor" LANG="en" CONTENT="Miller, Dick R.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Clarke, Kevin S.">
<META NAME="DC.Date.created" LANG="en" CONTENT="2002-04 to 2002-09">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
```

```
<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://laneweb.stanford.edu:2380/wiki/medlane/schema">
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

Metadata Harvesting and the Open Archives Initiative. Clifford A. Lynch. 2001

```
<META NAME="DC.Title" LANG="en" CONTENT="Metadata Harvesting and the Open Archives
Initiative">
<META NAME="DC.Creator" LANG="en" CONTENT="Lynch, Clifford A.">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Harvesting Protocol">
<META NAME="DC.Subject" LANG="en" CONTENT="Networked information services">
<META NAME="DC.Subject" LANG="en" CONTENT="Open Archives Metadata Harvesting Protocol">
<META NAME="DC.Subject" LANG="en" CONTENT="Information Organization">
<META NAME="DC.Description" LANG="en" CONTENT="This article describes the Open Archives
Metadata Harvesting Protocol">
<META NAME="DC.Publisher" LANG="en" CONTENT="The Association of Research Libraries">
<META NAME="DC.Date" LANG="en" CONTENT="2001-08">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.arl.org/newsltr/217/mhp.html">
<META NAME="DC.Relation" LANG="en" CONTENT="IsPartOf ARL Bimonthly Report 217">
<META NAME="DC.Rights" LANG="en" CONTENT="ARL policy is to grant blanket permission to reprint
any article in the Bimonthly Report for educational use as long as full attribution is made. Exceptions to
this policy may be noted for certain articles. This is in addition to the rights provided under sections 107
and 108 of the Copyright Act. For commercial use, a reprint request should be sent to ARL Director of
Information Services, Julia Blixrud <jblix@arl.org>.">
```

Discussion

Despite the fact that there is a wide range of metadata schemata in use today within various information communities, some commonalities still can be found among them. Perhaps chief among these is the practice of designing and structuring metadata records in such a way as to be machine-readable or computer processable. To this end, the data are labelled, arranged, and presented according to a prescribed structure so that a machine or computer can read and act upon the data efficiently. This structure is known as the schema's syntax.

Metadata schemata are built from one or more basic structures called elements, each of which is intended to identify or label a particular kind of data. For example, the ERC schema applies the "who" element to the name of a person or party responsible for the resource being described; Dublin Core has a similar element, known in that schema as "Creator". In either case, the element serves as the label for a particular kind of attribute, so that attribute is treated the same way consistently across all records. The elements in a metadata schema are defined by semantics that indicate what kind of attributes of an information object are meant to be associated with each element, making it possible to associate the proper data with the corresponding element. Not only does the structure make the metadata records machine-processable by allowing a computer to display the right data with the right label, it also serves the human creator of the records by allowing him to associate a descriptive attribute with the correct element.

While the ERC and DC schemata both spring from a common spirit and consensus—that metadata records should be simple, flexible, and semantically interoperable across domains—and often are used to describe the same kinds of information objects, each has developed and defined its own corresponding set of elements, semantics, and encoding systems. DC was developed to be a relatively uncomplicated system for resource description that could be used across many diverse disciplines by individuals who lack a formal cataloguing background. In turn, ERC has been proposed as an even simpler, streamlined

format for organising and presenting metadata that aims at being even more compact and extensible than Dublin Core.

Creating ERC records

At the centre of the ERC metadata schema are four fundamental elements that form a basis for a minimal, uniform object description that can be applied to any information object—or any object in the universe, for that matter, at least theoretically. These four elements and their semantics are:

1. who—a responsible person or party
2. what—a name or other human-oriented identifier
3. when—a date important in the object's lifecycle
4. where—a location or machine-oriented identifier

In order for an ERC record to be valid, each of these elements must be included in this order, and a value must be assigned to each. Appearing at the beginning of an ERC record, these four elements alone form the record's anchoring story, which serves as the most basic, general description of a resource.

However, other stories that describe an object may also be included in an ERC record; these stories may apply to more specific aspects of an object's description, such as its intellectual content, a statement of commitment of its support, or its provenance. The addition of these stories is achieved through the use of segments that include one or more of the four basic elements, with the meaning of each changed so as to apply to the segment. For example, in an "about" segment, the element "who" would describe a person that the information object is about, and the "what" element would describe the subject of the information object. These basic elements may also be qualified with the addition of the stroke character (/) and a qualifying term in order to add more explanation to the basic element; an example of the use of a qualified element is found in the ERC record above for the resource titled Introduction to Metadata: Pathways to Digital Information: in the `erc:rights` segment of the record, the "when" element is qualified by the addition of "copyrighted" to indicate that the important date referred to in this segment is when the resource was copyrighted (e.g. `when/copyrighted: 2000`).

The basic structure of an ERC record is as follows: the record begins with the label "erc:", followed by the labels of the four elements of the anchoring story—who: , what: , when:, and where: , and their corresponding values. This would constitute the simplest form of an ERC record and would require nothing more to be valid. Any stories subsequent to the anchoring story would follow a similar format, beginning with an appropriate label for the segment, such as `erc-about:` or `erc-from:`, then any element labels and values following. It should be noted that only the anchoring story requires the inclusion of all four basic elements; any subsequent stories may include as few as one or all of four. (Please refer to the section headed "Electronic Resource Citations Records" above for examples of ERC records and their structure.)

