Metadata Workshop
3 March 2006
Part 1

Metadata overview and guidelines

Amelia Breytenbach
Ria Groenewald

Overview

- What metadata is
- Types of metadata and their importance
- How metadata is stored, what metadata does and why we use it
- Metadata standards and schemas
- Dublin Core Metadata Standard
- Encoding
- General guidelines for writing metadata
- UPSpace and metadata
- Metadata Standard for UP Working document

Definition of metadata

Data about data. Metadata describes how and when and by whom a particular set of data was collected, and how the data is formatted. Metadata is essential for understanding information stored in data warehouses and has become increasingly important in XML-based Web applications.

Source: Webopedia http://www.webopedia.com/TERM/M/metadata.html

What is metadata?

- Standardised descriptions of resources that aid in the discovery and retrieval of resources, particularly in reference to information about electronic, or digital, material
- Describing individual files, single objects or complete collections
- Traditional library cataloging is a form of metadata and MARC 21 and the AACR2 used with it are metadata standards

Types of metadata

- **descriptive**
  - *title, author, extent, subject, keywords*
- **structural**
  - *unique identifiers, page numbers, special features (table of contents, indexes)*
- **administrative or technical**
  - *file formats, scanning dates, file compression format, image resolution*
  - **preservation**
    - *Archival information*
  - **rights management**
    - *Ownership, copyright, license information*

Importance of metadata types

- These metadata types are critical to establish an accurate understanding of the nature of the resource:

- Content (What the object contains or is about)
- Context (The who, what, why, where and how aspects associated with the object's creation)
- Structure (The formal set of associations within or among individual objects)

But how is metadata stored?

- Stored separately as a HTML, XML or MARC 21 document linked to the resource it describes
- As an integral part of the record by embedding the metadata in the Web pages
- In a database linked to the resource

What does metadata do?

Metadata:

- is the key to ensuring that resources will survive and continue to be accessible into the future
- is searchable and aids the identification and retrieval of resources
- helps the end user to do accurate searching and retrieval and will also help to evaluate a resource
- types also assists in managing, maintaining and preserving digital collections
- facilitate interoperability
- supports archiving, security and authentication of digital resources

Why use metadata?

- Metadata provides the essential link between the information creator and the information user
- We can ensure that this objective is met by using metadata in **accordance with international standards**

Metadata standards

Value of metadata standards

- Standards establish a uniform set of ground rules for tagging information.
- Standards help ensure consistency in metadata application.
- Standards support interoperability of applications and resource sharing.
- Standards pave the way for new technologies.

International metadata standards

- Standards are created by international or internationally recognised bodies such as IFLA, ISO, W3C, NISO.
- This is often a long, expensive and labour intensive process

  with expansion, adaptation and extension according to dynamic needs.

Metadata schemas 1

- Application of metadata is controlled by use of schemas consisting of defined fields for specific types of information
- Metadata elements are the individual components that make up a schema
- Each element will contain a particular category of information depending on the definition of that element
- Elements can vary and not all schemas contain the same elements as the needs of different communities vary

Metadata schemas 2

- **Dublin Core (DC)**
  - *for the description of simple textual or image resources*
- **Encoded Archival Description (EAD)**
  - *for encoding archival finding aids e,g. description of collections as a whole.*
- **Text Encoding Initiative (TEI)**
  - *for text markup such as novels, poetry, etc*
- **Visual Resources Association (VRA)**
  - *for the description of art, architecture, and other visual resources.*

Dublin Core Metadata Element Set

- Elements
  - are simple to understand and apply
  - optional and repeatable
  - international and cross-disciplinary in scope
- Unqualified
  - For coarse-grained discovery of resources
- Qualified
  - For richer descriptions to enable more refined resource discovery

Dublin Core Metadata Element Set

Content & about the resource	**Intellectual property**	**Physical manifestation**
Title	Author or Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Language		
Relation		
Coverage		

What is encoding?

Encoding allows the metadata to be processed by a computer program:

- This is done by the addition of markup to a document to store and transmit information about its structure, content or appearance.
- presentational markup relates to the visual appearance of a document e.g. fonts
- descriptive markup relates to the logical structure of a document e.g. to indicate a title, author etc
- according to international standards e.g. ISO, W3C

Encoding Schemes

Important schemes include:

- HTML (Hyper-Text Markup Language)
- XML (eXtensible Markup Language)
- RDF (Resource Description Framework)
- MARC (Machine Readable Cataloguing)
- SGML (Standard Generalised Markup Language)

Ways to write metadata

- **Natural** metadata is found in the source document
  - supports discovery of resources
  - includes the author's name, date, title

- **Added** metadata is added by an metadata editor or by software
  - supports resource selection
  - includes subject terms, abstracts, rights metadata

