Sara Adams
SLIS 5223
Spring 2005

Creating Metadata Records: Exploring Two Schemes

Introduction

Creating accurate and complete metadata records is imperative to users' abilities to locate the information packages that are most relevant to their specific searches. Accomplishing this task can be highly complex and often difficult due to several factors.

First, the lack of a single definitive standard for recording metadata results in a lack of consistency in the information recorded. Next, focus on the system database rather than user needs often leaves records incomplete or difficult to use. Last, the information packages themselves may lack organization and thus, make locating accurate information to input into the metadata records haphazard.

Over time, many approaches attempted to make information packages as accessible as possible. The Electronic Resource Citation Metadata Scheme (ERC) and the Dublin Core (DC) are two relatively new schemes under development. Several major points differentiate the two. ERC is being developed to encompass digital resources, while DC is designed for many kinds of media including print and digital.

In addition, ERC exemplifies simplicity with its elements "who," "what," "when," and "where" that are required and non-repeatable for the anchoring story. They must be listed in order. The anchoring story, as expressed by the ERC creator, John Kunze, is the basic set of entries of an information package that support the user needs of finding, identifying, selecting, and acquiring. Other stories may be added to further a user's understanding of an item's "aboutness," provenance, or other pertinent information. All stories are structured to include the four basic elements and to be easily read by both humans and machines. Syntax is simple and intuitive, allowing ERC to be highly versatile.

DC, on the other hand, is comprised of fifteen metadata elements, all of which are optional and repeatable and may be listed in any order. The elements include basic designations for creator, title, date, and identifier, but go on to more in-depth data such as publisher, contributor, and the like. Designed to be machine readable, semantics for each element are clearly stated while syntax is not. The purpose of DC's lack of syntax is to allow it to be adaptable to different organizations for different purposes.

The examples below demonstrate how ERC anchoring stories compare to DC records.

Electronic Resource Citation Records

ERC Record 1

```
who: Baca, Murtha | Gill, Tony | Gilleland-Swetland, Anne J. | Woodley, Mary
what: Introduction to Metadata Pathways to Digital Information
when: 2000
where:
http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html
```

ERC Record 2

```
who: UKOLN
what: Metadata
when: 2005 01 18
where: http://www.ukoln.ac.uk/metadata/
```

ERC Record 3

```
who: National Information Standards Organization
what: The Dublin Core Metadata Element Set
when: 2001
where: http://www.niso.org/standards/resources/Z39-85.pdf?CFID=179761&CFTOKEN=46984796
```

ERC Record 4

```
who: Lane Medical Library
what: The XML Organic Bibliographic Information Schema
when: 2003 10 22
where: http://laneweb.stanford.edu:2380/wiki/medlane/schema
```

ERC Record 5

```
who: Lynch, Clifford A.
what: Metadata Harvesting and the Open Archives Initiative
when: 2001 08
where: http://www.arl.org/newsltr/217/mhp.html
```

Dublin Core Records

DC Record 1

```
<META NAME="DC.Title" LANG="en" CONTENT="Introduction to Metadata (Getty Research Institute">
<META NAME="DC.Editor" LANG="en" CONTENT="Baca, Murtha">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata; Metadata mapping; metadata crosswalks; data standards; Dublin Core; web resources">
<META NAME="DC.Description"  LANG="en" CONTENT=" Professionals who are deeply involved in the development and implementation of information standards have contributed to this second version of Introduction to Metadata: one comes from academia, one from the cultural heritage information field, and the third is a practicing librarian.">
```

