

## Metadata Record Creation Report for SLIS 5223 – Spring 2005.

The creation of metadata is an important aspect to information management as it is the primary means of resource discovery. In other words good metadata allows users to quickly find the information that they need to answer their questions. As the bridge between question and answer its importance cannot be overestimated. Understanding metadata makes using metadata easier. Hence it is a useful exercise to learn how metadata is created, even for those who do not intend to be actively engaged in metadata creation in a professional sense.

The purpose of this document is to record and report the experiences of the author in the creation of ten metadata records. These records are meant to describe the content of five distinct items. Each item has 2 different records created for it. First a record was made according to the Electronic Resource Citation schema and then a separate record was created according to the rules of the Dublin Core schema. All ten records are included in this document in their entirety.

### **Electronic Resource Citation Records**

#### Record 1.

who: National Standards Information Organization  
what: Understanding Metadata  
when: 2004  
where: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

#### Record 2

who: Multimedia Access across Enterprises, Networks And Domains | Distributed Systems Technology Centre  
what: Metadata.net | Metadata Initiatives | Metadata Projects | Metadata tools and services | Metadata schema | Metadata schema registries | Metadata Journals  
when: 2004  
where: <http://metadata.net/>

#### Record 3

who: Dublin Core Metadata Initiative  
what: DCMI Frequently Asked Questions (FAQ)  
when: 2003  
where: <http://dublincore.org/resources/faq/index.shtml>

## Record 4

who: National Information Standards Organization | Van de Velde, Eric F.  
 what: Registry for the OpenURL Framework - ANSI/NISO Z39.88-2004  
 when: 2004  
 where: <http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=ListSets>

## Record 5

who: Moen, William E.  
 what: Resource Discovery Using Z39.50: Promise and Reality.  
 when: 2001-1-23  
 where: [http://lcweb.loc.gov/catdir/bibcontrol/moen\\_paper.html](http://lcweb.loc.gov/catdir/bibcontrol/moen_paper.html)

**Dublin Core Records**

## Record 1

```
<META NAME="DC.Title" LANG="en" CONTENT="Understanding Metadata">
<META NAME="DC.Creator" LANG="en" CONTENT="National Standards Information
Organization">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="Resource Discovery">
<META NAME="DC.Subject" LANG="en" CONTENT="Archiving and Preservation">
<META NAME="DC.Description" LANG="en" CONTENT="This article explains what
Metadata is and what it does. Also offers information on several metadata schemes and their
respective elements. Also discusses interoperability issues and future directions of Metadata ">
<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="text/html">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.niso.org/standards/resources/UnderstandingMetadata.pdf">
<META NAME="DC.Source" LANG="en" CONTENT="Metadata Made Simpler: A guide for
libraries">
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

## Record 2

```
<META NAME="DC.Title" LANG="en" CONTENT="Metadata.Net">
<META NAME="DC.Creator" LANG="en" CONTENT="Distributed Systems Technology
Centre (DSTC)">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Initiatives">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Schema">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Schema Registries">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Journals">
```

```

<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Tools">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata Projects">
<META NAME="DC.Description" LANG="en" CONTENT="Metadata.net is a website
maintained by the MAENAD Project of DSTC. It has information regarding Metadata initiatives,
projects, tools, schema and maintains a list of journals regarding metadata">
<META NAME="DC.Publisher" LANG="en" CONTENT="Multimedia Access across
Enterprises, Networks And Domains, DSTC">
<META NAME="DC.Contributor" LANG="en" CONTENT="Multimedia Access across
Enterprises, Networks And Domains">
<META NAME="DC.Date" LANG="en" CONTENT="2004">
<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://metadata.net/">
<META NAME="DC.Language" LANG="en" CONTENT="en">

```

## Record 3

```

<META NAME="DC.Title" LANG="en" CONTENT="DCMI Frequently Asked Questions
(FAQ)">
<META NAME="DC.Creator" LANG="en" CONTENT="Dublin Core Metadata Initiative.">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="Resource Discovery">
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core Metadata Initiative
DCMI">
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core Metadata Element Set">
<META NAME="DC.Subject" LANG="en" CONTENT="Warwick Framework">
<META NAME="DC.Subject" LANG="en" CONTENT="Open Metadata Registry">
<META NAME="DC.Subject" LANG="en" CONTENT="DCMI Intellectual Property">
<META NAME="DC.Subject" LANG="en" CONTENT="Qualified Dublin Core">
<META NAME="DC.Subject" LANG="en" CONTENT="Simple Dublin Core">
<META NAME="DC.Description" LANG="en" CONTENT="This web page answer general
and commonly asked questions regarding the Dublin Core Metadata Initiative. ">
<META NAME="DC.Publisher" LANG="en" CONTENT="DCMI">
<META NAME="DC.Date" LANG="en" CONTENT="2003-10-31">
<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://dublincore.org/resources/faq/index.shtml">
<META NAME="DC.Source" LANG="en" CONTENT="DCMI Website.
(http://dublincore.org/index.shtml)">
<META NAME="DC.Language" LANG="en" CONTENT="en">

```

## Record 4

