Metadata

Metadata is data about data, or information that makes locating, using, and storing an object easier. Metadata is often distinguished from traditional cataloging by the requirement that metadata be about electronic data (Greenberg, 2005), but in the broad sense, especially in the library environment, the term is used for any information relating to any object, digital or non-digital (National Information Standards Organization, 2004). The purpose of metadata is to provide a surrogate for the object, which gathers the important information in one place for the ease of the potential user. It describes the attributes and content of the original, and must describe it sufficiently for the user to understand the purpose, use, and location of the original (Milstead and Feldman, 1999).

Metadata is most useful when it provides some consistency in terms and labels, especially if the metadata is meant to be read by a computer (Milstead and Feldman, 1999). The principal reason for metadata is functionality (Greenberg, 2005), and that functionality is reduced if the user has no way of knowing what to expect. To provide this consistency, most metadata systems use a schema or multiple schemas.

A schema, or scheme, is a framework or plan that provides structure (Greenberg, 2005). Metadata schemas are structures, or sets of elements designed to describe a specific kind of data or for a specific purpose (NISO, 2004). Kant reasoned that schemas are based on experience and empirical analysis (quoted in Greenberg, 2005). Given this, metadata schemas are not designed in isolation, but as a reaction to a particular organizational or structural need. Many are formed by committees or initiatives

(Greenberg, 2005). Metadata schemas usually specify and define elements, as well as often giving content and representation rules and specifying allowable content (NISO, 2004). Another way of thinking of schemas is as a structured container to which data can be added, or as a data dictionary (Greenberg, 2005).

Metadata elements are the semantics of the schema (NISO, 2004), or the defined vocabulary that makes up the structure; for example, fields such as "title," "author," and "date." Attributes define the elements and how they're used. They specify things such as whether or not an element is mandatory, how many times it can be used, and the type of data that may be assigned to the element. Attributes provide a dictionary for the metadata elements.

The Metadata Object Description Schema (MODS) is an XML schema designed to carry data from Machine-Readable Cataloging (MARC) 21 records, as well as enable the creation of new records. It can be used to represent metadata for harvesting, or to represent simplified MARC records in XML. It provides a more user-oriented schema with a simpler element set than full MARC records, and uses language-based tags instead of numeric ones. MODS is also intended to complement other metadata formats, including Metadata Encoding and Transmission Standard (METS) and Search/Retrieval via URL (SRU) (Library of Congress, 2006).

The Text Encoding Initiative (TEI) is a project to develop guidelines for encoding electronic texts. The guidelines are not a single schema but an environment for creating

many schemas, which can be expressed using a number of different formal languages, such as SGML or XML. The TEI guidelines enable libraries, museums, publishers, and scholars to represent a variety of literary and linguistic texts for online research, teaching, and preservation. TEI Lite is a simpler subset that is often used in libraries. The guidelines can be used in the creation of new resources or in the interchange of existing ones. (Text Encoding Initiative, 2007)

The Visual Resources Association (VRA) Core is a standard designed for the cultural heritage community to describe visual works as well as the images that document them. The VRA Core is an XML standard, and makes use of several existing controlled vocabularies, such as the Thesaurus for Graphical Materials and the Art and Architecture Thesaurus (Visual Resource Association, 2007).

For each schema I would first determine how well it does what it was designed to do by studying several diverse sample records, both in the display format and in the raw code. The schema should include all information relevant to its purpose without including too much non-relevant information. For example, the TEI guidelines are designed to mark-up electronic text, so I would expect it to have elements to designate information such as paragraphs, chapters, stanzas, headings, authors, publishers, etc., but not detailed information about visual subjects such as artwork. Another consideration would be how easy to understand the elements and attributes are; in other words, how easy the schema is to learn and use. If the schema is designed to be read by humans, then the elements should be clear and distinct.

MODS Example:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<modsCollection xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.loc.gov/mods/v3"
xsi:schemaLocation="http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
 <mods version="3.0">
 <titleInfo>
 <title>FranUlmer.com -- Home Page</title>
 </titleInfo>
 <titleInfo type="alternative"><title>Fran Ulmer, Democratic candidate for
Governor, Alaska, 2002</title>
 </titleInfo>
 <name type="personal">
 <namePart>Ulmer, Fran</namePart>
 </name>
 <genre>Web site</genre>
 <originInfo>
 <dateCaptured point="start" encoding="iso8601">20020702
</dateCaptured>
 <dateCaptured point="end" encoding="iso8601">
20021203</dateCaptured>
 </originInfo>
 <language>
 <languageTerm authority="iso639-2b">eng</languageTerm>
 </language>
 <physicalDescription>
 <internetMediaType>text/html</internetMediaType>
 <internetMediaType>image/jpg</internetMediaType>
 </physicalDescription>
 <abstract>Web site promoting the candidacy of Fran Ulmer, Democratic
candidate for Governor, Alaska, 2002. Includes candidate biography, issue
position statements, campaign contact information, privacy policy and campaign
news press releases. Site features enable visitors to sign up for campaign email
list, volunteer, make campaign contributions and follow links to other internet
locations. </abstract>
 <subject>
 <topic>Elections</topic>
 <geographic>Alaska</geographic>
 </subject>
 <subject>
 <topic>Governors</topic>
 <geographic>Alaska</geographic>
 <topic>Election</topic>
```