<META NAME="DC.Publisher" LANG="en" CONTENT="Getty Research Institute">
<META NAME="DC.Contributor" LANG="en" CONTENT="Baca, Murtha"><META
NAME="DC.Contributor" LANG="en" CONTENT="Gill, Tony">
<META NAME="DC.Contributor" LANG="en" CONTENT="Gilleland-Swetland, Anne J.">

```
<META NAME="DC.Contributor"  LANG="en" CONTENT="Woodley, Mary">
<META NAME="DC.Date Copyrighted" LANG="en" CONTENT="2001">
<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.getty.edu/research/conducting_research/standards/intrometadata/index.html">
<META NAME="DC.Source" LANG="en" CONTENT="">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="">
<META NAME="DC.Coverage"  LANG="en" CONTENT="Metadata">
<META NAME="DC.Rights" LANG="en"
CONTENT="http://www.getty.edu/legal/copyright.html">
```

DC Record 2

```
<META NAME="DC.Title" LANG="en" CONTENT="Metadata">
<META NAME="DC.Creator" LANG="en" CONTENT="UKOLN">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata; Projects; Resources;
Initiatives; Metadata Registries">
<META NAME="DC.Description"  LANG="en" CONTENT="The focus of the metadata
group at UKOLN is to review current approaches to resource description and to
look at future options for metadata in the wider context of resource
discovery.">
<META NAME="DC.Publisher" LANG="en" CONTENT="UKOLN">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Day, Michael">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Powell, Andy">
<META NAME="DC.Date Updated" LANG="en" CONTENT="2005-01-18">
<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.Source" LANG="en" CONTENT="University of Bath">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="">
<META NAME="DC.Relation" LANG="en"  CONTENT="References Metadata in a
Nutshell">
<META NAME="DC.Coverage"  LANG="en" CONTENT=" ">
<META NAME="DC.Rights" LANG="en"
CONTENT="http://www.ukoln.ac.uk/ukoln/privacy/">
```

DC Record 3

```
<META NAME="DC.Title" LANG="en" CONTENT="The Dublin Core Metadata Element
Set">
<META NAME="DC.Creator" LANG="en" CONTENT="National Information and Standards
Organization">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata elements; Resource
description">
<META NAME="DC.Description"  LANG="en" CONTENT="Defines fifteen metadata
elements for resource description in a cross-disciplinary information
environment.">
<META NAME="DC.Publisher" LANG="en" CONTENT="National Information Standards
Organization">
<META NAME="DC.Contributor"  LANG="en" CONTENT="American National Standards
Institute">
<META NAME="DC.Approved Date" LANG="en" CONTENT="2001">
```

```
<META NAME="DC.Type"  LANG="en" CONTENT="text">
<META NAME="DC.Format"  LANG="en" CONTENT="pdf">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.niso.org/standards/resources/Z39-
85.pdf?CFID=179761&CFTOKEN=46984796">
<META NAME="DC.Source" LANG="en" CONTENT="">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="">
<META NAME="DC.Coverage"  LANG="en" CONTENT="United States">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright 2001">
```

DC Record 4

```
<META NAME="DC.Title" LANG="en" CONTENT="Xobis the XML Organic Bibliographic
System">
<META NAME="DC.Creator" LANG="en" CONTENT="Lane Medical Library">
<META NAME="DC.Subject" LANG="en" CONTENT="XML schema; MARC">
<META NAME="DC.Description"  LANG="en" CONTENT="There are many XML schemas
available for modeling MARC data. Most take a literal approach, naming
elements and attributes after their corresponding MARC fields, subfields, and
indicators. Others represent only a small subset of the data libraries use to
describe resources. XOBIS attempts to walk the middle path: describe the full
set of library information, but reorganize this information into a structure
that empowers the use of library data as just one more information resource
available in the digital domain.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Medlane">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Miller, Dick"><META
NAME="DC.Contributor"  LANG="en" CONTENT="Buttner, Mary">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Clarke, Kevin S.">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Feng, Maria">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Grappone, Todd C.">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Li, Ying">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Wang, Jo">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Wesley, Rebecca">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Woelfel, Randy">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Murnane, Pamela M.">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Yates, Charles">
<META NAME="DC.Contributor"  LANG="en" CONTENT="Zhu, Danjin">
<META NAME="DC.Date Last Edited" LANG="en" CONTENT="2003-10-22">
<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://laneweb.stanford.edu:2380/wiki/medlane/schema">
<META NAME="DC.Source" LANG="en" CONTENT="">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en"  CONTENT="Is PartOf The Medland Project">
```