Metadata requirements for collections

- Importance to balance metadata requirements against real cost of creating and managing the information
- Using a complex, multi-faceted metadata schema is more expensive than a simple schema
- Assigning terms from controlled vocabularies is more costly and time-consuming than using random keywords
- Level of metadata complexity used will correlate with the needs of the anticipated collection users

Controlling your language

- Metadata vocabularies are the “words” or “values” you enter into the subject element

- The Subject element can be assigned
  - Using an existing controlled vocabulary as it is
  - Adapting or customising a vocabulary
  - Developing your own vocabulary
  - Using free language

Choosing a vocabulary

- Please bear in mind when choosing a vocabulary:
  - The users
  - The nature and extent of the collection
  - The skills and available time of the submitters
  - Your community

Please take note of our list of thesauri, subject headings and word lists on your handout before deciding on a vocabulary

UPSpace and metadata

- **Qualified Dublin Core metadata standard**

➢ DSpace uses the Qualified Dublin Core metadata standard for describing textual or image resources

➢ But what does “Qualified” means?

Qualified Dublin Core employs additional qualifiers to the basic 15 elements to further refine the meaning of an element. Qualifiers allow applications to increase the precision of the metadata

Mandatory elements in UPSpace

- **Title:** A name given to the item/resource
- **Author:** This is the element “creator” and is the entity primarly responsible for making the content of the resource
- **Type:** The nature or genre of the content of the resource e.g. an image.
- **Subject:** Topic of the content of the item/resource. This can be free language keywords.
- **Language:** Language of the item/resource. This can be selected from a dropdown menu on the submission form.
- **Rights:** Information about rights held in and over the resource, e.g. the name of the copyright holder

Metadata and workflow in UPSpace

- Submitter
  - Submit baseline metadata
  - Can edit metadata for own submission
- Reviewer
  - Cannot edit metadata or change files
- Approvers
  - Can edit the submission metadata to fix obvious errors
- Metadata editor
  - Check and/or extend the submission’s metadata

Metadata Standard for UP Working Document

- Available at the URL:

  http://www.dspace.up.ac.za/metadata.doc

The document gives you:

- Guidelines to the DC Metadata fields using DSpace
- Explanation of DC elements and qualifiers
- Metadata Do's and Don't's
- Corresponding MARC21 fields
- Standards used for specific DC elements and qualifiers

Working document explanation (1)

- The DC element subject with the standards which will be used in UPSpace.
- DC element Date will be system generated by UPSpace according to the ISO 8601 standard
- DC element Type and Language standards are accommodated in dropdown menus on submission form in UPSpace
- DC element Format is accommodated in the Bitstream Format registry of UPSpace

Working document explanation (2)

- The next tabel explains the 15 DC elements and qualifiers
- In the description column you can see an interpretation of the meaning of each element/ qualifier, useful cataloguing notes and also the MARC fields
- In the examples column we try to illustrate what content to add to a specific element or qualifier field
- On page 16 please take note of the metadata Do’s and Don’ts

References

- Bass, Michael J. (2002) DSpace – a sustainable solution for institutional digital asset services (Functionality).
- Taylor, Chris. (2003) An Introduction to metadata http://www.library.uq.edu.au/iad/ctmeta4.html
- Technical Advisory Service for Images (TASI). Metadata and digital images. http://www.tasi.ac.uk/advice/delivering/metadata.html
- Technical Advisory Service for Images (TASI). Controlling your language – links to metadata vocabularies http://www.tasi.ac.uk/resources/vocabs.html
- Hodge, Gail. (2001) Metadata made simpler.
- Smith, MacKenzie. (2003) Dspace: an open source dynamic digital repository. D-Lib Magazine, January 2003.
- http://www.dlib.org/dlib/january03/smith/1smith.html

Exercise 1 and 2

Exercise 1. Answers

1) Are all of the six mandatory Dublin Core fields in UPSpace listed? Specify the elements lacking, if any.

No
Element Type and element Rights

2) What scheme has been used for DC subject headings?

LCSH (Library of Congress Subject Headings)

3) What agency is noted as dc.creator? Do you think the hierarchy for the name authority is correct?

Government of Canada. Population and Public Health Branch
Yes

4) What if any element refinement (s) / qualifiers are used for the date element?

Created, modified and accessioned

5) Are any elements repeated? Specify.

Yes
Element Subject

6) Are any other metadata fields (in addition to the six mandatory DC elements) included? Specify.

Yes
Description
Date

Exercise 2. Answers

Title	D
Creator/Author	D
Subject	D
Description	D
Publisher	R
Date	A
Type	A
Format	A
Identifier	S
Language	D