

Assignment: Creating Metadata, Spring 2005

Introduction

This document contains ten metadata records utilizing the Electronic Resource Citation (ERC) Scheme and the Dublin Core (DC) Simple Scheme. Five information objects were analyzed and one record for each type of scheme (two records apiece) is contained within the body of this document. Through this assignment the following goals were met:

- Used a basic metadata creation tool;
- Created a set of records using the Electronic Resource Citation (ERC) metadata scheme;
- Created a set of records using the Dublin Core (Simple) metadata scheme;
- Identified issues related to metadata creation;
- Explored issues of metadata quality.

Electronic Resource Citation Records

ERC Record 1

erc:
who: Dublin Core Metadata Initiative
what: Answers to frequently asked questions about the Dublin Core Metadata Initiative and Dublin Core.
when: 2003
where: <http://dublincore.org/resources/faq/index.shtml>

ERC Record 2

erc:
who: Moen, William E.
what: Resource discovery using Z39.50: promise and reality
when: 2001-01-23
where: http://www.loc.gov/catdir/bibcontrol/moen_paper.html

ERC Record 3

erc:
who: National Information Standards Organization [NISO]
what: Understanding metadata [booklet]
when: 2004
where: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

ERC Record 4

erc:
 who: NISO AX Committee for the OpenURL
 what: Registry for the OpenURL Framework website
 when: 2003-02-20
 where: <http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=ListSets>
 erc-about:
 who: OCLC
 what: interoperability protocol | repository interfaces | open-linking environment
 when: 2001-2002
 where: <http://www.oclc.org/research/projects/openurl/default.htm>
 erc-from:
 who: OCLC
 what: ALCME
 when: 2000-2002
 where: <http://www.oclc.org/research/projects/archive/default.htm>

ERC Record 5

erc:
 who: Distributed Systems Technology Centre (DSTC)
 what: Metadata website
 when: 2004
 where: <http://www.metadata.net>

Dublin Core Records**DC Record 1**

```
<META NAME="DC.Title" LANG="en" CONTENT="DCMI Frequently Asked Questions (FAC)">
<META NAME="DC.Title" LANG="en" CONTENT="Dublin Core Metadata Initiative Frequently Asked Questions">
<META NAME="DC.Description" LANG="en" CONTENT="Answers to Frequently Asked Questions about the Dublin Core Metadata Initiative and Dublin Core Metadata.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Dublin Core Metadata Initiative.">
<META NAME="DC.Date" LANG="en" CONTENT="2002-10-31">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="information classification">
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core">
<META NAME="DC.Subject" LANG="en" CONTENT="Cataloging of computer networked resources">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://dublincore.org/resources/faq/index.shtml">
```

DC Record 2

```

<META NAME="DC.Title" LANG="en" CONTENT="Resource Discovery Using Z39.50:
Promise and Reality">
<META NAME="DC.Creator" LANG="en" CONTENT="Moen, William E.">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="Z39.50 protocol">
<META NAME="DC.Subject" LANG="en" CONTENT="bibliographic control">
<META NAME="DC.Subject" LANG="en" CONTENT="web resources">
<META NAME="DC.Subject" LANG="en" CONTENT="intersystem communication">
<META NAME="DC.Description" LANG="en" CONTENT="This paper provides a
portrayal of Z39.50 that explains its flexibility in response to a variety of
information requirements in the networked environment">
<META NAME="DC.Date" LANG="en" CONTENT="2001-01-23">
<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.loc.gov/catdir/bibcontrol/moen_paper.html">
<META NAME="DC.Publisher" LANG="en" CONTENT="Library of Congress Cataloging
Directorate">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" http://www.loc.gov/catdir/bibcontrol/Paper
delivered at the Conference on Bibliographic Control in the New Millennium
[Library of Congress]">