```
 </subject>
 <subject>
 <topic>Democratic Party (AK)</topic>
 </subject>
 <relatedItem type="host">
 <titleInfo>
 <title>Election 2002 Web Archive</title>
 </titleInfo>
 <location>
 <url>http://www.loc.gov/minerva/collect/elec2002/</url>
 </location>
 </relatedItem>
 <location>
 <url displayLabel="Active  site (if
available)">http://www.franulmer.com/</url>
 </location>
<location>
<url displayLabel="Archived site">http://wayback-
cgi1.alexa.com/e2002/*/http://www.franulmer.com/</url>
 </location>
 </mods>
 </modsCollection>
```

TEI Example:
```
<pb n='474'/>
<div1 type="chapter" n='38'>

<p>Reader, I married him.  A quiet wedding we had: he and
I,
the parson and clerk, were alone present.  When we got back
from church, I went into the kitchen of the manor-house,
where Mary was cooking the dinner, and John cleaning the
knives, and I said &mdash;</p>

<p><q>Mary, I have been married to Mr Rochester this
morning.</q> The housekeeper and her husband were of that
decent, phlegmatic order of people, to whom one may at any
time safely communicate a remarkable piece of news without
incurring the danger of having one's ears pierced by some
shrill ejaculation and subsequently stunned by a torrent of
wordy wonderment.  Mary did look up, and she did stare at
me; the ladle with which she was basting a pair of chickens
roasting at the fire, did for some three minutes hang
suspended in air, and for the same space of time John's
knives also had rest from the polishing process; but Mary,
bending again over the roast, said only &mdash;</p>
```

VRA Core Example:

```xml
<?xml version="1.0" encoding="iso-8859-1"?>
<vra xmlns="http://www.vraweb.org/vracore4.htm"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.vraweb.org/vracore4.htm
http://gort.ucsd.edu/escowles/vracore4/vra-4.0-restricted.xsd">
  <work id="w_0005" refid="1" source="Ben's imaginary collection">
    <agentSet>
      <display>Anglo-Saxon</display>
      <agent>
        <name type="other">Unknown</name>
        <culture vocab="AAT" refid="300021024">Anglo-Saxon</culture>
      </agent>
    </agentSet>
    <dateSet>
      <display>between 710 and 721</display>
      <notes source="British Library Digital Catalogue of Illuminated Manuscripts"
refid="http://prodigi.bl.uk/illcat/record.asp?MSID=6469&amp;CollID=7&amp;N
Start=60404">The consensus of scholarly opinion (initially established in the
commentary volume to the first facsimile, Codex Lindisfarnensis) favours an
origin at Lindisfarne in 698 (the date of the translation of the relics of Cuthbert),
but Brown 2003 favours production there between 710 and 721 (721 being the
death date of Bishop Eadfrith, perhaps accounting for the fact that the work was
left incomplete in places).</notes>
      <date type="creation">
        <earliestDate>710</earliestDate>
        <latestDate>721</latestDate>
      </date>
      <date type="creation">
        <earliestDate>698</earliestDate>
        <latestDate>698</latestDate>
      </date>
    </dateSet>
    <locationSet>
      <display>British Library, London, Ms. Cotton Nero D IV</display>
      <location type="repository">
        <name type="corporate">British Library</name>
        <name type="geographic" vocab="TGN"
refid="7011781">London(inhabited place)</name>
        <name type="geographic" vocab="TGN" refid="7002445">England
(country)</name>
        <name type="geographic" vocab="TGN" refid="7008591">United Kingdom
(nation)</name>
        <refid type="accession">Cotton Nero D IV</refid>
      </location>
      <location type="creation">
<name type="corporate">Lindisfarne Monastery</name>
```

<name type="geographic" vocab="TGN" refid="7017534">Lindisfarne(inhabited place)
</name>
<name type="geographic" vocab="TGN" refid=" 7009989">Holy Island (peninsula)</name>
        <name type="geographic" vocab="TGN" refid="7002445">England (country)</name>
        <name type="geographic" vocab="TGN" refid="7008591">United Kingdom (nation)</name>
      </location>
   </locationSet>
   <materialSet>
      <display>ink and paint on vellum</display>
      <material type="medium" vocab="AAT" refid="300015012">ink</material><material type="medium" vocab="AAT" refid="300015029">paint</material><material type="support" vocab="AAT" refid="300011852">vellum (parchment)</material>
   </materialSet>
   <measurementsSet>
      <display>340 x 250 mm</display>
      <measurements type="height" unit="mm" extent="overall">340</measurements><measurements type="width" unit="mm" extent="overall">250</measurements>
   </measurementsSet>
    <relationSet>
       <relation type="largerContextFor" relids="w_0006">f.25v: St. Matthew</relation>
    </relationSet>
<rightsSet>
   <display>(c) The British Library</display>
   <rights><text>(c) The British Library</text></rights>
</rightsSet>
   <sourceSet>
      <display>British Library Digital Catalogue of Illuminated Manuscripts</display>
      <source>
         <name type="electronic">British Library Digital Catalogue of Illuminated Manuscripts</name>
         <refid type="URI"> http://prodigi.bl.uk/illcat/record.asp?MSID=6469&CollID=7&NStart=60404</refid>
      </source>
   </sourceSet>
   <stylePeriodSet>
      <stylePeriod vocab="AAT" refid="300021024">Anglo-Saxon</stylePeriod><stylePeriod vocab="AAT" refid="300021025">Hiberno-

