

Metadata Creation Report: ERC and DC

Introduction

This report describes the results of creating metadata based on a set of five digital objects. For creating metadata, two different metadata schemes are used: Electronic Resource Citation (ERC) and Dublin Core (DC). Besides Introduction, this report consists of four sections such as ERC records, DC records, Discussion, and Summary/Conclusion. The ERC and DC records sections show the actual metadata records applied to the five objects, respectively. On Discussion section, it identifies properties and attributes of each object, and describes decision making procedure for metadata elements, some observations, and issues pertinent to creating metadata records.

Electronic Resource Citation Records

- Introduction to Metadata: Pathways to digital Information**
erc:
who: Baca, Murtha | Gilliland-Swetland, Anne J. | Gill, Tony | Woodley, Mary
what: Introduction to metadata: Pathways to digital Information | Setting the Stage | Metadata and the World Wide Web | Crosswalks: The Path to Universal Access?
when: 2000
where: http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html
- Metadata**
erc:
who: UKOLN | UK Office for Library Networking | Day, Michael | Powell, Andy
what: metadata
when: 2005 01 18
where: <http://www.ukoln.ac.uk/metadata/>
- ANSI/NISO Z39.85 – 2001 Dublin Core Metadata Element Set**
erc:
who: National Information Standard Organization
what: The Dublin Core Metadata Element Set
when: 2001 09 10
where: <http://www.niso.org/standards/resources/Z39-85.pdf?CFID=179761&CFTOKEN=46984796>
- XML Organic Bibliographic Information Schema**
erc:
who: Stanford University Medical Center | Lane Medical Library | MedLane Project Team | MedLane
what: The XML Organic Bibliographic Information Schema
when: 2003 10 27
where: <http://laneweb.stanford.edu:2380/wiki/medlane/schema>
- Metadata Harvesting and the Open Archives Initiative**
erc:
who: Lynch, Clifford A.
what: Metadata Harvesting and the Open Archives Initiative
when: 2001 08 00
where: <http://www.arl.org/newsltr/217/mhp.html>

Dublin Core Records

□ Introduction to Metadata: Pathways to digital Information

```
<META NAME="DC.Title" LANG="en" CONTENT="Introduction to metadata: pathways digital
information">
<META NAME="DC.Title" LANG="en" CONTENT="Setting the stage">
<META NAME="DC.Title" LANG="en" CONTENT="Metadata and the World Wide Web">
<META NAME="DC.Title" LANG="en" CONTENT="Crosswalks: the path to universal access?">
<META NAME="DC.Creator.Editor" LANG="en" CONTENT="Baca, Murtha">
<META NAME="DC.Creator.Author" LANG="en" CONTENT="Gilliland-Swetland, Anne J.">
<META NAME="DC.Creator.Author" LANG="en" CONTENT="Gill, Tony">
<META NAME="DC.Creator.Author" LANG="en" CONTENT="Woodley, Mary">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata"; "Standards"; "Information
Standards"; "world wide web"; "crosswalks"; "cataloging"; "Art"; "Architecture"; "Cultural Information";
"Dublin Core"; "USMARC"; "EAD"; "ISAD"; "Encoded Archival"; "International Standard Archival
Description"; "DC">
<META NAME="DC.Description" LANG="en" CONTENT="As an introduction to metadata, it contains
three articles, crosswalks, glossary, acronyms, and urls">
<META NAME="DC.Publisher" LANG="en" CONTENT="The Getty Research Institute">
<META NAME="DC.Contributor" LANG="en" CONTENT="Gill, Tony">
<META NAME="DC.Contributor" LANG="en" CONTENT="Gilliland-Swetland, Anne J.">
<META NAME="DC.Contributor" LANG="en" CONTENT="Woodley, Mary S.">
<META NAME="DC.Date" LANG="en" CONTENT="2000">
<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.getty.edu/research/conducting_research/standards/intrometadata/index.html"
>
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en" CONTENT="Term of Use/Copyright">
<META NAME="DC.Rights" LANG="en" CONTENT="http://www.getty.edu/legal/copyright.html">
```

