Kelly Robertson
SLIS 5223
Spring 2005

Creating Metadata using Simple Dublin Core and Electronic Resource Citation

Introduction

There are many different metadata schemes available, some are specific to certain fields or types of information objects while others were designed to encompass the entire world of information. This paper looks at two different schemes, simple Dublin Core (DC) and Electronic Resource Citation (ERC). Both can be applied to a wide variety of objects, but are usually used in the digital information world. The loosely defined nature of many fields in these schemes can result in quality issues, but at the same time allow for greater application and interoperability.

Electronic Resource Citation Records

erc:
who: Brand, Amy | Daly, Frank | Meyers, Barbara | Sheridan Press | Niso Press
what: Metadata demystified: a guide for publishers
when: [:] 2003
where: http://www.niso.org/standards/resources/Metadata_Demystified.pdf
erc-about:
who: (:unap)
what: Metadata standards | Metadata practices
when: (:unap)
where: (:unap)
erc-from:
who: Robertson, Kelly
what: Brand_Demystified_ERC
when: [:] 2005 01 29
where: (:unav)

erc:
who: Library of Congress
what: Mets: an overview & tutorial
when: [:] 2004 09 23
where: http://www.loc.gov/standards/mets/METSOverview.v2.html
erc-about:
who: (:unap)
what: Metadata encoding and transmissions standards | Metadata schema
when: (:unap)
where: (:unap)
erc-from:
who: Robertson, Kelly
what: METS_ERC
when: [:] 2005 01 29

where: (:unav)

erc:
who: DCMI-Libraries Working Group | Clayphan, Robina | Guenther, Rebecca
what: DC-Library Application Profile (DC-Lib)
when: [:] 2004 09 10
where: http://dublincore.org/documents/library-application-profile/
erc-about:
who: (:unap)
what: Dublin Core | Application profiles | Metadata schema
when: (:unap)
where: (:unap)
erc-from:
who: Robertson, Kelly
what: DCLIB_ERC
when: [:] 2005 01 29
where: (:unav)

erc:
who: Lynch, Clifford A.
what: Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age
when: 2003 02
where: http://www.arl.org/newsltr/226/ir.html
erc-about:
who: (:unap)
what: Institutional repositories | Scholarly communities | Digital preservation
when: (:unap)
where: (:unap)
erc-from:
who: Robertson, Kelly
what: LynchERC
when: [:] 2005 01 27
where: (:unav)

erc:
who: Powell, Andy | Lyon, Liz
what: JISC Information Environment Architecture | Joint Information Systems Committee Information Environment Architecture
when: [:] 2003 07 03
where: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/
erc-about:
who: (:unap)
what: Resource discovery protocols | Digital publication standards | Portals
when: (:unap)

where: United Kingdom

erc-from:
who: Robertson, Kelly
what: Powell_JISC_ERC
when: [:] 2005 01 27
where: (:unav)

Dublin Core Records

<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.Subject" LANG="en" CONTENT="Metadata standards ; Metadata practices">
<META NAME="DC.Description" LANG="en" CONTENT="Explains what metadata is and is not and is not and the benefits of metadata for the publishing industry and readers. Gives brief explanations of Dublin Core, XML, and SGML.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Sheridan Press">
<META NAME="DC.Publisher" LANG="en" CONTENT="Niso Press">
<META NAME="DC.Contributor" LANG="en" CONTENT="Daly, Frank ">
<META NAME="DC.Contributor" LANG="en" CONTENT="Meyers, Barbara">
<META NAME="DC.Date" LANG="en" CONTENT="2003">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/pdf">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.niso.org/standards/resources/Metadata_Demystified.pdf">
<META NAME="DC.Language" LANG="en" CONTENT="en">

<META NAME="DC.Title" LANG="en" CONTENT="Mets: an overview & tutorial">
<META NAME="DC.Subject" LANG="en" CONTENT="Metadata encoding and transmissions standards ; Metadata schema">
<META NAME="DC.Description" LANG="en" CONTENT="Explains the METS metadata scheme.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Library of Congress">
<META NAME="DC.Date" LANG="en" CONTENT="2004-09-23">
<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/standards/mets/METSOverview.v2.html
">
<META NAME="DC.Language" LANG="en" CONTENT="en">