```

DC Record 3

```

<META NAME="DC.Title" LANG="en" CONTENT="Understanding Metadata">
<META NAME="DC.Creator" LANG="en" CONTENT="National Information Standards
Organization (NISO)">
<META NAME="DC.Subject" LANG="en" CONTENT="Resource discovery">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="information organization">
<META NAME="DC.Description" LANG="en" CONTENT="NISO's general introduction
to metedata">
<META NAME="DC.Publisher" LANG="en" CONTENT="NISO Press">
<META NAME="DC.Date" LANG="en" CONTENT="2004">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">
<META NAME="DC.Identifier" LANG="en" CONTENT="ISBN: 1880124629">
<META NAME="DC.Source" LANG="en"
CONTENT="http://www.niso.org/standards/resources/UnderstandingMetadata.pdf">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" Metadata made simpler: a guide for
librariesRevision of 2001 booklet published by NISO Press">
<META NAME="DC.Rights" LANG="en" CONTENT="This booklet is available for free
on the NISO website (www.niso.org) and in hardcopy from NISO Press">

```

DC Record 4

```

<META NAME="DC.Title" LANG="en" CONTENT="Registry for the OpenURL Framework">
<META NAME="DC.Title" LANG="en" CONTENT="OpenURL Framwork Repository">
<META NAME="DC.Creator" LANG="en" CONTENT="NISO AX Committee for the
OpenURL">
<META NAME="DC.Description" LANG="en" CONTENT="The website provides
information on the components of OpenURL framework">
<META NAME="DC.Publisher" LANG="en" CONTENT="National Information Standards
Organization">
<META NAME="DC.Date" LANG="en" CONTENT="2003-02-20">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Publisher" LANG="en" CONTENT="OCLC, Inc."> LANG="en"
CONTENT="text/html">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=ListSets">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="ANSI/NISO Z39.88-2004">
<META NAME="DC.Relation" LANG="en" http://alcme.oclc.orgALCME: Advanced
Library Collection Management Environment">
<META NAME="DC.Subject" LANG="en" CONTENT="registry">
<META NAME="DC.Subject" LANG="en" CONTENT="openURL">
<META NAME="DC.Subject" LANG="en" CONTENT="opensource tools">
<META NAME="DC.Subject" LANG="en" CONTENT="repository interface">
<META NAME="DC.Subject" LANG="en" CONTENT="open-linking environment">
<META NAME="DC.Subject" LANG="en" CONTENT="network resources">

```

DC Record 5

```

<META NAME="DC.Title" LANG="en" CONTENT="Metadata.Net Home Page">
<META NAME="DC.Creator" LANG="en" CONTENT="Australian Government's
Cooperative Research Centre's Program. Distributed Systems Technology Centre.
Resource Discovery Unit.">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata initiatives">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata projects">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata tools and services">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata schema registry">
<META NAME="DC.Subject" LANG="en" CONTENT="Internet">
<META NAME="DC.Subject" LANG="en" CONTENT="information storage and
retrieval">
<META NAME="DC.Subject" LANG="en" CONTENT="information management">
<META NAME="DC.Description" LANG="en" CONTENT="Metadata.net aims are to
investigate and develop tools, technologies, and information management
processes that will allow organizations to locate, access, retrieve, and
manage information in highly distributive heterogeneous networks.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Distributed Systems Technology
Centre (DSTC)">
<META NAME="DC.Contributor" LANG="en" CONTENT="MAENAD (Multiple Access
Across Enterprises, Networks and Domains)">
<META NAME="DC.Date" LANG="en" CONTENT="2004-12-10">
<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.metadata.net">
<META NAME="DC.Language" LANG="en" CONTENT="en">

```

Discussion

As I looked at the five resources, Clifford Lynch's suggestion that a metadata record can be considered a collection of assertions about something became very clear to me. The basic components of a metadata record are easy to decipher from looking at a resource. It was not too difficult to discern what information went with what element, and to follow the syntax and semantics of Dublin Core (DC) and the Electronic Resource Citation (ERC) schemes, such as they are. The difficulty really came in with the subject or "aboutness" representation in the records.

The primary concern in metadata creation for the purpose of record discovery is to ensure that users can discover, identify, select, and access information objects. In order to accomplish this, one has to know what attributes are important to the users in a particular information community. This assignment helped me discover the limitations of Machine-Readable Cataloging (MARC) in this area. MARC is a remarkable accomplishment. As I read the articles prior to working on the assignment I was not convinced that new schemas had to be developed at all. I felt that MARC was being overlooked only because it was developed for one particular community, that of library users. What the assignment made me realize, though, is that MARC was not developed for library users at all. It was simply developed as a way for information to be exchanged between library systems. I knew this intellectually, but the implications became crystal clear as I worked on this assignment.