□ Metadata

```
<META NAME="DC.Title" LANG="en" CONTENT="Metadata">
<META NAME="DC.Creator" LANG="en" CONTENT="metadata group at UKOLN">
<META NAME="DC.Creator" LANG="en" CONTENT="UKOLN">
<META NAME="DC.Creator" LANG="en" CONTENT="Day, Michael">
<META NAME="DC.Creator" LANG="en" CONTENT="Powell, Andy">
<META NAME="DC.Creator" LANG="en" CONTENT="UK Office for Library Networking">
<META NAME="DC.Subject" LANG="en" CONTENT="metadata"; "Dublin Core"; "Schema"; "Z39.50";
"DC"; "Metadata portal"; "UKOLN">
<META NAME="DC.Description" LANG="en" CONTENT="Metadata related projects, resources,
tutorials, initiatives, metadata registries by UKOLN">
<META NAME="DC.Publisher" LANG="en" CONTENT="UKON">
<META NAME="DC.Publisher" LANG="en" CONTENT="University of Bath">
<META NAME="DC.Contributor" LANG="en" CONTENT="Dowdell, Pete">
<META NAME="DC.Contributor" LANG="en" CONTENT="Duke, Monica">
<META NAME="DC.Contributor" LANG="en" CONTENT="Guy, Marieke">
<META NAME="DC.Contributor" LANG="en" CONTENT="Heery, Rachel">
<META NAME="DC.Contributor" LANG="en" CONTENT="Hunter, Philip">
<META NAME="DC.Contributor" LANG="en" CONTENT="Johnston, Pete">
<META NAME="DC.Contributor" LANG="en" CONTENT="Patel, Manjula">
<META NAME="DC.Contributor" LANG="en" CONTENT="Robinson, Bridget">
```

```
<META NAME="DC.Date" LANG="en" CONTENT="2005-01-18">
<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.ukoln.ac.uk/metadata/">
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

□ **ANSI/NISO Z39.85 – 2001 Dublin Core Metadata Element Set**

```
<META NAME="DC.Title" LANG="en" CONTENT="The Dublin Core Metadata Element Set">
<META NAME="DC.Creator" LANG="en" CONTENT="National Information Standards
Organization">
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core"; "Metadata", "Element Set";
"Standard"; "Information Standard"; "resource description"; "Cataloging of computer network
resources">
<META NAME="DC.Subject.LCC" LANG="en" CONTENT="Z695.24 N38 2001">
<META NAME="DC.Subject.DCC" LANG="en" CONTENT="025.3'44—dc21">
<META NAME="DC.Description" LANG="en" CONTENT="Defines fifteen metadata elements for
resource description in a cross disciplinary information environment">
<META NAME="DC.Publisher" LANG="en" CONTENT="National Information Standards
Organization">
<META NAME="DC.Contributor" LANG="en" CONTENT="American National Standards Institute">
<META NAME="DC.Date" LANG="en" CONTENT="2001-09-10">
<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/Z39-
85.pdf?CFID=179761&CFTOKEN=46984796">
<META NAME="DC.Source" LANG="en" CONTENT="ISBN: 1-880124-53-X">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsFormatOf ANSI/NISO Z3985 -2001 Dublin
Core Metadata Element Set">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright ©2001 by the National Information
Standards Organization All rights reserved under International and Pan-American Copyright
Conventions. No part of this book may be reproduced or transmitted in any form or by any means,
electronic or mechanical, including photocopy, recording, or any information storage or retrieval
system, without prior permission in writing from the publisher. All inquiries should be addressed to
NISO Press, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814.">
```

□ **XML Organic Bibliographic Information Schema**

```
<META NAME="DC.Title" LANG="en" CONTENT="xobis: the xml organic bibliographic information
schema">
<META NAME="DC.Creator" LANG="en" CONTENT="Lane Medical Library, Stanford University
Medical Center">
<META NAME="DC.Subject" LANG="en" CONTENT="Standard"; "XML"; "Bibliographic Information";
"Information Standard"; "Medical Information"; "MARC"; "xobis">
<META NAME="DC.Description" LANG="en" CONTENT="XML schema available for modeling full
set of MARC data">
<META NAME="DC.Date" LANG="en" CONTENT="2003-10-27">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Format" LANG="en" CONTENT="application/pdf">
<META NAME="DC.Identifier" LANG="en"
CONTENT="http://laneweb.stanford.edu:2380/wiki/medlane/schema">
<META NAME="DC.Language" LANG="en" CONTENT="en">
```

□ **Metadata Harvesting and the Open Archives Initiative**