```xml
Saxon</stylePeriod><stylePeriod vocab="AAT"
refid="300021018">Insular</stylePeriod>
    </stylePeriodSet>
    <subjectSet>
      <subject>
        <term type="descriptiveTopic" vocab="AAT"
refid="300026466">Evangeliaries</term>
        <term type="descriptiveTopic" vocab="LCSH">Bible. N.T.
Gospels</term>
        <term type="descriptiveTopic" vocab="LCSH">Lindisfarne
Gospels</term>
        <term type="descriptiveTopic" vocab="AAT"
refid="300265483">Illuminated manuscripts</term>
      </subject>
    </subjectSet>
    <titleSet>
      <display>Lindisfarne Gospels</display>
      <title type="repository" pref="true">Lindisfarne Gospels</title>
    </titleSet>
    <worktypeSet>
      <display>manuscript</display>
      <worktype vocab="AAT" refid="300028569">manuscript</worktype>
    </worktypeSet>
</work>
<work id="w_0006" refid="8" source="Ben's imaginary collection">
    <descriptionSet>
      <display>St Matthew is depicted as a scribe and identified by his
accompanying evangelist symbol, the man (in this case shown as a winged,
trumpeting angel, recalling Last Judgement iconography). The figure peeping
from behind the curtain has been variously interpeted as representing Christ and
the integration of Old and New Testaments, God the Father, Moses, an inspiring
muse</display>
      <description source="British Library Digital Catalogue of Illuminated
Manuscripts"
refid="http://prodigi.bl.uk/illcat/ILLUMIN.ASP?Size=mid&amp;IllID=2224&am
p;MSID=6469">St Matthew is depicted as a scribe and identified by his
accompanying evangelist symbol, the man (in this case shown as a winged,
trumpeting angel, recalling Last Judgement iconography). The figure peeping
from behind the curtain has been variously interpeted as representing Christ and
the integration of Old and New Testaments, God the Father, Moses, an inspiring
muse.  </description>
    </descriptionSet>
    <relationSet>
      <relation type="partOf" relids="w_0005">Lindisfarne Gospels</relation>
      <relation type="imageIs" relids="i_0007">full view</relation>
    </relationSet>
```

