David Denny
SLIS 5223
Spring 2005

Creating Metadata for Digitized Resources

Introduction

This paper explores many of the challenges and results underlying the metadata creation process in the context of two metadata schemes: The Electronic Resource Citation (ERC) and the Dublin Core Element Set (DC). The importance of providing consistent organization and access to sets of resources as large and diverse as those appearing in electronic format is discussed. Four core issues, determined to merit special attention in the digital environment, are identified and elaborated on.

Electronic Resource Citation Records

Understanding Metadata
who: National Information Standards Organization (NISO) Press
what: Understanding Metadata
when: 2004
where: http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

Metadata.net
who: Distributed Systems Technology Centre (DSTC)
what: Metadata.Net
when/modified: 2004 12 10
where: http://metadata.net/

Dublin Core Metadata Initiative (DCMI) Frequently Asked Questions (FAQ)
who: Dublin Core Metadata Initiative
what: Dublin Core Metadata Initiative Frequently Asked Questions (FAQ)
when: 2005
where: http://dublincore.org/resources/faq/index.shtml

Registry for the OpenURL Framework
who: National Information Standards Organization (NISO)
what: Registry for the OpenURL Framework
when: 2003 02 20
where: http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=Identify

Resource Discovery Using Z39.50: Promise and Reality
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

Dublin Core Records

Understanding Metadata

<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="metadata, schemes, element sets, interoperability and exchange">
<META NAME="DC.Description" LANG="en" CONTENT="A guide to understanding metadata and specific schemes associated with it">
<META NAME="DC.Contributor" LANG="en" CONTENT="Guenther, Rebecca ">
<META NAME="DC.Contributor" LANG="en" CONTENT="Radebaugh, Jacqueline">
<META NAME="DC.Date" SCHEME="ISO8601" 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" SCHEME="URI" LANG="en" CONTENT="http://www.niso.org/standards/resources/UnderstandingMetadata.pdf">
<META NAME="DC.Language" SCHEME="RFC1766" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsBasedOn Metadata Made Simpler: A guide for libraries">
<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">

Metadata.net

<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, projects, tools, schema, registry, journals">
<META NAME="DC.Description" LANG="en" CONTENT="provides sources of information about metadata">
<META NAME="DC.Date.Modified" SCHEME="ISO8601" 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" SCHEME="URI" LANG="en" CONTENT="http://metadata.net/">
<META NAME="DC.Language" SCHEME="RFC1766" LANG="en" CONTENT="en"

Dublin Core Metadata Initiative (DCMI) Frequently Asked Questions (FAQ)

<META NAME="DC.Title" LANG="en" CONTENT="Dublin Core Metadata Initiative (DCMI) Frequently Asked Questions (FAQ)">
<META NAME="DC.Creator" LANG="en" CONTENT="Dublin Core Metadata Initiative (DCMI)">

<META NAME="DC.Subject" LANG="en" CONTENT="metadata, frequently asked questions, faq, dublin core, dcmi">
<META NAME="DC.Description" LANG="en" CONTENT="questions and answers concerning metadata and the dublin core metadata initiative">
<META NAME="DC.Date" SCHEME="ISO8601" LANG="en" CONTENT="2005"><META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" SCHEME="URI" LANG="en" CONTENT="http://dublincore.org/resources/faq/index.shtml">
<META NAME="DC.Language" SCHEME="RFC1766" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsPartOf Dublin Core Metadata Initiative Website">
<META NAME="DC.Rights" SCHEME="URI" LANG="en" CONTENT="http://dublincore.org/about/copyright/index.shtml#copyright">

Registry for the OpenURL Framework
<META NAME="DC.Title" LANG="en" CONTENT="Registry for the OpenURL Framework">
<META NAME="DC.Creator" LANG="en" CONTENT="National Information Standards Organization (NISO)">
<META NAME="DC.Subject" LANG="en" CONTENT="URL framework, URL registry, Z39.88, OAI handler">
<META NAME="DC.Description" LANG="en" CONTENT="provides a means of describing a referenced resource and transporting these descriptions between networked systems">
<META NAME="DC.Date" SCHEME="ISO8601" LANG="en" CONTENT="2003-02-20">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Identifier" SCHEME="URI" LANG="en" CONTENT="http://alcme.oclc.org/openurl/servlet/OAIHandler">
<META NAME="DC.Language" SCHEME="RFC1766" LANG="en" CONTENT="en">

