Metadata, Metadata Schemas & Metadata Standards

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Overview

- What metadata is
- How metadata is stored (encoding), what metadata does and why we use it
- Types of metadata and their importance
- Metadata standards and schemas
- Dublin Core Metadata Standard
- General guidelines for writing metadata

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

What is metadata?

"The digital catalogue card"

- Standardised descriptions of resources that aid in the discovery and retrieval of information resources in electronic or digital format

- Describing individual files, single objects or complete collections

- Traditional library cataloguing is a form of metadata and MARC 21 and the AACR2 used with it are metadata standards

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

What does metadata do?

Metadata:

- is the key to ensuring that resources will survive and continue to be accessible in 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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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**

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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*

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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 e.g subject headings)

➢ Context (The who, what, why, where and   how aspects associated with the object's creation e.g authors)

➢ Structure (hierarchical and organisational design of the data and the parts thereof e.g. chapters, articles)

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Encoding schemes

Important schemas include:

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Metadata standards/schemas

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.

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Metadata schemas 2

- **Dublin Core Metadata Schema (DC)**

  A simple generic element set applicable to a variety of digital object types. Use for the description of simple textual or image resources.

  http://dublincore.org/

- **Encoded Archival Description (EAD)**

  A set of rules for the encoding of the intellectual and physical parts of archival finding aids e.g. description of collections as a whole.

  http://www.loc.gov/ead/

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Metadata schemas 2 (cont.)

- **Text Encoding Initiative (TEI)**

  To assigned text markup for literary and linguistic texts such as novels and poetry, etc.

  http://www.tei-c.org/

- **VRA Core Categories 4.0**

  A scheme developed by the Visual Resources Association for the description of art, architecture, artifacts, and other visual resources.

  http://www.vraweb.org/projects/vracore4/index.html

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Metadata schemas 2 (cont.)

- **MODS (Metadata Object Description Schema)**

  An XML schema for descriptive metadata compatible with the MARC 21 bibliographic format.

  http://www.loc.gov/standards/mods/

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Dublin Core Metadata Element Set

- **15 Elements**
  - are simple to understand and apply
  - optional and repeatable
  - international and cross-disciplinary in scope

- **Unqualified Dublin Core**
  - For coarse-grained discovery of resources

- **Qualified Dublin Core**
  - For richer descriptions to enable more refined resource discovery

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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		

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Qualified Dublin Core metadata standard

- **What does "Qualified" means?**

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

- **Classes of qualifiers**

  -Element refinement

  -Encoding scheme

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Ways to write metadata

- **Natural** metadata is found in the source document and created by the researcher or submitter

  - 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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Which metadata schema to choose

- 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
- Plan for future uses of the metadata you create
- Focus on content and value standards as much as possible

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Consistency in assigning metadata

- Metadata registry

  - Provides standardised information for the definition, identification, and use of each data element.

  - ensure that a metadata schema and data elements in use by an organization can be applied consistently within the organization or community, reused by other communities, and interpreted by computer applications and human users.

- Personal names and controlled vocabulary

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Personal names

- Authority control using OCLC's Name Authority File
- Format names consistently within a collection, according to authority files or standards
- Use of composite identifiers that combine author name, publication date, and author affiliation
- Asking authors to input the variants of their name at the time of depositing articles

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Putting it all into practice

- Planning is as important as the implementation

- Written documentation on do's and don't are essential

- Develop a quality control workflow for metadata creation

- Share your findings, joys and sorrows with others

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Closing remarks

"Metadata" means many different things:

- It involves applying traditional library principles to new environments

- Good metadata practitioners use fundamental cataloging principles in non-MARC environments

- There is *always* more to learn

- Good metadata promotes good digital collections

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

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

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

References (cont.)

- Disa Workshop: Digital collections management, University of KwaZulu-Natal, 2004.

- NISO Framework Advisory Group. A Framework of Guidance for Building Good Digital Collections. 2nd ed. Bethesda, MD: National Information Standards Organization, 2004. http://www.niso.org/framework/framework2.html

- Xia, Jingfeng. Personal name identification in the practice of digital repositories. Electronic library and information systems. Vol. 40, no. 3 2006. pp. 256-267 .

Ina Smith
Amelia Breytenbach
Ria Groenewald

IGBIS Seminar
"Digital Library Standards & Metadata – the basics"
June 20th, 2007

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA