

## Introduction

This exercise began the extraction of metadata from five collection objects into two metadata frameworks, ERC and Dublin Core. It includes a brief analysis of the method used and issues encountered; and concludes with a summary of the lessons learned.

## Electronic Resource Citation Records

Metadata Demystified: A Guide for Publishers. by Amy Brand, Frank Daly, and Barbara Meyers. 2003.

```
erc:
  what:      Metadata Demystified: A Guide for Publishers
  who:       Amy Brand | Frank Daly | Barbara Meyers
  when:      20030715063455
  where:     http://www.niso.org/standards/resources/Metadata_Demystified.pdf

erc-about:
  who:       publishers | readers
  what:      Metadata | Dublin Core | ONIX | OAI-PMH
  when:      2001
  where:     http://www.niso.org/ | http://www.sheridanpress.com/
```

Canadian Heritage Information Networked (CHIN). Metadata Standards for Museum Cataloguing.

```
erc:
  who:       Canadian Heritage Information Networked
  what:      Metadata Standards for Museum Cataloguing
  when:      20040820
  where:     http://www.chin.gc.ca/English/Standards/metadata_description.html
```

Using Dublin Core. Diane Hillman. 2003.

```
erc:
  who:       Diane Hillman
  what:      Using Dublin Core
  when:      20030826
  where:     http://dublincore.org/documents/usageguide/
```

XML Software [Website]. 2003.

```
erc:
  who:       James Tauber | Linda van den Brink
  what:      XMLSoftware
  when:      20050211
  where:     http://www.xmlsoftware.com/
```

International Metadata Initiatives: Lessons in Bibliographic Control. Priscilla Caplan. 2000.

```
erc:
  who:       Priscilla Caplan
  what:      International Metadata Initiatives: Lessons in Bibliographic Control
  when:      20020307
  where:     http://www.loc.gov/catdir/bibcontrol/caplan_paper.html
```

## Dublin Core Records

Metadata Demystified: A Guide for Publishers. by Amy Brand, Frank Daly, and Barbara Meyers. 2003.

```

<META NAME="DC.Title" LANG="en" CONTENT="Metadata Demystified: A
Guide for Publishers" />
<META NAME="DC.Creator" LANG="en" CONTENT="Amy Brand" />
<META NAME="DC.Creator" LANG="en" CONTENT="Frank Daly" />
<META NAME="DC.Creator" LANG="en" CONTENT="Barbara Meyers" />
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata" />
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core" />
<META NAME="DC.Subject" LANG="en" CONTENT="ONIX" />
<META NAME="DC.Subject" LANG="en" CONTENT="OAI-PMH" />
<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." />
<META NAME="DC.Publisher" LANG="en" CONTENT="NISO Press" />
<META NAME="DC.Publisher" LANG="en" CONTENT="Sheridan Press" />
<META NAME="DC.Date" LANG="en" CONTENT="2003-07-15" />
<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:1880124599" />
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://www.niso.org/standards/resources/Metadata_Demystifie
d.pdf" />
<META NAME="DC.Language" LANG="en" CONTENT="en" />

```

#### Canadian Heritage Information Networked (CHIN). Metadata Standards for Museum Cataloguing.

```

<META NAME="DC.Title" LANG="en" CONTENT="Metadata Standards for
Museum Cataloguing" />
<META NAME="DC.Creator" LANG="en" CONTENT="Canadian Heritage
Information Networked" />
<META NAME="DC.Subject" LANG="en" CONTENT="metadata" />
<META NAME="DC.Description" LANG="en" CONTENT="An overview of
general and specific standards for the description of museum
collections, and specific objects within collections " />
<META NAME="DC.Publisher" LANG="en" CONTENT="Government of Canada,
Canadian Heritage" />
<META NAME="DC.Date" LANG="en" CONTENT="2004-08-20" />
<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.chin.gc.ca/English/Standards/metadata_descriptio
n.html" />
<META NAME="DC.Language" LANG="en" CONTENT="en" />
<META NAME="DC.Language" LANG="en" CONTENT="fr" />

```

#### Using Dublin Core. Diane Hillman. 2003.

```

<META NAME="DC.Title" CONTENT="Using Dublin Core" />
<META NAME="DC.Creator" CONTENT="Diane Hillmann" />
<META NAME="DC.Subject" LANG="en" CONTENT="metadata" />
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin core" />
<META NAME="DC.Description" 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" 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/usageguide/" />
<META NAME="DC.Language" LANG="en" CONTENT="en" />