I began by looking at the information objects as a cataloger would approach the project. It soon became evident that this approach was counter-productive. The reason that Dublin Core was developed is because producing MARC record elements takes too long and requires a higher level of expertise than costs will allow. I created the Electronic Resource Record #4 for the website *Registry for the OpenURL Framework* in great detail. It took me two hours of work to read through the different levels of the website provenance. If information objects are to be made accessible to users, then the aim of finding out the least amount of information users need to find the source is the quest. The Dublin Core (DC) group tried to get down to the most useful fifteen elements. The Electronic Resource Citation (ERC) was developed to whittle down Dublin Core even more, down to four elements. The main point between the two schemes, though, is that the ERC has four *essential* elements and flexible semantics depending on the context of the story. I found the story feature of the ERC to be the most provocative concept.

The most difficult part of creating the records was to decide on the "aboutness" of the information object. I thought that using some kind of controlled vocabulary would be the best method. However, this was not an option for most objects. I was outside of the information community for most of the sources in my group. The best course of action was to use terms that were in the source itself. I could be pretty sure that the users who would benefit from discovering this resource, would be using some of the same terms that the creator of the resource used. This took me to an increased understanding of the definition of metadata used for this course. I really had to look at a resource from all angles to determine which terms to use. I used the design of the resource display to determine a title. I used the underlying source code to try to find information such as creators or dates. The definition or description of metadata captured the actions I had to take to describe or define a particular information object for the specific purpose

of making it discoverable and discernable to someone interested in the information contained in the object.

From this experience, I think that there is a place for metadata terms to be supplied by the creator of an information object. This does not mean that additional processing of the resource would not be needed. Although I do consider that unstructured metadata such as the full-text information retrieval (IR) systems comprise metadata, I do not think it would be the most beneficial to users. I think there is value in human-generated structured metadata for information retrieval because humans can discern relationships between information objects that computers can not. However, there is a cost to human-generated structured metadata and so I think that it is best to combine human-generated and computer-generated metadata creation. Since the creator of information objects can be expected to be in the same information community as the users, I think it would be beneficial to have metadata subject terms created at that level.

There is still the issue of metadata quality to be resolved. As good as MARC is, there are still some libraries in the OCLC information community whose records are not used by “those in the know.” I believe in this would be true in any information community. What is needed then is to work on having registries of controlled vocabulary for particular information groups. I find myself using ODLIS: Online Dictionary of Library and Information Science when I need to explain a library activity to someone outside of the library information community. I am sure that others do the same within their own communities. However, I did not find any of these “dictionaries” as I tried to work on the record creation. The Library of Congress Subject Headings do not use terms in the vernacular of information communities so I did not rely on them much in the creation of metadata records.

Summary and Conclusion

I found using the Generic Metadata Editor to be a useful tool. I think it would be more useful if I added selections to it as I used it. For example, I had difficulty in figuring out what label to use for a **pdf** document. One of my classmates pointed out to me that this is considered an application file. The Generic Metadata Editor tool would have been more useful if that term had been an option in a drop-down menu. Both the Dublin Core and the Electronic Resource Citation appear to be simple schemes to follow. In fact, they are not difficult to use but they require knowledge on the part of the person creating the records. Some of the knowledge of the intrinsic information, such as intellectual content, can be supplied by the authors of networked resources. I depended heavily on the user guide, Using Dublin Core, for creating the DC records. I understand why syntax independence is considered beneficial but then it is necessary for each information community to define its own syntax and has implications for interoperability.

In conclusion, both DC and ERC records were able to be created in a short amount of time and allow for machine processing. Through the assignment, an understanding of the pros and cons of each schema was developed. I found that the Dublin Core schema was easier to use. The four core elements of the ERC were adequate for description purposes. However, the use of qualifiers would have made the schema difficult to use and more costly. In particular, I found that the use of the syntax, in particular for dates, is necessary and the use of controlled

vocabularies may be preferred, as pointed out in the article *Metadata Principles and Practicalities* (Principles, C. Refinement). It would have been interesting to have an information object not in English and attempt to create a metadata record. The assignment did allow me to gain valuable experience in comparing and using two metadata schemes.