

## Report on Metadata Records

### Introduction

This document contains ten metadata records, five using the Electronic Resources Citation schema and five using the Dublin Core schema. These records are reproduced in this report for reference purposes. After the records is a Discussion concerning their creation. The section relates decisions and problems I faced, particularly related to using the text editor, finding data, what data to include, and input rules. After the Discussion appears the Summary and Conclusion, in which I review the material and relate what I learned from this assignment.

### Electronic Resource Citation Records

erc:

who: Multimedia Access across Enterprises, Networks And Domains | Distributed  
Systems Technology Centre  
what: Metadata.net | Metadata.Net Home Page  
when: 2004  
where: <http://metadata.net/index.html>

erc:

who: National Standards Information Organization  
what: Registry for the OpenURL Framework - ANSI/NISO Z39.88-2004 | OpenURL  
Framework Repository  
when: 2003  
where: <http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=ListSets>

erc:

who: Moen, William E.  
what: Resource Discovery Using Z39.50: Promise and Reality | Conference on  
Bibliographic Control in the New Millennium (Library of Congress)  
when: 2001 01 23  
where: [http://www.loc.gov/catdir/bibcontrol/moen\\_paper.html](http://www.loc.gov/catdir/bibcontrol/moen_paper.html)

erc:

who: National Information Standards Organization | NISO Press  
 what: Understanding Metadata  
 when: 2004  
 where: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>  
 | ISBN 1-880124-62-9

# ---- new segment ----

erc-about:

who: (:none)  
 what: metadata | metadata schemes  
 when: (:null)  
 where: (:null)

## Dublin Core Records

<META NAME="DC.Title" LANG="en" CONTENT="Metadata.net">  
 <META NAME="DC.Title" LANG="en" CONTENT="Metadata.Net Home Page">  
 <META NAME="DC.Contributor" LANG="en" CONTENT="MAENAD: Multimedia Access across  
 Enterprises, Networks And Domains">  
 <META NAME="DC.Contributor" LANG="en" CONTENT="DSTC: Distributed Systems Technology  
 Centre">  
 <META NAME="DC.Subject" LANG="en" CONTENT="metadata">  
 <META NAME="DC.Subject" LANG="en" CONTENT="metadata initiatives">  
 <META NAME="DC.Subject" LANG="en" CONTENT="metadata schema">  
 <META NAME="DC.Description" LANG="en" CONTENT="A website containing information on  
 metadata schema, projects, initiatives, tools, services, journals, and other resources.">  
 <META NAME="DC.Date" LANG="en" CONTENT="Modified 2004-12-10">  
 <META NAME="DC.Type" LANG="en" CONTENT="text">  
 <META NAME="DC.Type" LANG="en" CONTENT="website">  
 <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">

<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="The NISO AX Committee for the  
 OpenURL">  
 <META NAME="DC.Subject" LANG="en" CONTENT="Z39.88">  
 <META NAME="DC.Date" LANG="en" CONTENT="Created 2003-02-20">  
 <META NAME="DC.Date" LANG="en" CONTENT="Modified 2004">  
 <META NAME="DC.Publisher" LANG="en" CONTENT="National Information Standards  
 Organization">  
 <META NAME="DC.Type" LANG="en" CONTENT="text">

```

<META NAME="DC.Type" LANG="en" CONTENT="website">
<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">

<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="information retrieval">
<META NAME="DC.Subject" LANG="en" CONTENT="resource discovery">
<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 retrieval
  requirements in the networked environment.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Library of Congress">
<META NAME="DC.Date" LANG="en" CONTENT="dateCreated 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">
<META NAME="DC.Relation" LANG="en" CONTENT="References William E. Moen's 2000 article
  Interoperability for information access: Technical standards and policy considerations">

<META NAME="DC.Title" LANG="en" CONTENT="Understanding Metadata">
<META NAME="DC.Creator" LANG="en" CONTENT="National Information Standards
  Organization">
<META NAME="DC.Publisher" LANG="en" CONTENT="National Information Standards
  Organization. NISO Press">
<META NAME="DC.Contributor" LANG="en" CONTENT="Rebecca Guenther">
<META NAME="DC.Contributor" LANG="en" CONTENT="Jacqueline Radebaugh">
<META NAME="DC.Contributor" LANG="en" CONTENT="staff members of the Library of
  Congress Network Development and MARC Standards Office">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata schemes">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata">
<META NAME="DC.Description" LANG="en" CONTENT="tableOfContents="What is Metadata;
  What Does Metadata Do; Structuring Metadata; Metadata Schemes and Element Sets;
  Creating Metadata; Interoperability and Exchange of Metadata; Future Directions; More
  Information on Metadata; Glossary"">
<META NAME="DC.Description" LANG="en" CONTENT="This article describes what metadata is,
  does, describes a variety of metadata schemes, and includes other helpful information that