```

<META NAME="DC.Title" LANG="en" CONTENT="Metadata Harvesting and the Open Archives Initiative">
<META NAME="DC.Creator" LANG="en" CONTENT="Lynch, Clifford A.">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata"; "Harvesting"; "Open Archives Initiative"; "Protocol"; "distributed networked information services";>
<META NAME="DC.Description" LANG="en" CONTENT="Open Archives Metadata Harvesting Protocol, an infrastructure component for supporting distributed networked information services">
<META NAME="DC.Publisher" LANG="en" CONTENT="Association Research Libraries">
<META NAME="DC.Publisher" LANG="en" CONTENT="ARL">
<META NAME="DC.Date.Created" LANG="en" CONTENT="2001-08">
<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.arl.org/newsltr/217/mhp.html">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright © 2001 Clifford A. Lynch">

```

Discussion

The main idea of ERC is to keep it simple and compact with a minimal description, so it can be easily developed without using complicated mediation software or high skilled abilities. This report keeps to follow the principle of the simplicity and the requirement of using four elements for anchoring story, who, what, when, and where. On the other hand, DC has more complicated purposes to achieve resource description and resource discovery. To accomplish both purposes, there are fundamental principles for DC: intrinsicality, extensibility, syntax independence, optionality and repeatability. The core principle of intrinsicality resembles the procedure of cataloging in libraries in describing elements in hands. To hold the needs from different communities, the extensibility of DC comes up. One way of achieving extensibility is to define additional elements fit in specific community needs. The other way of extensibility is to refine each of the core 15 elements. This report uses the refinement to elements as needed. The syntax of DC is not fixed, but this report uses the html metatags brought from the NoteTab Light application. The last two principles, optionality and repeatability, are massively used in this report. While the optionality is that any of elements can be not used at all, the repeatability is that any of elements can be repeated as many as needed.

For a set of five objects, each object has its own attributes and properties. Based on the identification of the attributes and properties of each object, observations and discussions follow for creating metadata records using ERC and DC. The basic idea of decision making procedure in creating metadata records is to pursue the earlier mentioned fundamental principles of both metadata schemes: ERC and DC.

1) "Introduction to metadata: Pathways to digital Information" object can be considered as an edited collection, edited by Murtha Baca, and mainly contains three articles by different authors. As the articles in it weigh as much as the object itself, it is reasonable to describe them as metadata elements. In the case of ERC, they can be described as multiple titles and multiple authors for them on "who" and "what" elements by using the syntax of putting "|" between data. In DC, "DC.title" for the title of the articles is used as many times as it needs to be used. It comes from one of the principles of DC metadata scheme, repeatability. In addition to the multiple titles, even though different authors for three articles are described as "DC,Contributor" on the object and they are assigned as so, they are assigned by "DC.Creator.Author" as well. Since the extensibility/modifiability of DC principles provide the description with some extent of refining capability, so it is possible for metadata creators to describe them more specifically. In this case, metadata creators are able to specify the name of separate articles' authors as "DC.Creator.Author" and "DC.Creator.Editor". The other crucial topic of this object deals with crosswalk for different metadata schemes and presents them with tables between different schemes. To provide this information as access points with users, the subject terms are assigned as the name of each metadata scheme and

crosswalk as well. On the other hand, as this object contains a lengthy term of use and copyright statements, it is described as an URL pointing to the copyright statement.

2) "Metadata" object is a portal to the metadata subject and maintained by mainly Michael Day, Andy Powell and staffs of UKOLN (UK Office for Library Networking). The topics of this object contain metadata related current and past projects, publications, resources, tutorials, and metadata registries. For ERC, in case of "who" as a responsible person or party, multiple persons and party are assigned. For DC, "DC.Creator" is used repeatedly for describing multiple representations for the Creator element. In addition to "DC.Creator", the object contains a list of staffs working on metadata related projects. As this staff list seems to be considered as contributors for this object, eight names are assigned as "DC.Contributor". The "DC.Subject" is assigned with more comprehensive terms like "metadata portal" and "metadata".