```

#### XML Software [Website]. 2003.

```

<META NAME="DC.Title" LANG="en" CONTENT="XMLSoftware" />
<META NAME="DC.Creator" LANG="en" CONTENT="James Tauber">
<META NAME="DC.Creator" LANG="en" CONTENT="Linda van den Brink">
<META NAME="DC.Subject" LANG="en" CONTENT="XML (software)" />
<META NAME="DC.Description" LANG="en" CONTENT="A collection of
software related to the development of systems that implement XML
and XML related technologies, as well as news about such software."
/>
<META NAME="DC.Date" LANG="en" CONTENT="2005-02-11" />
<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" />

```

#### International Metadata Initiatives: Lessons in Bibliographic Control. Priscilla Caplan. 2000.

```

<META NAME="DC.Title" LANG="en" CONTENT="International Metadata
Initiatives: Lessons in Bibliographic Control">
<META NAME="DC.Creator" LANG="en" CONTENT="Priscilla Caplan" />
<META NAME="DC.Subject" LANG="en" CONTENT="metadata" />
<META NAME="DC.Subject" LANG="en" CONTENT="ONIX" />
<META NAME="DC.Description" LANG="en" CONTENT="A high level overview
of technical metadata initiatives with brief descriptions." />
<META NAME="DC.Date" LANG="en" CONTENT="2002-03-07" />
<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

*There was a little girl,  
 Who had a little curl,  
 Right in the middle of her forehead.  
 When she was good,  
 She was very good indeed,  
 But when she was bad she was horrid.*

-- Henry Wadsworth Longfellow

Like the poet's little girl, good metadata is very good indeed, but when it is bad, it is horrid. In examining two (of many) formats for describing collections of intellectual output and related ephemera – the Dublin Core Metadata Initiative, and John Kunze's Electronic Resource Citation – it helps to bear in mind that while the selection of a format that is appropriate to a given, specific task can be a boon to the usability and value of the resulting metadata.

In late 1996, I was producing the California State Library Web site, largely by hand, when my client and mentor, John Jewell, told me about Dublin Core, and how great that it would be if we could implement it. I investigated it and found some issues: It would require a large commitment of time from library staff and me to implement and maintain; it did not fit into our relatively inflexible, but necessary content management scheme; and there was no practical way to get our relatively inflexible, but necessary indexing tool – early Verity – to recognize it. While the technical issues are now manageable, the human resource issue has remained intractable, and has remained the major impediment to broad and consistent implementation.

Another aspect of this issue is that the data that is directly available from electronic files including the Web pages and sites, Adobe Acrobat (.pdf) documents, word processing documents (.doc, .rtf, .wpd, etc.), image files, and text files found on the Internet and from other sources has no consistency whatsoever. Most of these formats, save text files, have some provision for metadata, but no rules for their use, and most of the tools for creating and maintaining these items have provision for entering and editing metadata. Unfortunately, few of these tools provide much assistance or encouragement in this task.

In extracting metadata from the five assigned objects, it became clear to me that the most book-like object, the NISO guide, was the easiest to work with. The DCMI guide, with its attached RDF document, was also easy to work with, but did not use as many of the Dublin Core elements as I would have expected. On the other hand, the CHIN document had embedded metadata, but it was largely unusable. For example, the embedded creation/revision date is “2002014”, whatever that might mean. The XMLSoftware Web site, on the other hand required investigation, interpretation, and informed guesswork to provide all but the most basic information.

Although I did not use the full extent of ERC’s capability to build stories in this exercise, the difference between ERC’s hierarchical, extensible structure and Dublin Core’s flat structure influenced my approach to extracting metadata from these objects. I found it somewhat surprising that, while I thought that this would lead to a clear preference for one structure over the other, it did not. It did lead me to ponder what alternative structure(s) could carry the benefits of both, and build on them.

## Summary and Conclusion

Lessons learned in completing this exercise:

- Simple tools like the Generic Metadata Editor can be a useful starting place, resulting in some time savings, but they seldom produce a finished product on their own.
- Manual metadata extraction, without controlled vocabularies, and strict syntax control, leads to inconsistent results.
- It is not possible to determine an optimum balance between detail and concision in creating records without having the context of their intended use, or a ruleset that would render such consideration superfluous.
- Just because the data comes from a generally reliable source, there is no guarantee that metadata from the same source will be reliable, or even useful.
- Placing metadata in two dissimilar contexts, the flat Dublin Core and the hierarchical ERC, can give the editor a nuanced view of the source material.

Perhaps the quip, *If it was easy, anyone could do it*, should be updated to *If it was describable with a simple ruleset, a machine could do it*.