```

introduces the topic. References are provided for further reading, and a glossary is included.">

```
<META NAME="DC.Relation" LANG="en" CONTENT="IsBasedOn Metadata Made Simpler: A
guide for libraries by NISO Press 2001">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<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/UnderstandingMetadata.pdf">
<META NAME="DC.Identifier" LANG="en" CONTENT="ISBN 1-880124-62-9">
<META NAME="DC.Date" LANG="en" CONTENT="2004">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright 2004 National Standards
Organization">
```

## Discussion

NoteTab Light was a nice editor to work with. There was no fancy color-coding or similar features, but I appreciated using the schema library files that enabled drag and drop functionality. I think that this function affected my attitudes toward each schema. Because all of the elements and some qualifiers were available for Dublin Core, I was able to appreciate how the elements could best be used without the frustration of hand-coding it. I think that if I had had to hand-code each DC record, my appreciation of it would have been hampered—as would the ability of any novice user who did not have the use of such a text editor & library. However, because ERC is already so simple, the drag-and-drop function added very little to its ease. I would have liked to have had some additional story tags available (“erc: about,” etc.), because a significant amount of my frustration with ERC was the desire to include additional data but little guidance with how to present it. I suspect that there are dramatic differences between the attitudes of metadata creators who use editors similar to NoteTab with the schema libraries, and those who do not. I did not read much about the part such an editor plays in metadata, but I think that there is a lot to be researched—and possibly changed.

One of the general issues I faced with all the records—the Dublin Core records in particular—was punctuation. As mentioned in the lectures, this could affect the machine-readability of the records. This was on my mind even before reading the lecture on rules, as I am currently learning how to script PHP in another class, and am learning that when you print strings to the screen, sometimes to print punctuation without causing script errors you have to “escape” certain characters or punctuation. In the end, I ignored this issue, because I do not know enough about the various machine processes that would read these records. I used punctuation as it was shown in the documents, for the most part. The exception was that any titles that were presented in sentence format I re-formatted to be completely capitalized.

In general, I disliked the extreme simplicity of the ERC format. It was nice that anchoring story was easy to define, but I felt uncertain about going further. I was not confident in defining other stories, and what the elements would signify in the new story. I did try this method with the

"Understanding Metadata" document; I created "erc: about." I used the What element to describe the subject of the document. I placed the Who, When, and Where elements in the story, even though I did not need to include them because I had no data to place in them, because I wanted to experiment with the codes for no value, and had not felt that these were appropriate in any of the other records. I placed ":none" in the Who element to convey that there was no information for this element: the document was not about a person. I used the ":null" code in the When and Where elements to indicate that there was not a value in the element without connoting that this information did not exist. While I did not consider these elements to be relevant, another person might consider that this document's When was the twenty-first century and that its What was the United States. Therefore, I felt the ":null" code was more appropriate than the ":none" code.

After some thoughtful consideration of how to create ERC records, I can see that this system offers a lot of flexibility and ease of use, particularly for non-trained professionals. The simplicity of the anchoring story enables untrained people to create good ERC records. Also, the Who/What/When/Where format is an instinctive set of elements with an instinctive order—at least for Americans. This makes it easy to remember and use. My favorite part of ERC, however, was that it provides the discipline that DC lacks—a specific order, and the four key elements that are required. However, I prefer the pre-defined format of DC, which gives me a pre-labeled place for all my information.