3) "ANSI/NISO Z39.85-2001 the Dublin Core Metadata Element Set" object is a periodical publication for international standard published electronically. One feature of this object contains multiple identifiers such as URL, ISBN, and ISSN. Instead of using multiple of "DC.Identifier", one identifier is chosen and it is URL for this object. This decision is made by the DC's one-to-one correspondence between a record and an object. For ISBN and ISSN version, another record should be described. Instead, there is a relation between the printed version and this object. It is described as "DC.Source" and "DC.Relation". For ERC, the same principle with DC is applied, so only one identifier which is the electronic version location is used. Also, subject and keywords for "DC.Subject" are chosen from three different categories such as keywords from the contents, Library of Congress Classification (LCC) number, and Dewey decimal classification (DCC) number since it has LCC and DDC number in it. Assigning LCC and DDC numbers to "DC.Subject" is following the best practice in DC.

4) "XML Organic Bibliographic Information Schema" object mainly contains one specific project, XML schema for MARC data, related information. When trying to identify a responsible person or party, it is vague within the object itself. However, using related links, the "who" for ERC and "DC.Creator" for DC can be identified to the organization, "Lane Medical Library, Stanford University Medical Center", according to the parent URL for this object. Also, as this object has both html and pdf format contents, accordingly, "DC.Format" are used repeatedly for both formats.

5) "Metadata Harvesting and the Open Archives Initiative" is an article published in ARL Bimonthly Report 217 and similar to the traditional journal article. The elements from ERC and DC are directly assigned from the object, respectively. For the date of this object, it is different in nature with other objects. While other objects are supposed to be updated anytime when needed, this object pretty much resembles the printed information environment. This object has its fixed creation (printed, in words of the traditional information environment) date and will not allow change its contents without notifying revision date. In case of DC, it has the capability to refine the element as a qualified term, "DC.Date.Created".

Secondly, in this records creation, keywords from the contents of objects are used for "DC.Subject". This decision is made based on the mentality of metadata initiatives. The metadata creation is for general users who can create objects, but not necessarily skilled with assigning terms to objects with controlled vocabularies. Probably, the most reasonable approach to accommodate this mentality is to assign plain keywords from the contents to objects. By this practices, however, there have been vocabulary issues whether used terms represent its topicality and they can be used for interchanging each other between different systems, especially for "DC.Subject" of DC and "ERC: about, what" of ERC. Those issues are closely related with choosing terms representing the object contents. For metadata creators, there are options for their choices. According to the DC and ERC standards or best practices, they are able to use terms from controlled vocabularies such as Library of Congress Subject Headings (LCSH), any domain specific thesauri or classification numbers (e.g. LCC or DDC). Just plain keywords from the object

contents are allowed to use as well. Since applying subject terms to objects are not strictly defined, there have existed several dialects for describing the same object in terms of subject. Problems happen to be serious when the data try to interoperate between different systems or institutions. As metadata are born with the properties of the networked information environment and have to interchange data between different institutions, organization, or systems, dissonance of subject terms or keywords for the same objects could be relatively a huge issue compared with the traditional information environment. Lately, some approaches have tried to reduce the dissimilarity in subject terms or keywords between different systems. One approach of them is to build a hierarchical umbrella thesaurus¹ (e.g. WordNet) for them to connect each other, even though they use different vocabulary. The other approach is to use the machine learning algorithms to predict subject terms² based on documents analysis with previously assigned subject terms or pure clustering techniques.

Summary and Conclusion

This report shows the results of creating metadata records for five objects using two different metadata schemes: ERC and DC. The procedure for creating records in both schemas is discussed. Based on the identification of each object with the properties and attributes, the data of each element are assigned. Decision made during records creation is mainly embedded on the fundamental principles of schemas, rules, and best practices found in the related literatures. One crucial issue of assigning subject terms to objects is discussed. Even though the issue has existed at least since the classification scheme has existed, metadata in networked information environment has more weighed on it. A couple of recent approaches to the issue are mentioned as an umbrella thesaurus approach and machine learning technique approach. The metadata creation is hard to keep only in control of catalogers or people in library communities, because the web based information with quality has been increasing exponentially. The metadata to improve the quality of resource description and discovery could evolve in the way of embracing billions of naïve metadata records into a controlled and quality structure.

¹ Mihalcea, Rada. (2003). Turning WordNet into an information retrieval resource: systematic polysemy and conversion to hierarchical codes. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(5):689-704.

² Sebastiani, Fabrizio. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1):1-47