```

<META NAME="DC.Title" LANG="en" CONTENT="Registry for the OpenURL Framework -
ANSI/NISO Z39.88-2004">

```

```

<META NAME="DC.Title" LANG="en" CONTENT="OpenURL Framework Repository">
<META NAME="DC.Creator" LANG="en" CONTENT="NISO AX Committee">
<META NAME="DC.Subject" LANG="en" CONTENT="OpenURL Framework">
<META NAME="DC.Subject" LANG="en" CONTENT="Namespaces">
<META NAME="DC.Subject" LANG="en" CONTENT="Character Encoding">
<META NAME="DC.Description" LANG="en" CONTENT="The OpenURL Framework for
Context-Sensitive Services Standard provides a method of describing a referenced resource
along with a description of the context of the reference. This site contains the specific standards
used to employ OpenURL.">
<META NAME="DC.Publisher" LANG="en" CONTENT="NISO">
<META NAME="DC.Contributor" LANG="en" CONTENT="Van de Velde, Eric F.">
<META NAME="DC.Date" LANG="en" CONTENT="2004">
<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://alcme.oclc.org/openurl/servlet/OAIHandler?verb=ListSets">
<META NAME="DC.Language" LANG="en" CONTENT="en">

```

#### Record 5

```

<META NAME="DC.Title" LANG="en" CONTENT="Resource Discovery Using Z39.50:
Promise and Reality">
<META NAME="DC.Title" LANG="en" CONTENT="Conference on Bibliographic Control in
the New Millennium (Library of Congress)">
<META NAME="DC.Creator" LANG="en" CONTENT="Moen, William E.">
<META NAME="DC.Subject" LANG="en" CONTENT="Z39.50">
<META NAME="DC.Subject" LANG="en" CONTENT="Resource Discovery">
<META NAME="DC.Subject" LANG="en" CONTENT="Information Retrieval">
<META NAME="DC.Subject" LANG="en" CONTENT="Interoperability">
<META NAME="DC.Subject" LANG="en" CONTENT="Cross Domain Searching">
<META NAME="DC.Description" LANG="en" CONTENT="This paper reviews the history of
Z39.50 and proposes ways in which it is suited to be used in the current networked
environment.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Library of Congress">
<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.Language" LANG="en" CONTENT="en">

```

#### **Discussion of, and Reflection on, the Record Creation Process:**

Having never been involved in the creation of metadata prior to this exercise I had only a vague idea of what to expect. The first step for me was setting up the Generic Metadata Editor in NoteTab. I found the lists of elements in the two schemas to be disorientating when viewed in the editor. At the time I was overwhelmed and suddenly felt that perhaps I should have read the metadata scheme documentation first. The ERC elements appeared rather straight forward (who, what – no problem), however the Dublin Core elements were somewhat more enigmatic (Relation? To what?). In retrospect the review of the elements without knowing what they meant gave me a framework to plug in the answers as I found them in the documentation, so it worked out for the best.

The next challenge was to learn the basics of the schemas. The Electronic Resource Citation schema was the one I choose to tackle first. The article, “A Metadata Kernel for Electronic Permanence,” by John Kunze did an excellent job of explaining how ERC works. In the confines of this exercise, it seems that Dublin Core gives more information per record; however, I feel that ERC has greater utility. ERC’s anchoring story provides enough information for quality resource discover by answering four questions (who, what, when, and where). In four lines of text you can make a decision on the content of a document. Its greatest advantage is its easy human readability. Someone with no prior knowledge could read an ERC anchoring story and understand it.

In creating the records under the ERC schema I had little difficulty in deciding what information needed to be entered in each field. Each element seemed fairly self-explanatory. Finding the correct information was sometimes challenging. On records 2, 3 and 4, it was difficult to decide what to put in the “Who” field. In the end I went with the organization that publish the document. It seems likely that there was an individual author, or list of authors, but I could not find such information included in the documents. On record 4, I was able to find the name of chair of the committee that compiled the document and so I included his name as a second entry in the “who” field.

It was often very difficult to pin down the date of the “When” field. On those items that had clear publishing dates it was natural to go with that date. However, some of the items were Websites. Websites are constantly revised and edited and often do not include a date of original creation. This made it unclear as to how to date the item. According to Kunze, the “when” has to be a date significant to the items “lifecycle”<sup>1</sup>. Therefore, I used the “last updated” date on the page in question when available.