```xml
<rightsSet>
   <display>(c) The British Library</display>
   <rights><text>(c) The British Library</text></rights>
</rightsSet>
   <sourceSet>
      <display>British Library Digital Catalogue of Illuminated
Manuscripts</display>
      <source>
         <name type="electronic">British Library Digital Catalogue of
Illuminated Manuscripts</name>
         <refid
type="URI">http://prodigi.bl.uk/illcat/ILLUMIN.ASP?Size=mid&amp;IllID=222
4&amp;MSID=6469</refid>
      </source>
   </sourceSet>
   <subjectSet>
      <display>St. Matthew; Evangelists</display>
      <subject>
       <term type="iconographicTopic" vocab="ICONCLASS"
refid="11H(MATTHEW)">St. Matthew</term>
       <term type="iconographicTopic" vocab="ICONCLASS"
refid="11I4">Evangelists</term>
      </subject>
   </subjectSet>
   <titleSet>
      <display>f.25v: St. Matthew</display>
      <title type="repository" pref="true" refid="folio 25 verso">St.
Matthew</title>
   </titleSet>
   <worktypeSet>
      <display>manuscript page</display>
      <worktype>manuscript page</worktype>
   </worktypeSet>
</work>
<image id="i_0007" refid="10" source="Ben's imaginary collection">
   <relationSet>
      <relation type="imageOf" relids="w_0006">Lindisfarne Gospels, f.25v: St.
Matthew</relation>
   </relationSet>
<rightsSet>
   <display>(c) The British Library</display>
   <rights><text>(c) The British Library</text></rights>
</rightsSet>
   <sourceSet>
      <display>British Library Digital Catalogue of Illuminated
Manuscripts</display>
```

