

Creating Metadata Records

Introduction

This report is a description of the work performed to create metadata records for five example objects. The example objects are electronic documents in various forms including HTML text, PDF text, and a Web page. Two metadata records are created for each object, using both the Electronic Resource Citation (ERC) and the Dublin Core (DC) schemes. The ERC records are created using a simple syntax of "<element>: <data>". The DC records are created using a very basic HTML syntax, but is in no way intended to represent functioning HTML code. This report contains the ten metadata records created (5 ERC and 5DC for each object) and a discussion of the positive and negative experiences had in creating them.

Electronic Resource Citation Records

- 1) **erc:**

who: National Information Standards Organization (NISO)
what: Understanding Metadata
when: 2004
where: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
#-----
erc-about:

what: Metadata | Metadata Schemes | Element Sets | Dublin Core | DC | TEI | METS | MODS | EAD | LOM | <indec> | CDWA | ONIX | VRA | MPEG | FGDC | DDI
#-----
erc-contribution:

who:, Guenther, Rebecca | Radebaugh, Jacqueline
#-----
erc-from:

who/Revision:, Hodge, Gail
what/Revision: Metadata Made Simpler: A guide for libraries
when/Revision: 2001
where/Revision: http://www.niso.org/news/Metadata_simpler.pdf

- 2) **erc:**

who:, Le Boeuf, Patrick
what: Brave New FRBR World
when: (:unkn)
where: http://www.ddb.de/news/pdf/papers_leboeuf.pdf
#-----
erc-about:

what: functional requirement for bibliographic records | frbr | IFLA

- 3) **erc:**
 ###
who: The International DOI Foundation (IDF)
what: The Digital Object Identifier System
when/Updated: 2004 11 23
where: <http://www.doi.org/index.html>
 #-----
erc-about:
 #####
what: digital object identifiers | doi | idf | handle system | persistent identifier
- 4) **erc:**
 ###
who: Research Libraries Group, Inc.
what: Trusted Digital Repositories: Attributes and Responsibilities
when/published: 2002 05
where: <http://www.rlg.org/longterm/repositories.pdf>
 #-----
erc-about:
 #####
what: digital repository | open archival information system | oais | certification
- 5) **erc:**
 ###
who:, Miller, Eric
what: An Introduction to the Resource Description Framework
when: 1998 05
where: <http://www.dlib.org/dlib/may98/miller/05miller.html>
 #-----
erc-about:
 #####
what: resource description framework | rdf | xml

Dublin Core Citation Records

- 1) #Dublin Core
 #####
 <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; Metadata Schemes; Element
 Sets; Dublin Core; DC; TEI; METS; MODS; EAD ; LOM; <indec>; CDWA; ONIX; VRA;
 MPEG; FGDC; DDI">
 <META NAME="DC.**Description**" LANG="en" CONTENT="An introduction to the concept of
 Metadata and its purposes. Includes discussions on structuring Metadata, a look at some
 Metadata schemes and element sets, and the interoperability of Metadata">
 <META NAME="DC.**Publisher**" LANG="en" CONTENT="NISO Press">
 <META NAME="DC.**Contributor**" LANG="en" CONTENT="Guenther, Rebecca; Radebaugh,
 Jacqueline">
 <META NAME="DC.**Date**" 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" 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.Language" LANG="en" CONTENT="en">
<META NAME="DC.Relation" LANG="en" CONTENT="IsVersionOf Revision and expansion of
  Gail Hodge's 'Metadata Made Simpler: A guide for libraries', 2001">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright NISO 2004">

```

2) # Dublin Core

```

#####
<META NAME="DC.Title" LANG="en" CONTENT="Brave New FRBR World">
<META NAME="DC.Creator" LANG="en" CONTENT="Patrick Le Boeuf">
<META NAME="DC.Subject" LANG="en" CONTENT="functional requirement for bibliographic
  records; frbr; IFLA">
<META NAME="DC.Description" LANG="en" CONTENT="A discussion of the Funtional
  Requirements for Bibliographic Records (FRBR) as developed for IFLA. Includes a history,
  current implementations, and problems associated with FRBR.">
<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.ddb.de/news/pdf/papers\_leboeuf.pdf">
<META NAME="DC.Language" LANG="en" CONTENT="en">

```

3) #Dublin Core

