

Creating Metadata Records

Introduction

The complexity of creating metadata is deceptive. On the surface, recording an object's title, creator, subject, and other information seems trivial. To select precisely the correct data that will promote the user tasks of finding, identification, selection, and acquisition, however, can be difficult, especially in the context of today's multiple (and often incompatible) metadata standards.

Indeed, creating metadata is not a mere issue of plugging square pegs into square holes and round pegs into round holes. Rather, some metadata standards have holes of only one shape, others have ones of many different shapes, and still others have ones that are ambiguously shaped. Thus, this document explores the difficulties of creating metadata, especially for electronic resources, and the ways in which different types of metadata standards can affect the metadata creation process and the resulting records' effectiveness. It compares records created about the same information objects using different standards—Electronic Resource Citation (ERC) and Dublin Core (DC). First, records of five electronic resources created using ERC are presented, followed by records of the same resources created using DC, and then a discussion of the pertinent metadata creation issues.

Electronic Resource Citation Records

Record 1

erc:

who: Brand, Amy | Daly, Frank | Meyers, Barbara

what: Metadata Demystified: A Guide for Publishers

when: 2003-07

where: http://www.niso.org/standards/resources/Metadata_Demystified.pdf

Record 2

erc:

who: Canadian Heritage International Network

what: Metadata Standards for Museum Cataloging

when: 2002-04-27

where: http://www.chin.gc.ca/English/Standards/metadata_description.html

Record 3

erc:

who: Hillmann, Dianne

what: Using Dublin Core

when: 2003-08-26

where: <http://dublincore.org/documents/2003/08/26/usageguide/>

Record 4

erc:

who: Van den Brink, Linda | Tauber, James

what: XMLSoftware

when: 2005

where: <http://www.xmlsoftware.com/>**Record 5**

erc:

who: Caplan, Priscilla

what: International Metadata Initiatives: Lessons in Bibliographic Control

when: 2001-01-23

where: http://www.loc.gov/catdir/bibcontrol/caplan_paper.html**Dublin Core Records****Record 1**

<META NAME="DC.Title" LANG="en" CONTENT="Metadata Demystified: A Guide for Publishers">

<META NAME="DC.Creator" LANG="en" CONTENT="Brand, Amy">

<META NAME="DC.Creator" LANG="en" CONTENT="Daly, Frank">

<META NAME="DC.Creator" LANG="en" CONTENT="Meyers, Barbara">

<META NAME="DC.Subject" LANG="en" CONTENT="Metadata">

<META NAME="DC.Subject" LANG="en" CONTENT="Publishing">

<META NAME="DC.Subject" LANG="en" CONTENT="Metadata standardization">

<META NAME="DC.Subject" LANG="en" CONTENT="Metadata syntaxes">

<META NAME="DC.Description" LANG="en" CONTENT="This guide presents an overview of evolving metadata conventions in publishing, as well as related initiatives designed to standardize how metadata is structured and disseminated online. Its focus is on strategic rather than technical considerations.">

<META NAME="DC.Publisher" LANG="en" CONTENT="NISO Press">

<META NAME="DC.Publisher" LANG="en" CONTENT="The Sheridan Press">

<META NAME="DC.Contributor" LANG="en" CONTENT="Harris, Pat">

<META NAME="DC.Contributor" LANG="en" CONTENT="Parente, Susam">

<META NAME="DC.Contributor" LANG="en" CONTENT="Pirkey, Kevin">

<META NAME="DC.Contributor" LANG="en" CONTENT="Suprock, Greg">

<META NAME="DC.Contributor" LANG="en" CONTENT="Witkowski, Mark">

<META NAME="DC.Date" LANG="en" CONTENT="2003-07">

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

<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">

<META NAME="DC.Format" LANG="en" CONTENT="19 pages">

<META NAME="DC.Format" LANG="en" CONTENT="428 kb">

<META NAME="DC.Identifier" LANG="en"

