Information Management Resource Kit

Module on Management of Electronic Documents

UNIT 3. METADATA STANDARDS AND SUBJECT INDEXING

LESSON 1. GENERAL OVERVIEW OF METADATA STANDARDS

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.

Objectives

After this lesson you will:

- understand **what metadata are**;
- understand the different **purposes of metadata**;and
- be familiar with the **areas of application** of different metadata standards.

Introduction

Imagine that your community's library is the largest and most extensive in the world.

Imagine that its collections include books on every area of study known to humankind.

Now, imagine that the only way to find any one book in the collection is **by looking at every other book** on the shelves.

Imagine that there is no index or catalogue with information about the title, author, or subject to enable you to find a specific book…

What are metadata

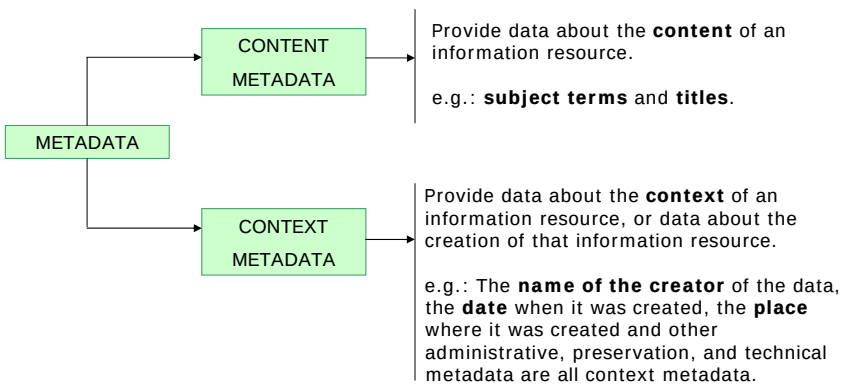
This massive library with its enormous collection of books, all without any identifying information, is rather like the **Internet**.
The Internet is a huge **electronic collection of information**.
Although search engines can find information by searching the content of all these resources, it is not a very efficient way of finding information…

…that's where metadata come in!

Metadata are **structured data** that provide a **short summary about any information resource**, print or electronic, and facilitate the **location, identification, or discovery** of that resource.

What are metadata

Metadata can provide data about the content or the context of an information resource.



What are metadata

The following form presents sample metadata regarding a FAO publication.

Can you identify which of the following items are content or context metadata?

		Content Metadata	Context Metadata
Title	Global Forest Resources Assessment 2000. Main report	☐	☐
Year	2002	☐	☐
ISBN	9251046425		
Author	Robert Smith	☐	☐
Number in series	140		
Language	En		
Number of pages	512		

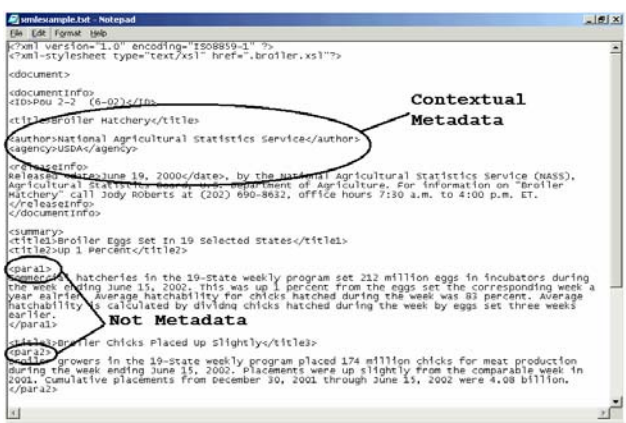
Select the correct answer for each item.

When you have finished, click on the confirm button.

What metadata are not

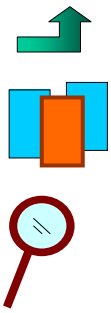
The example provided shows the html code for a web page. As you can note, the tags that are used to separate paragraphs are not metadata.

Metadata are **not document markup**: they are not used to format a document for display or exchange.



Purposes of metadata

Metadata have a number of different purposes:



Location. Metadata can indicate where an information resource is located, either physically or virtually.

Identification. Metadata can distinguish one information resource from another without describing the entire collection of information resources.

Resource discovery. Metadata can link a user's queries about a particular subject with those information resources about the same subject.

Let's now look at the details…

Purposes of metadata

The first purpose for metadata is:

Location. Metadata can indicate where an information resource is located, either physically or virtually.

For example:

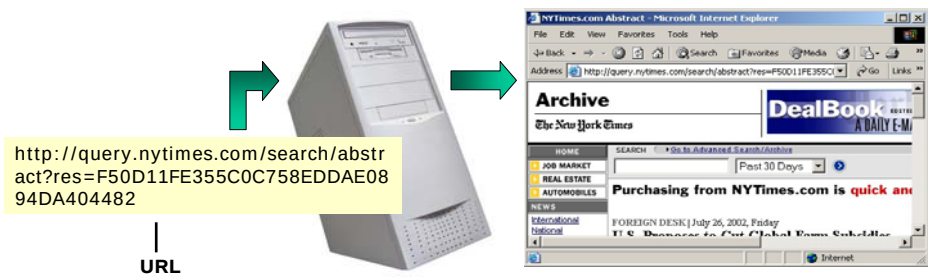
• library holdings; and
• classification information, i.e. Dewey Decimal Classification.