Dublin Core posed a very different set of challenges. DC is a flexible schema, but after working with ERC it appears more structured. Learning the element set took a significant amount of time and I am certain that my grasp of the desired content for some elements is still weak. Several fields seem unnecessary in my mind; particularly, the Type and Format fields seem redundant. It seems a very fine distinction between the two, one that is not necessarily useful. I realize that there are times when they are different but I cannot think of time when they could not be combined without any loss of record usability.

---

<sup>1</sup> Kunze, John. A Metadata Kernel for Electronic Permanence.

The wide selection of fields to input data into made the creation of records for Dublin Core more time consuming as well. However the completed records are richer and more full than a simple ERC anchoring story. Dealing with the Title, Creator, Date and Subject fields was simple, as I had reviewed the items for this information previously to do the ERC records. Filling in the remaining fields was the interesting part. The most challenging fields were the Description, Source, and Coverage.

The description field is doubtlessly the most important field for resource evaluation by human end users. Therefore I felt it was important to provide accurate description but also to keep them short and simple. Balancing detail and brevity is always challenging. I avoided the easy route of lifting the thesis statement out of the documents. I felt that was inadequate to the task of describing the document as a whole. With websites there was no clear “thesis” statement to be had. The OpenURL website (Record 4) proved the most difficult for me to describe. This was primarily due to not having any knowledge of this topic prior visiting the website during the record creation process. This experience crystallized the problems of a non-expert describing an items so that both experts and non-experts can find the information they are seeking. The only solution I could find was to invest a significant amount of time reading the literature on the site to develop vocabulary necessary to explain what was on the site. However, if I had hundreds or thousands of such sites to make records for this would not be a viable option.

The Source field was challenging because my initial reaction was that everything has a source. However after reviewing the literature on the Dublin Core elements I revised my evaluation of the items and decided that only record 1 and 3 had information that would be placed in the Source field. Record 1 states that it is a “revision and expansion” of a previous document. So I placed the title information of the previous document in this field. Record 3, is a FAQ on a webpage explaining the Dublin Core. The source of the information in the FAQ is the information contained on the primary content pages of the webpage. The information is simply reformatted in a questions and answers format. I used the Source field to point back to the original, more fleshed out information as a means of providing end users with additional information.

Most of the documents from which I was creating records did not appear to require the Coverage field. The one exception was Record 5, which reviewed the evolution of the Z39.50 in a historical fashion, so the dates of 1979-2001 were entered in the Coverage field. Further reflection led me to decide that including this field was not helpful information to end-users. The document primarily dealt with the current standing of Z39.50 in the year 2001. Thus the usefulness of listing it as a resource for someone interested in how Z39.50 was being used in 1984, for example, would be minimal. Having the coverage listed from 1979-2001 gave the impression of a more through coverage across that time frame than the document actually provided.

### **Summary and Conclusion:**

The creation of metadata is a fascinating, time-consuming process that I am looking forward to learning more about. The most challenging aspect of creating metadata is learning the schema to be used. Once one is familiar with what types of information to plug into the various element fields the actual creation process is straightforward. There were sometimes, in my experience, where finding the information was difficult, but not as a rule. The ERC schema, having a more basic kernel, has a significant advantage over Dublin Core when being used by non-experts. This I feel will allow it to gain dominance if it is standardized and promoted.

No lesson stands out more clearly to me than that the process of creating quality metadata requires a serious investment of time and mental energy. That said I do believe that it is highly possible, given wide acceptance of a standard form such as ERC or DC, for individual information generators to create their own metadata as part of their normal routine. Creating metadata at the time of original information creation seems the most efficient and promising option. However adopting such a course only solves future problems, and leaves millions of untagged records that have already been created. It will be interesting to see how the Information management community decides what records are worth tagging and finds the resources and methods to implement their plan.