DC Record 5

```
<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="Open Archives Metadata Harvesting
Protocol">
<META NAME="DC.Description"  LANG="en" CONTENT=" This article describes the
Open Archives Metadata Harvesting Protocol, an important new infrastructure
component for supporting distributed networked information services. The
```

```
Metadata Harvesting Protocol—a mechanism that enables data providers to
expose their metadata—is seeing very rapid deployment, and enables a
fascinating array of new services and system architectures for a diverse set
of communities. I will speculate about some of these services and discuss
issues involved in their development. This article is not intended to be a
definitive technical summary of the protocol; documents providing such a
discussion can be found at <http://www.openarchives.org/>. Rather, the focus
here is on the uses of the protocol and its strategic significance as an
enabling technology.">
<META NAME="DC.Publisher" LANG="en" CONTENT="ARL Bimonthly Report 217">
<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.Source" LANG="en" CONTENT="">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsPartOf ARL Bimonthly Report
217">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright 2001">
<META NAME="DC.Date" LANG="en" CONTENT="2001-08">
```

Discussion

In creating the metadata records above, several strong points and weaknesses were revealed.

Strong points include the ease of use of the Generic Metadata Editor software. For DC, its simple fill-in-the-blank design allows the data creator to focus on the values to be entered, rather than the html structure of the input. For ERC, the Generic Metadata Editor assisted in maintaining the proper and consistent order of the elements, though the order is so intuitive as to require little or no assistance.

Another strong point relating directly to DC is the guidance the element semantics lend to the data creator to locate and differentiate the appropriate information. For example, elements that clearly define where to place publisher, creator, and contributor elements eliminate some of the confusion associated with making distinctions between these three “who” entities. In addition, a conscientious data creator will follow each DC element carefully, resulting in very complete records without the need to formulate additional stories, as with ERC.

ERC’s strong points relate to its simplicity and intuitive nature, as previously mentioned. Following the simple “who,” “what,” “when,” and “where” sequence could not be easier to follow. Its extensibility by employing additional stories according to user needs makes the scheme adaptable to virtually any organization or database project. In addition, the codes provided to explain information that does not exist or is unavailable clarify to the end user that elements are not simply omitted by accident or laziness.

By the same token, one of DC’s weaknesses is its allowance of any or all elements to be optional. This can create a situation where vital elements are omitted without offering the user any explanation to an element’s absence.

In general, the difficulty in identifying quality metadata elements exists for all information packages. A data creator must carefully examine information packages to locate accurate information. If it is not readily apparent, the creator must attempt to find it through other channels, such as viewing the page source. Taking information from the page source can be questionable if the information it provides cannot be corroborated elsewhere in the document. In addition, it can create conflicts between the data creator's opinions of what is important to his particular user group and the reasons for a page creator's labels. In ERC and DC Record 2, for example, the page source lists the creators as Michael Day and Andy Powell. The information item itself, however, emphasizes UKOLN as the creator instead. It is a matter of individual judgment based upon knowledge of the user group that the data creator must make.

Summary and Conclusion

The search for consistency in creating accurate and complete metadata records is an ongoing one. Its value is demonstrated by the sheer number of initiatives that continue to explore this issue, exemplified in this report by both ERC and DC.

ERC and DC share the qualities of extensibility by allowing flexibility for different organizations and databases. ERC tends to operate in a more intuitive fashion, while DC encourages very complete records with its specificity in identifying elements. DC can function as a thorough guide for those less versed in creating metadata records, and it is possible that record creators using ERC could benefit from such a guide as they develop different stories for individual documents.

Unfortunately, these, nor any other schemes, can solve the dilemma of identifying valuable metadata in every information package. This is left up to the data creator and his knowledge of the user group.