<META NAME="DC.Title" LANG="en" CONTENT="DC-Library Application Profile (DC-Lib)">
<META NAME="DC.Creator" LANG="en" CONTENT="Clayphan, Robina">
<META NAME="DC.Subject" LANG="en" CONTENT="Dublin Core ; Application profiles ; Metadata schema ">
<META NAME="DC.Description" LANG="en" CONTENT="Outlines standards to develop application profiles by requiring specific Dublic Core elements. Also covers the concept of using multiple metadata schemes.">
<META NAME="DC.Publisher" LANG="en" CONTENT="DCMI-Libraries Working Group">
<META NAME="DC.Contributor" LANG="en" CONTENT="Guenther, Rebecca">
<META NAME="DC.Date" LANG="en" CONTENT="2004-09-10">
<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/library-application-profile/">
<META NAME="DC.Language" LANG="en" CONTENT="en">

<META NAME="DC.Title" LANG="en" CONTENT="Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age">
<META NAME="DC.Creator" LANG="en" CONTENT="Lynch, Clifford A.">
<META NAME="DC.Subject" LANG="en" CONTENT="Institutional repositories ; Scholarly communities ; Digital preservation">
<META NAME="DC.Description" LANG="en" CONTENT="Defines and discusses institutional repositories and their potenial role in academic communities. Potential problems with repositories are also discussed, especially long-term sustainability issues.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Association of Research Libraries">
<META NAME="DC.Date" LANG="en" CONTENT="2003-02">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Format" LANG="en" CONTENT="42kB">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.arl.org/newsltr/226/ir.html">
<META NAME="DC.Language" LANG="en" CONTENT="en">

<META NAME="DC.Title" LANG="en" CONTENT="JISC Information Environment Architecture">
<META NAME="DC.Creator" LANG="en" CONTENT="Powell, Andy">
<META NAME="DC.Subject" LANG="en" CONTENT="Resource discovery protocols ; Digital publication standards ; Portals">

<META NAME="DC.Description" LANG="en" CONTENT="Promotes the advantages of the JISC's information environment architecture. Gives its framework and standards and provides scenarios where it would be useful to apply it.">
<META NAME="DC.Publisher" LANG="en" CONTENT="Joint Information Systems Committee">
<META NAME="DC.Contributor" LANG="en" CONTENT="Lyon, Liz">
<META NAME="DC.Date" LANG="en" CONTENT="2003-07-03">
<META NAME="DC.Type" LANG="en" CONTENT="text">
<META NAME="DC.Type" LANG="en" CONTENT="image">
<META NAME="DC.Format" LANG="en" CONTENT="text/html">
<META NAME="DC.Format" LANG="en" CONTENT="image/gif">
<META NAME="DC.Identifier" LANG="en" CONTENT="http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/">
<META NAME="DC.Language" LANG="en" CONTENT="en">

Discussion

There are many issues related to creating metadata records of any kind. One of the most important universal concerns is the user and how the metadata will be most useful to him. An important question to ask is what kind of information the user is looking for about the document and how to create that information consistently for many information objects. However, the user is not the only concern in the creation of metadata, how the records are stored and retrieved needs to be examined as well as the need to create administrative metadata. Administrative metadata assists the people who work with the back end of the system in maintaining it and gives them knowledge about the creation of the record. Preservation metadata needs to be taken into consideration as well and structural metadata is required for some electronic resources. Another universal issue is whether to perform authority control and on what fields or elements authority control would be most useful. Authority control is related, but independent of metadata creation. While useful, authority control can be time consuming (hence expensive) and thought needs to be given whether its benefits outweigh the costs. Another consideration of using authority control is how to denote when it is being used and which thesauri are being used.

A big question that all metadata schemes need to answer is how narrowly should the fields or elements be defined. What is the largest chunk of information that can still be meaningful to the user and how broadly can it be defined? While narrow definitions can lead to better precision, they also require more training and time on the part of the metadata creator. It can also lead to a scheme being used in only a few disciplines or applied to only a few kinds of information objects. However, how meaningful can loosely defined fields be and what is their impact on searching? DC and ERC both try to be as broad-based as possible, but they allow for the creation of narrower fields. While this can improve searching precision, it can also cause interoperability problems. When fields are qualified, consistency is lost and this can cause misinterpretation of content within the field or the field ignored altogether by another system that does not recognize the

changes. Since most narrowing of fields will only be useful locally it can be questioned whether they are worth the time to define and implement or whether a different scheme that can be applied with more uniformity should be chosen.