```xml
        <source>
            <name type="electronic">British Library Digital Catalogue of
Illuminated Manuscripts</name>
            <refid type="URI">
http://prodigi.bl.uk/Illimages/BLCD/mid/C166/C1663-03.jpg</refid>
        </source>
    </sourceSet>
    <titleSet>
        <display>full view</display>
        <title type="generalView">full view</title>
    </titleSet>
    <worktypeSet>
        <display>digital image</display>
        <worktype vocab="AAT" refid="300215302">digital image</worktype>
    </worktypeSet>
</image>
</vra>
```

References:

Greenberg, J. (2005). Understanding metadata and metadata schemes. *Metadata: A Cataloger's Primer*. 17-36.

Library of Congress. (July 7, 2006). *Metadata Object Description Schema official Web site*. Retrieved October 2, 2007, from http://www.loc.gov/standards/mods/.

Milstead, J. & Feldman, S. (1999). Metadata: cataloging by any other name. Online, v23(1), 24-30.

Text Encoding Initiative. (2007). Retrieved October 3, 2007, from http://www.tei-c.org/.

Understanding Metadata. (2004). *NISO Press*.

Visual Resources Association. (April 9, 2007). *VRA Core 4.0*. Retrieved October 3, 2007, from http://www.vraweb.org/.

Metadata Standards

Metadata standards provide rules for what content is allowed and how content should be included (Chan and Zeng, 2006). Standards evolve as a response to real or anticipated problems, although technology standards are often developed to anticipate product development (Moen, 2000). Standards are intended to reduce uncertainty and lead to more predictable results. They provide structure and guidelines for data entry and organization. Standards are necessary for exchange of information between computer systems, to ensure that data is read and interpreted the correct way (Moen, 2000).

Unlike most industry standards, which change slowly over a long period of time, metadata standards are in an almost constant state of flux as technology evolves (Moen, 2000). However, the ever evolving state of technology is one of the main reasons standards are needed. Without them collections can quickly fall into a chaos of data. Standards also provide a structure on which data can be transferred from one technology to another as software and hardware is updated.

Interoperability is the ability of different systems to work together and exchange information in a useful and meaningful way (Moen, 2000). Because different schemas are developed for different problems and to serve different audiences, it would be impractical to try to adopt a single standard across all platforms and situations (NISO, 2004). This solution is more feasible on a smaller scale, such as the widespread use of MARC format in libraries, but is not ideal for different collections containing different types of resources, and where different specialized schemas are already in use (Chan and

Zeng, 2006). However, different schemas can often be mapped using crosswalks and other methods, which can help bridge the gaps caused by using different standards.

Despite the problems interoperability poses, it is important for the same reason that metadata is important; information is not useful if it cannot be found. Interoperability increases the usefulness of information by increasing the ways it can be found and viewed, and makes it possible to exchange and share data that has been prepared to different schemas (Chan and Zeng, 2006). It is a key concept for networked systems; for example, multiple databases on different information servers can be searched simultaneously if they use the same standard, even if the information is in different data formats (Moen, 2000). Even within a single community it may not be possible to adopt a single, uniform standard, so it becomes important to find other ways of achieving interoperability (Chan and Zeng, 2006).

The most complete interoperability involves open systems, which require information systems to be completely open to other systems on the same network; however, this causes problems with proprietary software and privileged information. Standards provide a way to share data that does not require that the information systems be open. For example, the Web environment uses the standards of hypertext transfer protocol (HTTP) and hypertext markup language (HTML), which can be read by any Web browser (Moen, 2000).

Problems can arise when the standards change or new standards start being used, and existing readers (in this case, Web browsers) haven't caught up. Changes in technology

standards may be inevitable, but problems occur when one technology replaces another, and interoperability is limited if the standard currently in use is not taken into account, or the new standard is not backward compatible (Moen, 2000).

Interoperability can take place at different levels. At the schema level, the focus is on the elements of the schema. Sometimes new schemas are developed from existing ones, different guidelines or policies are established for specific needs, or one schema is mapped to another (Chan and Zeng, 2006).

However, the benefits of interoperability must be weighted against economic and organizational concerns (Moen, 2000). Database owners may not want their information to be freely available for anyone to search. Some systems will not allow the use of third-party software. In this case, the use of standards to display the information could make it searchable without having to integrate another software, and user authentication mechanisms can be used to control access (Moen, 2000).

References:

Chan, L. & Zeng, M. (2006). Metadata interoperability and standardization - a study of

methodology part I. *D-Lib Magazine*. 12 (6). Retrieved September 21, 2007, from

http://www.dlib.org/dlib/june06/chan/06chan.html.

Greenberg, J. (2005). Understanding metadata and metadata schemes. *Metadata: A*

Cataloger's Primer. 17-36.

Milstead, J. & Feldman, S. (1999). Metadata: cataloging by any other name. Online,

v23(1), 24-30.

Moen, W. (2000). Interoperability for information access: technical standards and policy

considerations. *Journal of Academic Librarianship*, 26 (2), 129-132.

Understanding Metadata. (2004). *NISO Press*.

Metadata Elements and Standards

Attributes describe metadata elements by providing guidelines and definitions, such as the acceptable types of data, whether or not the element is mandatory, how many times it can be used. Attributes can also provide administrative information such as the version or history of the element, the language, and status of the element.

According to the PBCore user guide, PBCore is based on Dublin Core, but it's clear the attributes have been specialized for a different purpose or audience. The PBCore metadata dictionary describes the attributes in more detail than the Dublin Core, but both lists have the same number of attributes. The Dublin Core specifies which attributes are mandatory and which are optional, while the PBCore makes no such specification. Both lists include the attributes "Name" and "Definition." The Dublin Core element "Label" appears to be similar to the PBCore attribute "Element Label," but is defined as the "human-readable label assigned to the term," while PBCore's "Element Label" is an administrative attribute that is "used to indicate the exact manner in which an element is referenced." The two attributes have similar titles, but appear to be different in practice.

Dublin Core has more attributes that distinguish the term and the term's place in the hierarchy, such as "Broader Than" and "Narrower Than," as well as an attribute for additional information not included in the description, and one for a link to authoritative documentation. In PBCore these are covered by the single attribute "Element Interdependencies." PBCore has more elements that seem to be for technical information, instead of descriptive information, including five administrative attributes,

which have technical information; they are "Element Label," "Language of the Element," "Namespace Identifier," "Element Version," and Registration Authorities," all of which, the descriptions state, can be ignored by a cataloger. Dublin Core does not designate any of their attributes as administrative.

Common attributes:

Name

The name or unique token of the element.

Definition

A brief statement or description of the element.

PBCore:

Element Interdependencies

A description of the relationship of this element to other elements.

Refinements & Encoding Scheme

A way to control the data entered for certain elements, using authority lists, controlled vocabularies, or syntax rules.

Guidelines for Usage

A brief guide for the intended meanings and use of the elements.

Obligation to Use

Whether an element is mandatory, mandatory if applicable, optional, or recommended.

Repeatable Element

How many times an element may be used within a document.

Type of Data Entry

What type of data is permitted, such as text, numerical, date, time, etc.

Examples

Real world examples to illustrate how an element is to be used.

Element Label

An administrative attribute that indicates the exact manner in which the element is referenced. Usually the same as the Name attribute.

Element Version

An administrative attribute that indicates which version of an element is being used.

Namespace Identifier

A unique name that identifies the XML namespace being used with the schema. Administrative attribute.

Registration Authority

An administrative attribute that identifies the metadata registry associated with the schema.

Language of the Element

An administrative attribute that specifies the language used to define the element.

Dublin Core:

URI

The uniform resource identifier used to identify an element.

Label

The human-readable name of an element.

Type of Term

The type of an element, as described in the DCMI Grammatical Principals.

Status

The status of an element, as described in the DCMI Usage Board Process.

Date Issued

The date the element was first declared.

Comment

Any additional information about the element.

See

A link to an authoritative document about the element.

References

A citation for resources referenced in the Definition or Comment.

Qualifies

A reference to an element qualified by the encoding schema.

Broader Than

A reference to a more specific element in the vocabulary.