CONTENT="http://www.niso.org/standards/resources/Metadata_Demystified.pdf">

<META NAME="DC.Identifier" LANG="en" CONTENT="ISBN:1880124599">

<META NAME="DC.Language" LANG="en" CONTENT="en">

<META NAME="DC.Relation" LANG="en" CONTENT="Requires Adobe Acrobat Reader">
 <META NAME="DC.Rights" LANG="en" CONTENT="Copyright 2003, The Sheridan Press and NISO Press">

Record 2

<META NAME="DC.Title" LANG="en" CONTENT="Metadata Standards for Museum Cataloging">
 <META NAME="DC.Title" LANG="en" CONTENT="Collections Description Standards">
 <META NAME="DC.Creator" LANG="en" CONTENT="Canadian Heritage International Network">
 <META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
 <META NAME="DC.Subject" LANG="en" CONTENT="Cataloging">
 <META NAME="DC.Subject" LANG="en" CONTENT="Descriptive metadata">
 <META NAME="DC.Subject" LANG="en" CONTENT="Museums">
 <META NAME="DC.Subject" LANG="en" CONTENT="Collection metadata">
 <META NAME="DC.Description" LANG="en" CONTENT="Describes and contains links to many different standards for creating museum-related metadata. Standards are related to the type of museum or collection.">
 <META NAME="DC.Date" LANG="en" CONTENT="2002-04-27">
 <META NAME="DC.Type" LANG="en" CONTENT="Text">
 <META NAME="DC.Format" LANG="en" CONTENT="text/html">
 <META NAME="DC.Format" LANG="en" CONTENT="36 kb">
 <META NAME="DC.Identifier" LANG="en" CONTENT="http://www.chin.gc.ca/English/Standards/metadata_description.html">
 <META NAME="DC.Language" LANG="en" CONTENT="en">
 <META NAME="DC.Rights" LANG="en" CONTENT="© CHIN 2004. All rights reserved">
 <META NAME="DC.Rights" LANG="en" CONTENT="http://www.chin.gc.ca/English/Notices/index.html">

Record 3

<META NAME="DC.Title" LANG="en" CONTENT="Using Dublin Core">
 <META NAME="DC.Creator" LANG="en" CONTENT="Hillmann, Dianne">
 <META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
 <META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core">
 <META NAME="DC.Subject" LANG="en" CONTENT="Practical use of Dublin Core">
 <META NAME="DC.Subject" LANG="en" CONTENT="Metadata semantics">
 <META NAME="DC.Description" LANG="en" CONTENT="This document is intended as an entry point for users of Dublin Core. For non-specialists, it will assist them in creating simple descriptive records for information resources (for example, electronic documents). Specialists may find the document a useful point of reference to the documentation of Dublin Core, as it changes and grows.">
 <META NAME="DC.Publisher" LANG="en" CONTENT="Dublin Core Metadata Initiative">
 <META NAME="DC.Date" LANG="en" CONTENT="2003-08-26">
 <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/documents/2003/08/26/usageguide/">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsVersionOf
http://dublincore.org/documents/2001/04/12/usageguide/">
<META NAME="DC.Relation" LANG="en" CONTENT="IsBasisFor
http://dublincore.org/resources/translations/">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright © 1995-2005 DCMI All
Rights Reserved. DCMI liability, trademark/service mark, document use and software
licensing rules apply. Your interactions with this site are in accordance with our privacy
statements. Please feel free to contact us for any questions, comments or media inquiries.">
<META NAME="DC.Rights" LANG="en"
CONTENT="http://dublincore.org/about/copyright/index.shtml">

```

Record 4