There are also issues specific to the metadata scheme used. For example in ERC and simple DC there is concern about the date field, unfortunately neither is very specific as to the exact nature of the date (date created vs. last update vs. copyright date). While on the surface, ambiguity about dates does not seem to be a matter of great concern, it can potentially change what records are retrieved in a search. Many users limit their searches by the date field. If what the date field represents is unclear, users may not retrieve information relevant to them. With both of these schemes, there is a frustration of trying to enter very specific information into broad elements or fields. Interpreting the information in the field can be difficult for the same reason.

However, broadly defined fields can also be a strength, as they can be applied to wide variety of information objects and across different kinds of databases, allowing for greater interoperability. This allows the user to search many different types of information with one search. Their simplicity also means more people can create metadata, allowing large databases and repositories to be developed with greater efficiency. DC and ERC both have very simple input rules and allow a lot of flexibility in formatting the information in the elements. This makes it easier for people with little training to input records but it also has a negative effect on the quality of the records. With the potential for a wide variety of input styles, there is a decrease in uniformity of records, which can lead to lower precision and recall when searching. This can be contrasted to MARC, which requires a high level of training, but provides greater consistency allowing for more accurate search returns.

ERC is elegant in its simplicity, but that same simplicity can be frustrating and confusing. An important part of creating ERC records is determining the most important, or anchoring, story and how many stories are needed to describe the object properly. For most information objects the main story is needed, which contains the essential descriptive metadata (title, author, date, location) and the about story which contains the subject analysis of the object. Depending on the needs of the agency creating the records, from and support stories may be added for administrative metadata.

The lack of places to enter descriptive metadata is frustrating in ERC. For example, there are no good places to enter publisher information or affiliated corporate bodies. Publisher information and corporate bodies could be added into the who field, although the fit is not good as they are not content providers. Another problematic but likely scenario is a record created for a journal article, the title of the journal does not neatly fit into any of the ERC fields, an added entry under title would probably be the best place for such information. The ability to create stories does help solve this problem to some extent, as a story could be created to contain this information. However inventing stories carries its own problems, particularly when it comes to interoperability. While the body that created this additional metadata would be able to use it, it is likely others who harvest it would not.

One of the benefits of ERC is its modularity, it is easy for the metadata creator to choose how much or how little to use. For example if a large quantity of information objects need metadata, a decision is made to tell only the main story to save time. A knowledgeable user can readily determine that is all the metadata created, the other stories intentionally left untold. In a similar vein, each field in ERC is required to contain some information, even if it is only to tell the user the information is unknown, not applicable or intentionally left blank. Other metadata scheme leave the user wondering if fields were deliberately left blank, are irrelevant or if it is simply an example of a poor metadata record.

Simple Dublin Core uses 15 basic elements to describe the object. None of the elements are required so the metadata creator must determine which elements are necessary for the object. These elements cover the basic types of metadata, including descriptive, subject analysis, and administrative. DC is similar to MARC in that no elements are required and both were developed with text-based information in mind, although they can be applied to any type of information object. However, DC is much easier to use than MARC and easier to apply to non-text objects.

Simple Dublin Core is more complex than ERC and thus can provide a richer record in most respects. For example, DC can differentiate between a primary creator or content provider and secondary ones while ERC cannot. DC can better describe what kind of information object it represents than ERC. While ERC only states where the object is located (an URL or call number for a physical object), DC can provide that information and tell what the object is, be it a PDF file, book or painting. When it comes to subject analysis though, the ERC record is a bit more specific. While simple Dublin Core only provides one subject element that all subjects must fit into, the ERC-about story provides fields allowing subject analysis to be broken into who, what, where and when. Making it easier to determine who the item is about or what time period the information is about. On the other hand, DC does have a content element that allows a summary to be added to the record.

Summary and Conclusion

The decision on how loosely fields can be defined often depends on what type of information the metadata will describe and how it will be used. Because of the information explosion, metadata needs to be applied quickly and easily but still be meaningful. While they have their problems, ERC and DC are good metadata schemes in many ways for the digital world.

Overall, ERC and DC can be used to create good metadata records and are easy enough to use that they can be applied over large quantities of documents. Unfortunately, the broad elements that allow them to be applied over many different types of information and make them easy to use causes confusion over what exactly each field or element contains. This means different types and levels of information can be entered into the same field for different objects, leading to more confusion and lowering the

precision of retrieval. While authority control can be used with these (or any) metadata schemes, in this exercise, it was not used and that created frustrations of its own, especially with subject analysis. Applying authority control can be difficult, especially when many different governing bodies are creating records according to their own standards and this can decrease searchability as well.