I really enjoyed working with the Dublin Core schema. I appreciated having clear-cut, distinct elements for the shades of variance between Creator, Publisher, and Contributor. The fifteen elements could provide a variety of well-documented information, and the qualifiers really add to the level of data distinction. However, I was frustrated that DC does not require any particular elements. This seems like a chaotic system—I could create records using only Creator, Date, and Subject elements, and my colleague could create records using only Title and Publisher—how can any system reliably present and interpret documents without common elements? But at the same time, this flexibility has its uses. For instance, in my recent thesis for my Master's in Art History, I referred to 43 objects. Because these were medieval objects, some had no title and some had no artist. In fact, some had neither, and finding a system with which to categorize and refer to these objects was problematic. The DC schema would enable me to define these objects by Date, Description, and Identifier—in this case, a museum catalog number.

I also do not like how the DC elements are organized by default; it makes little sense to me why Creator, Contributor, and Publisher are not closer together, instead of being separated by Subject and Description. However, since the elements do not have to appear in a particular order, I have rearranged some of the elements in the "Understanding Metadata" record as an experiment (I wanted to try some different input techniques in order to assess the schemes, so I did not use the same rules for each record). While I find this order easier to read and understand, I would think that regularity in the order of elements is a good standard. Otherwise, when I read Jane Doe's DC records and she begins with Date, Language, and Identifier, I may become frustrated or confused by her method. I suppose that a computer could read the elements in any order and automatically re-arrange them, so the machine-readability would be affected little.

The unchangeable order of elements in ERC is one of its qualities that I appreciate. The Who/What/When/Where order is intuitive for English-speakers and thus is something easily

grasped by the untrained. The simplicity afforded by only four elements and the non-changing order ensures that all ERC records can easily be read and understood, no matter who created them. This is a definite improvement over DC. (I can see the complexity inherent in choosing an appropriate metadata scheme, as I obviously cannot make up my mind about the “better” scheme between these two!)

I had problems determining which data to use in the Title/Who elements in both schemes with two of the HTML documents (“OpenURL Registry” and “Resource Discovery”). I found it difficult to define titles for web pages, as the HTML names one title, and the top of the page presents a different title. In both schemes, I used two titles for these documents. In ERC, I separated the titles with a pipe and placed them in one line; in DC, I entered each into a separate Title element. I don’t think that it makes a difference for machine-readability between separating with a pipe and beginning a new element, but I find starting a new element a much easier format to read. For example, in the “Resource Discovery” document, the HTML title was very long, and when added to the page title already in the metadata record, I found the ERC pipe-separated value frustrating to read.

In “Understanding Metadata” I was confused by what to enter in the DC record for the Creator and Publisher fields. I understand Hillman’s rule of thumb that when in doubt, individuals are classified as Creator and organizations as Publisher, but my problem with this document was that two related organizations were listed, but no individuals. The concept was further confused for this document because NISO Press appears to be a smaller section of NISO itself—however, the term “Press” as well as its location in the print information section of the page seems to indicate that this sub-section is the Publisher. I debated back and forth about including NISO as both Creator and Publisher, and naming NISO Press as a second Publisher. Ultimately, I used Hillman’s method of presenting hierarchies of organizations by presenting the parent organization first, followed by a period, a space, and the smaller organization/sub-section. I initially felt that to present the Publisher as “National Information Standards Organization. NISO Press” lacked the precision and brevity that “NISO Press” possessed. However, I later decided that this format was appropriate based on Hillman’s guidelines. I now appreciate that Hillman’s method provides a level of explanation—the second organization name is clearly a subset of the previous one. That does provide an additional nugget of data itself, the relationship between two organizations.