```

<META NAME="DC.Title" LANG="en" CONTENT="XMLSoftware">
<META NAME="DC.Title" LANG="en" CONTENT="Welcome to XMLSoftware">
<META NAME="DC.Creator" LANG="en" CONTENT="Van den Brink, Linda">
<META NAME="DC.Creator" LANG="en" CONTENT="Tauber, James">
<META NAME="DC.Subject" LANG="en" CONTENT="XML">
<META NAME="DC.Subject" LANG="en" CONTENT="XML creation software">
<META NAME="DC.Subject" LANG="en" CONTENT="XML editors">
<META NAME="DC.Description" LANG="en" CONTENT="A website containing links to
and descriptions of shareware/freeware/Open Source software that is somehow XML-
related.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Van Oven, Barry">
<META NAME="DC.Date" 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" LANG="en"
CONTENT="http://www.xmlsoftware.com/">
<META NAME="DC.Language" LANG="en" CONTENT="en">

```

Record 5

```

<META NAME="DC.Title" LANG="en" CONTENT="International Metadata Initiatives:
Lessons in Bibliographic Control">
<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="Caplan, Priscilla">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata">
<META NAME="DC.Subject" LANG="en" CONTENT="International metadata
initiatives">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata standards comparison">
<META NAME="DC.Description" LANG="en" CONTENT="This paper looks at a subset
of these metadata schemes in more detail: the TEI header, EAD, DCMES, and VRA Core. It
looks at why they developed as they did, major points of difference from traditional library
cataloging, and what advantages they offer to their user communities. It also discusses
challenges to implementers of these schemes and possible future developments. It goes on to

```

identify some commonalities among these cases, and to attempt to generalize from these some lessons for developers of metadata element sets. It concludes by suggesting we also look carefully at emerging schemes being developed by publishers in support of electronic commerce and rights management, and think seriously about the implications of commodity metadata upon our traditional bibliographic apparatus.">

```
<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/caplan_paper.html">
```

```
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

Discussion

The primary issues surrounding record creation in this exercise stem from two major sources: the nature of the original information objects themselves and the metadata schemes used to create the records.

Creating metadata for web-based digital resources is notoriously challenging. Unlike traditional information objects, which typically have a standard format, web-based resources are wildly inconsistent. For example, finding the title, author, publisher, and copyright date on most books is straightforward. Web-based resources, however, have no standardized title page from which to draw such information. Some might have an “about” section set aside for that purpose, but others have the information tucked away in their source code and still others refuse to list it at all. To borrow the previous metaphor, such digital resources hide the appropriately shaped pegs that should fit into the requisite holes—or, at best, they offer amorphous pegs that may not seem to fit any hole well.

Record number 4 above particularly illustrates this point. The resource itself is a website entitled *XMLSoftware*, and the only unambiguous information that it presents is its title and its URL. For the purpose of creating both the ERC and the DC record, the rest of the information had to be inferred—the statement at the top of the website’s home page indicates that Linda van den Brink and James Tauber worked together to create the website and that Barry van Oven helped to design it, but it does not state as much outright. Dates associated with the website are equally unclear—updates, which seem to occur several times a week, are listed, but not dates associated with the resource’s creation. As the latest update’s full date changes every few days, only the year (2005) seemed appropriate to list in the metadata records.

Interestingly, the digital resources that most closely resemble traditional ones were the easiest for which to create metadata: e.g., the resource associated with record number 1, an article entitled *Metadata Demystified: A Guide for Publishers*, contains recognizable and easy-to-find self-descriptive information. This particular resource is familiar because it appears to have been scanned directly from a print article, but to derive similarly useful information metadata creators must plumb the depths of resources that have no such print doppelgangers.

Besides that deriving metadata from web-based resources is naturally tricky, the particular standards that this exercise employed only added to the challenge. ERC and DC both have vague syntax and vague semantics. Unlike a standard such as MARC, which has hundreds of elements defined in painstaking detail, ERC and DC have very few elements, all of which are left open to interpretation. One could speculate that, because MARC was originally created with traditional resources in mind (whose formats are better defined than those of electronic resources, as discussed above), its elements and semantics are specifically tailored to that context. On the other hand, ERC and DC were created for electronic resources, thus justifying those standards' ambiguity. To again use the metaphor, ERC and DC comprise amorphous holes able to accommodate pegs with similar—but different—shapes.