```

#####
<META NAME="DC.Title" LANG="en" CONTENT="The Digital Object Identifier (DOI) System">
<META NAME="DC.Creator" LANG="en" CONTENT="The International DOI Foundation (IDF)">
<META NAME="DC.Subject" LANG="en" CONTENT="digital object identifiers; doi; idf; handle
  system; persistent identifier">
<META NAME="DC.Description" LANG="en" CONTENT="A Web site devoted to the digital
  object identifier system, which creates persistent identifiers for digital objects using handles.
  The site is run by the International DOI Federation (IDF).">
<META NAME="DC.Date.modified" LANG="en" CONTENT="2004-11-23">
<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.doi.org/index.html">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright International DOI Foundation -
  All rights reserved.">

```

4) #Dublin Core

```

#####
<META NAME="DC.Title" LANG="en" CONTENT="Trusted Digital Repositories: Attributes and
  Responsibilities">
<META NAME="DC.Creator" LANG="en" CONTENT="RLG/OCLC Working Group on Digital
  Archive Attributes">
<META NAME="DC.Subject" LANG="en" CONTENT="digital repository; open archival
  information system; oais; preservation">
<META NAME="DC.Description" LANG="en" CONTENT="An investigation of trusted digital
  repositories--their definition, attributes, resposibilites and certification. Concludes with
  recommendations. Contains an OAIS Technical Overview in appendix.">
<META NAME="DC.Publisher" LANG="en" CONTENT="RLG, Inc.">

```

```

<META NAME="DC.Contributor" LANG="en" CONTENT="<META NAME="DC.Contributor"
  LANG="en" CONTENT="Neil Beagrie; Dr. Marianne Doerr; Dr. Margaret Hedstrom; Maggie
  Jones; Anne Kenney; Catherine Lupovici; Kelly Russell; Colin Webb; Deborah Woodyard;
  Robin Dale; Meg Bellinger">
<META NAME="DC.Date" LANG="en" CONTENT="2002-05">
<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.rlg.org/longterm/repositories.pdf">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright Research Libraries Group 2002
  - All rights reserved.">

```

5) #Dublin Core