Resource Discovery Using Z39.50: Promise and Reality
<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="Z39.50, information retrieval, interoperability, multiple library catalogs">
<META NAME="DC.Description" LANG="en" CONTENT="Discusses the Z39.50 standard and its development, along with the complexity of information retrieval problems it addresses, and how the goals for its use has changed over time.">
<META NAME="DC.Date" SCHEME="ISO8601" 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" SCHEME="URI" LANG="en"
CONTENT="http://www.loc.gov/catdir/bibcontrol/moen_paper.html">
<META NAME="DC.Language" SCHEME="RFC1766"
LANG="en" CONTENT="en">

Discussion

Inevitably, there were numerous issues associated with the metadata creation exercise. The first major question instigated by this exercise was “which scheme is most effective - and how much does the function required of the metadata factor in to this effectiveness”? For example, while ERC was developed with “electronic permanence” in mind, its value as a descriptive metadata scheme is clearly evident because it focuses on the more basic (thus most likely to be searched) attributes of the *majority* of information objects.

Even though this exercise worked with two similar and relatively simple schemes (out of many available), it became evident early that digital resource description is not akin to classifying books. Perhaps the biggest problem was how to handle the diversity of the resources coming from a networked environment, especially while trying to maintain consistency and accuracy.

While there were many difficulties encountered during this exercise that are commonly associated with general resource description, this discussion focuses on four areas of metadata creation perceived to merit special consideration in an electronic resource environment:

- Granularity
- Ambiguity
- Complexity
- Consistency

Granularity
The difficulty in deciding at what level to create metadata - for a group of resources that make up a collection (i.e. pages in a website) - was encountered with the *Registry for the OpenURL Framework* resource. Distinctions between “whole” and “component” should be a consideration. Are individual metadata records required for each page in this small “theme specific” collection, or could we manage with a few collection-level records? Solving granularity problems appears to require a procedure for “inheriting” metadata from “parent” sources.

Ambiguity
Obviously, some elements are more difficult (and more subjective) to determine than others. Some, like subject, are most effective when extracting values from a controlled vocabulary. Title can be a surprisingly difficult area for certain digital sources. Determining the title of a published paper or thesis may be relatively straightforward, but

many resources have more than one possible (ambiguous) title. For example, what is the actual title of the following website? Is it *Registry for the OpenURL Framework* **or** *OpenURL Framework Repository*? Both are viable candidates and both are indicated in at least one area of the website. Other sources, particularly non-textual resources, may have no title at all. In these cases, it is important to determine who decides what the title of a given resource is and according to what criteria.

Complexity
Adding qualifiers to a schema, widely regarded as a necessary procedure for enriching the description by adding semantic specificity, was a problematic consideration and may turn a relatively simple task into a more complex one. ERC, which lacks an established core set of terms (semantics) for its "qualifier stories", makes it even more difficult to create machine-readable records of a consistent nature. Even with schemes that provide established qualifiers, any attempt to mix semantic simplicity with the complexity of qualifying elements compromises the goal of providing a simple, easy-to-use metadata scheme.

Consistency
The simplicity of the two schemes used in this exercise (especially ERC) helped to establish an important fundamental principle that it is better to be accurate and consistent in creating metadata for a few keys fields (e.g. title, author, date…) than to be so daunted by complexity that user satisfaction with retrieval results become compromised. In order to achieve metadata consistency, a scheme should be open and amenable to all communities, easy to use, and robust enough to handle an increasingly diverse set of resources.

Summary and Conclusion

The ultimate purpose of a metadata record is to make published resources easy to find and reliable to access. It should be possible for individuals as well as organizations to publish anything they want. The tremendous amount of information on the Internet is both a great boon and an (often) overwhelming burden. It's a burden in the sense of ever-increasing information management problems as resources exponentially grow. A simple-to-use, universally accepted scheme, like the Dublin Core, for creating (and managing) metadata offers a solution to the problems associated with the explosion of digital information.

However, in order to improve information retrieval results, networked resources have to be described by the same rules. Metadata schemes and semantics must be suitable for a wide variety of resources, and scaleable to enormous proportions. Procedures for creating the associated metadata should be automated in order to make it easy for creators (publishers) to describe, create, submit and administer their own resource descriptions. It must be made easy for creators because the sheer volume of information indicates the improbability of any one group or organization (such as librarians) acting as "custodians" of information resources.

While information professionals will most likely play a diminishing role in the creation of digital resource descriptions, they will continue to be at the forefront in the development of metadata standards and, as such, should be able to:

- Think universally. Solutions devised now and in the future need to be global and independent of discipline and language.

- Stay current with metadata trends. Many people or organizations are thinking about the same issues (i.e. Kunze and ERC), and different perspectives can be very enlightening.