As intended, the ERC schema is indeed quite simple and basic, as the amount of information required to create a valid ERC record is quite minimal. So long as the data associated with the elements is available from the resource, the record can be created quite easily and quickly. By allowing unknown or unsupplied data to be represented by one of the standard missing value codes, such as `(:unkn)` for an unknown value or `(:null)` for an explicitly empty value, a valid ERC record conceivably could contain nothing more than an anchoring story constructed from element labels with missing value codes. However, such a record would be of little value to an end user and would not provide any information to assist with the basic user tasks of finding, identifying, selecting, and obtaining a resource. The system of missing value codes does, however, provides a means for offering at least some kind of an explanation for any data that is not supplied. The extension mechanism of allowing additional story segments to be created to add to the four basic elements of the anchoring story is quite useful, as it provides a way to include more information to make for a richer, more useful record. For example, most of the ERC records shown above have a subject story, created by appending an "erc-about" segment. Mechanisms for extending a record to include copyright information or details about contributors were easy to construct, as they followed the same general structure and could be qualified as needed.

The issue of entity level did, however, present a challenge, as evidenced by the record for the resource titled Introduction to Metadata: Pathways to Digital Information. This resource—a Web site based on a

book edited by one individual that includes several articles by different authors—was somewhat problematic to create. While the most basic representation of this resource was the anchoring story that described the site as a whole, each article warranted a segment of its own as well, and had to be handled through the inclusion of an erc-contributor segment for each. If, however, it were to be decided that the entity level would be the individual articles rather than the entire site, each article would have to be represented in a single ERC record with its own anchoring story. Other issues involved in creating this record included finding the correct date (“when” element) and copyright (“erc-copyright:” segment) information associated with the Web site rather than the book. The decisions involved in assigning a description of intellectual content to each of the resources also involved careful consideration. In constructing “erc-about” segments for the records, there are no guidelines in the ERC schema’s semantics that dictate a controlled vocabulary to use for subject description. A record creator is completely free to use whatever terms and any depth of indexing he chooses, ultimately leaving the door open to inconsistency and arbitrary assignment of subjects. Therefore, while the ERC metadata schema may indeed be general and basic, this simplicity can prove to be somewhat deceptive.

Creating DC records

The Dublin Core is a type of descriptive metadata that is meant to serve as for information resource description. It was conceived as a way to create descriptions of resources that were simple, flexible, and applicable within and across domains in the interest of reducing the cost of metadata creation and increasing interoperability. In fact, DC is not intended to supplant or replace any particular metadata schema or standard, but rather to exist alongside them.

Like ERC, Dublin Core focuses on a set of basic elements, in its case fifteen elements, that were initially intended to facilitate the discovery of information objects. Early in its development, DC began to expand by way of adding extensions and qualifiers for the elements in the interest of being used to create more precise resource descriptions. A key feature of DC is its lack of any syntax, which was a deliberate decision intended to allow different communities to use its own encoding format. As a result, DC records may be created in HTML, XML, MARC DC format, as well as other formats.

The DC Element Set is made up of these fifteen elements: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights; these elements were developed by the Dublin Core Metadata Initiative (DCMI) and are formally defined in the document [The Dublin Core Metadata Element Set](#) (ANSI/NISO Z39.85-2001) by the National Information Standards Organization and approved by the American National Standards Institute. The work of the DCMI in refining and developing DC is ongoing.

In creating a DC record, all of the elements are optional, they may appear in any order, and they may be repeated. Many of the elements also may be extended with a qualifier. A DC record, therefore, may be as simple or as complex as the record’s creator chooses. This flexibility makes the creation of DC records quite easy. As was discussed regarding the ERC, one area of concern was the issue of entity level, especially vis-à-vis the resource titled [Introduction to Metadata: Pathways to Digital Information](#). A decision would have to be made whether the appropriate entity level for a resource such as this would be the site as a whole or the individual records. However, the DC record structure’s “Contributor” element does make the inclusion of the article authors’ names more straightforward, yet it does not address how to include the article titles. For the most part, however, DC’s elements and qualifiers do provide for a more precise description of the resource—especially the “Description” element, which allows for a full-text description and may be qualified to include a table of contents or an abstract. The semantics are more explicit, providing more guidelines for a record’s creator, which can be especially helpful for someone who has less experience working with descriptive metadata. Other elements, such as “Rights”, may accept longer strings of text, allowing the record to include a fuller description than would otherwise be possible. While attributes described in a few words (Creator, Title, Subject) may make finding an information resource easier, these elements that allow for longer descriptions can be especially useful for the tasks of identifying and selecting a resource. Essentially, the structure of DC records makes richer resource description possible.

Generic Metadata Editor

The ERC and DC metadata records listed at the beginning of this report were all created using a Generic Edit Tool. This editing tool provided templates for entering the data values; selecting an element provided a text box for entering the correct data. This was a helpful feature, especially for creating the DC records, as a pull-down menu was provided in some cases that allowed the record creator to select a value from a list. Text could be edited within a record so that the creator could add qualifiers, more text, and special characters as needed. In the case of the DC record editor, HTML tags were automatically supplied. The record creator who is less familiar with HTML mark-up certainly would appreciate this step, but to create DC records in another format, adjustments would be required.

Summary and Conclusion

Creating metadata records using the Electronic Resource Citation scheme and Dublin Core Metadata Elements provides the creator with valuable insight and experience in working with descriptive metadata. Each of these aims to provide a way for a user with little or no formal cataloguing experience to create simple, uniform metadata records for information resource description by defining a structure built upon basic elements that may be extended or qualified as needed. Both systems strive to use elements that are flexible enough to be applicable across subject domains in the interest of promoting interoperability. ERC, with its four basic elements and anchoring story, is a more basic metadata schema, but could benefit from more specific semantics. DC, with its fifteen elements and extensibility, allows for the development of more refined metadata records, and has the added flexibility of being encodable in a variety of formats, including HTML and XML. Both, however, continue to be confronted with some of the same challenges, such as entity level, the varying quality and reliability of available metadata for a given resource, and the varying objectives of different information communities.