```
#####
```

```

<META NAME="DC.Title" LANG="en" CONTENT="An Introduction to the Resource Description
  Framework">
<META NAME="DC.Creator" LANG="en" CONTENT="Miller, Eric">
<META NAME="DC.Subject" LANG="en" CONTENT="resource description framework; rdf;
  xml">
<META NAME="DC.Description" LANG="en" CONTENT="An introduction to the Resource
  Description Framework (RDF), including its background, data model, syntax and schema.">
<META NAME="DC.Publisher" LANG="en" CONTENT="D-Lib Magazine">
<META NAME="DC.Date" LANG="en" CONTENT="1998-05">
<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.dlib.org/dlib/may98/miller/05miller.html">
<META NAME="DC.Identifier" LANG="en" CONTENT="ISSN: 1082-9873">
<META NAME="DC.Language" LANG="en" CONTENT="en">
<META NAME="DC.Rights" LANG="en" CONTENT="Copyright Eric Miller 1998">

```

Discussion

The Note Tab Light editing tool was used to create the metadata records. Using the tool was intuitive and it provided basic syntax and a general framework for the metadata's presentation. For the ERC records this was fairly simple as there were only four elements—who, what, when and where. The tool was much more helpful with DC record creation, providing a basic HTML syntax and interactive menus to add qualifiers to some elements.

For all of the objects locating appropriate bibliographic information was straightforward, though some objects did require more hunting than others. In all cases though, enough information was found to provide a distinctive and accurate description of the objects. The ability to fully utilize the bibliographic information available for the objects varied however depending on the metadata scheme used.

Implementation of the ERC scheme is simple and straightforward. The ERC elements of who, what, when and where are clear and generally understood, but open-ended enough to represent different data depending on the surrounding story. The requirement that all four elements be present in the anchor story (or the reason for an element's absence be noted) provides a solid base of information for each object. In the records encoded for this report this basic information was always available in the object with the exception of one object's publication date (when:). The generality of the meaning of the four elements, while versatile, also prove problematic, especially when deciding what information to provide in the four elements of the anchoring story. A good

example of this is contributors to a work. Should all contributors to an object's creation be included in the anchor story's *who* element, or should another story be created to cover the contributors? And if another story is created, what is it called, and what other information can be represented in this story, if any?

These questions demonstrate a further weakness of ERC. The ERC metadata scheme is still new and does not appear to have been developed or used much beyond its conceptualization. ERC's concept of story and the use of multiple stories which can modify the four elements in different ways to share an object's metadata is a powerful concept and empowers ERC with great versatility and extensibility. But these stories as well as element qualifiers remain largely undefined. As yet, basic tools such as a library of allowed or preferred story types and their semantics is not available. This means that often there are several possible ways to encode one or more of an object's metadata. This creates much confusion and difficulty for an encoder trying to choose between them. Even a set of recognized conventions of the best ways to represent certain types of information would be very helpful. Without this kind of standardization, encoding an object's metadata with ERC feels like hedging at best and ad hoc rule making at worst. It is clear that ERC was meant to provide a simple and basic way to describe electronic objects, but limiting an object's description to just its anchor story leaves an encoder feeling that the object is only vaguely described.

The DC metadata scheme on the other hand did not possess this difficulty. DC is an established metadata scheme consisting of 15 well defined, optional elements. DC was created and has grown via dialogue between groups representing various interests. It is also promoted and somewhat controlled by the DC Metadata Initiative (DCMI) which provides a well-documented and regulated forum for the scheme's growth and description. DC's larger element set, well-defined semantics and moderated libraries are in sharp contrast with ERC's vague notions and lack of specificity. Encoding the above five objects with DC was much less ambiguous than with ERC. For each object, DC's 15 elements were more than adequate to describe its salient characteristics. Populating the elements was seldom more difficult than locating the object's bibliographic data and then doing a cut-and-paste into the appropriate DC element fields.

This is not to say that the DC scheme is not without its problems. DC was created by committee and is meant to be versatile enough to be used in many different object environments. This does lead to some lack of specificity in element usage on DC's part as well. This is especially true when describing hybrid objects or objects of mixed media or content, like Web sites (object #3). However, the availability of a multitude of usage examples, a list and description of element qualifiers, and a litany of available control vocabularies mitigates some of this concern. But these features of DC are more evidence of its maturity and the greater design effort behind it than its superiority of design. ERC could also greatly benefit from a similar effort. It should also be noted that while DC worked well in describing the five objects above it is uncertain whether encoding different kinds of objects with DC would be likely as straightforward.

Summary and Conclusion

The Electronic Resource Citation metadata scheme shows a lot of promise. It's simplicity and the self-evident nature of its elements makes it very easy to use. However, more work and some standardization are necessary in order to meet its stated goal of encouraging widespread use of a simple metadata scheme for electronic objects on the Web. Promotion and regulation by a group in the manner of the DCMI for DC would go a long way to providing ERC with the breadth and clarity it needs. At the moment however, using ERC is not so clear cut. Deciding how to represent all of an object's information via ERC metadata is not always easy. More work on semantics and the standardization of qualifiers and story types is necessary to make ERC a really valuable metadata scheme.

ERC was created to serve as a basic and easy-to-use metadata scheme that would hopefully encourage the widespread creation of metadata. However, this simplicity comes at the cost of descriptive power. The absence of a more fully defined set of story types and qualifiers restrains ERC's scope of use. Without clear and precise ways to modify the four basic elements, ERC's descriptive power is limited. Metadata creators are unable to encode an object's more nuanced yet important information elements or describe relationships between those elements. Without a centralized set of such modifiers users will be forced to create their own as they go, essentially negating the efficacy of using a scheme at all. It may be that ERC is simply not meant to provide this kind of detailed description, in which case metadata creators who desire more descriptive power will have to use another scheme.

The Dublin Core metadata scheme on the other hand is clearly a more established and mature system. Its growth and application continues to expand. Its longtime use in various fields and its regulated growth has led to the creation of libraries, vocabularies, and examples which greatly enhance DC's robustness. The exercise of creating the metadata records above demonstrates DC's ability to accurately and fully describe electronic objects of this type. However, it is also a more complicated system, which undoubtedly limits its widespread use among the creators of electronic information entities. The Note Tab Light editor though does relieve the user of some of that burden. The development of additional automated or computer-aided metadata creation tools in the DC scheme could further aid metadata creators in the future.

Any attempt to describe an object using only a subset of its characteristics is bound to fall short. Metadata schemes are no exception. Likewise, a metadata scheme, no matter how clever or elegant, is of little value if it is not used. Usability and descriptive power are not necessarily mutually exclusive of each other. Dublin Core is a relatively successful metadata scheme but its use is somewhat uncertain and complicated. Many additional metadata schemes continue to be created, with varying success. By melding some of DC's more powerful features with newer schemes' simplicity and/or elegance (like ERC), widespread metadata creation may eventually be an achievable goal.