With the “Metadata.Net” document, I had problems finding the data for the Creator field. Although two related organizations were listed on the website, their relationship to the document was not explicitly expressed. Both MAENAD and DSTC appeared prominently on the page, but their relationship was not spelled out. I clicked on the links for each, and those web pages indicated that MAENAD was sub-section of DSTC. Because of that, in the DC record I initially placed MAENAD as the Creator and DSTC as the Publisher. However, after reading through the lectures again and contemplating the Contributor element, I decided that since I could not accurately determine which party was Creator or Publisher, that it was best to place both names in the Contributor field. Hillman states that this element should be used when responsibility is unknown. I did not use Hillman’s method for representing organization hierarchies, however, because I was still unsure as to the precise nature of the relationship between MAENAD and DSTC.

In the ERC version of the "Metadata.Net" record, I toyed with the idea of leaving out the Who element and putting in the code for either unknown or unassigned. However, although I am not sure of the role that MAENAD and DSTC played in this website (creator, publisher, contributor), they both appear responsible, which means that they both fit the description of the Who element. I therefore input both and separated them with the pipe.

I like the Description element in DC, because it provides a nice abstract, as in the "Understanding Metadata" record. However, I experienced difficulty with the "OpenURL Registry" record because I did not understand the document enough to write a description of it. This also made finding subject keywords difficult in both schemes. An aspect of the description element that I enjoyed was the tableOfContents qualifier. For the "Understanding Metadata" document, I enjoyed being able to enter the name of each section. I felt that this could be a great help to someone filtering through records that had been returned by a search. In past searches, when I have been able to read a list of the table of contents for a book or article, I have often been able to quickly and accurately determine whether the material contains the information that I need. This is a powerful search and decision tool for students and researchers.

My favorite example of using DC qualifiers was in the "Resource Discovery" document. I wanted to show that an article was referred to in this document, and easily chose References from the list of qualifiers for Relation. The notion of relation is something that I have personally struggled with defining in my own classification, and I dislike systems that leave it out, because I think that it is an important concept. I am impressed by the number of qualifiers available for this element that precisely identify the nature of the relation between the objects. It comes in particularly handy for HTML documents that have been transcribed from a print article.

The "Understanding Metadata" document also exhibited another example of the treatment of more than one value. For the Identifier in the DC record, I used both the ISBN and the URL as separate lines. Again, like the documents for which I used two titles, this is much easier to read than in the ERC record where I included both values on one line, separated by a pipe. Although ERC is cleaner to read because it does not include the code as DC does, human reading is still hampered by the inclusion of multiple values in a single element.

## Summary and Conclusion

Creating the records taught me a lot about how to interpret the rules and to think about how the user will be using the information. Since these records were created as an exercise, I did not have a particular audience to keep in mind, but some information that could have been included seemed irrelevant for users (for instance, when I left out the When and Where elements in the "erc: about" section of the "Understanding Metadata" record). This assignment also showed me that a familiarity with LC Subject Headings or some similar controlled vocabulary will be necessary to create these records and represent their subjects accurately. I felt at a loss for describing the subjects, for the most part, because I do not have a background in any controlled vocabulary of the sort.

I am torn between flexibility/creativity and structure in each schema. I like the structure of the DC elements and the fact that reading the element labels prompts me to seek out specific information (for instance, I would not have thought to include a separate field for Relation). However, ERC's flexibility of creating stories enables me to record information for which DC may not have accounted. ERC is very structured in the order of its elements, which makes it very easy to both create and read ERC records. DC's order leaves something to be desired for organization's sake, although I like the freedom to place elements where I feel they belong. There is a clear give and take between these two schemes, and although I personally like DC better—probably because I am more familiar with its format—I cannot say which scheme is best overall. Their striking differences suggest that they are best used in different circumstances. For my own personal uses—cataloging my books, DVDs, and HTML files—I would choose DC, because it is a schema which will provide space for a lot of data and is designed for such resource description, but it is also easy in that I do not have to create a new story to record additional data.

In creating the ERC records, I gained a new respect for this schema. Although I disliked it initially, the more I created records with it, the more I appreciated how it uniquely treats information, and in particular how helpful its rules or order and required elements can be. I think that although I would prefer to use DC, that I may find that ERC comes in handy more often; particularly for any uses that involve human reading of the records.