To further complicate matters, ERC's vagueness differs from DC's vagueness. Indeed, the two do share many similarities: e.g., both are meant for electronic resources and both have minimal versions that can be extended when further definition is necessary. Their differences become apparent, however, when one attempts to use them to create metadata records. For the purpose of comparison, using each metadata standard's most basic element set seemed appropriate; even so, the experience of creating ERC records varied considerably from that of creating DC records.

ERC's basic element set contains only four nebulously defined elements—"who", "what", "when", and "where." The simplest ERC record, which contains one instance of these four elements (the "anchoring story"), can log any aspect of the information object that the record creator deems the most fundamental. Despite these specifications' fuzziness, they did not prove difficult to employ. In fact, creating the five ERC records took little time at all, except where noted earlier regarding the *XMLSoftware* website. Because the resources involved are relatively similar to one another in purpose, standardizing the four elements' semantics was helpful. "Who" translated into "author(s)," "what" translated into "title," "when" translated into "creation date," and "where" translated into "URL." All five resources seemed to share these as their most basic attributes.

Despite the ease with which they were created, the ERC records do not seem particularly useful. They are suited for rough identification and resource acquisition, but without any subject-related details or more specific identifiers, the records would probably not support the finding, selection, or identification tasks on all but the most superficial of levels.

While ERC's elements and semantics are extremely vague, those of DC have proven to be only moderately so—DC's basic element set comprises fifteen elements that are more precisely defined than those of ERC yet still remain quite broad. Additionally, they are both repeatable and optional. Thus, if a particular element does not apply to a resource, then including it in the metadata record is not mandatory.

The difficulty of using DC seems to lie in its moderate ambiguity—its semantics are not so fuzzy that they can mean almost anything (as in ERC), but they are vague enough to be unclear and therefore impractical without further explanation. Indeed, creating the DC records required the use of a helper document, *Using Dublin Core*, which is also the resource tied to the third record in this exercise's record set. With the aid of this guide, the DC semantics seemed clearer, but record creation still took significantly more time using DC

than it did using ERC—presumably because even DC’s simple form uses a richer set of elements than does that of ERC.

Due to its semantic richness, the DC records seem better suited than the ERC records to support the four basic user tasks. The subject-related metadata elements, the “description” element, and the “type” and “format” elements provide much more information about each resource than the basic ERC format allows. The DC elements that provide related resources add yet another level of depth, although these elements also add considerably to the time required for record creation.

Based on this analysis, then, ERC seems better for creating quick and dirty metadata records. When time or training is a factor in metadata creation, or when a short record is needed only to identify an item, ERC would be a natural choice. DC, on the other hand, seems better for creating records that support resource discovery.

One important caveat should be noted: using the extended versions of these two schemes would provide a much different experience. Both were created with flexibility and extensibility in mind (again, because they are to be used for describing a wide range of electronic resources), which would explain why their basic element sets seem anemic compared to the element sets of other standards.

For the same reason, ERC and DC are as syntactically vague as they are semantically vague. Due to the use of the generic metadata tool in this exercise, however, syntax was not much of an issue—except that care had to be taken to format names and dates consistently.

Summary and Conclusion

As with many endeavors, this exercise’s most interesting—and difficult—aspect was dealing with ambiguities. Besides the standard ambiguity that creating metadata for electronic resources typically involves, the chosen metadata schemes seem to exploit semantic and syntactic ambiguity in different ways for different effects. Thus, not only did the experience of creating ERC records vary from that of creating DC records, but the ERC and DC records themselves also differ and offer varied levels of utility.

For those who do not know any better, creating metadata for digital resources is generally a mere afterthought. After all, what could be simpler than putting a title into a “title” field, an author’s name into an “author” field, and a few subject-related keywords into a “subject” field? A trained monkey could do as much, right? On the contrary, as this exercise has shown, creating *useful* metadata records is a complex task: one must consider an appropriate metadata scheme to use, gain a complete understanding of the chosen scheme, and thoughtfully populate each element.