Library holdings and classification information are used in a bibliographic record to point the user to where in a library or library system the information resource is actually located.

Purposes of metadata

Another **example** of metadata being used to **locate data** is the uniform resource locator (**URL**) that distinguishes one World Wide Web resource from another.

Although many URLs look as if they are identifying, or naming, web resources, they are actually the name of a file on a server, or a server itself where the web resource is located.



Purposes of metadata

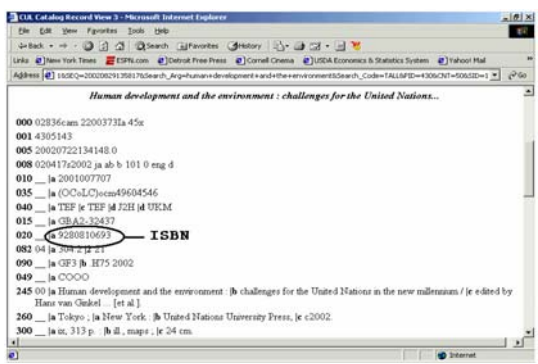
Another metadata purpose is:

Identification. Metadata can distinguish one information resource from another without describing the entire collection of information resources.

For example:

The International Standard Book Number (**ISBN**).

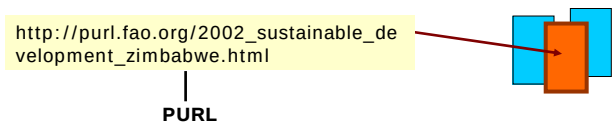
The ISBN is an internationally recognized machine-readable ten digit number used to distinguish one monographic publication from another.



Purposes of metadata

Another **example** of metadata being used to **identify data** is the persistent uniform resource locator (**PURL**).

A PURL identifies a resource **no matter where it is located** on the Web or no matter where that resource is moved on the Web.



For example, this PURL could always identify a FAO document written in 2002 on sustainable development in Zimbabwe, no matter where on the FAO web site it was located. It could reside, hypothetically speaking, on: http://www.fao.org/documents/2002/zimbabwe/sustainable_development.html or it could reside on http://www.fao.org/documents/2002/africa/zimbabwe/sustainable_development.html.

Purposes of metadata

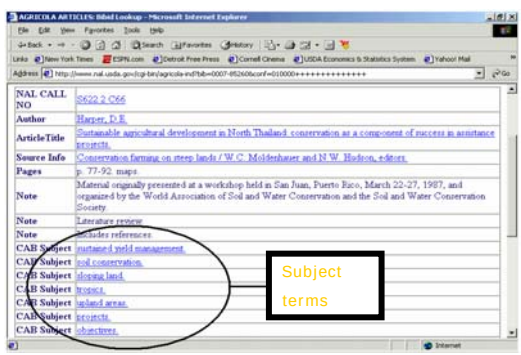
Lastly, metadata can also be used for:

Resource discovery. Metadata can link a user's queries about a particular subject with those information resources about the same subject.

For example:

Subject terms, which are sometimes referred to as **subject headings**, are taken from a controlled vocabulary, such as FAO's AGROVOC thesaurus.

These vocabularies are established by governing organizations and set out to carefully define a set of terms that can be consistently used to describe a large set of resources.

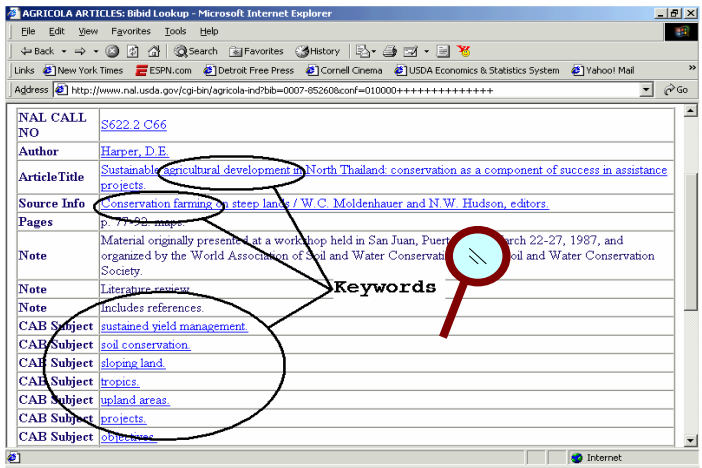


Purposes of metadata

Another **example** of metadata being used for **resource discovery** are **keywords**.

A search on a metadata record using keywords is performed on the *entire* metadata record, not just the sections that have subject terms.

This means that all words in the record are searched.



Purposes of metadata

Now it's your turn to try to identify the different purposes of metadata.

What is the purpose of the following metadata?

http://fao.org/forestry/index.jsp

- ○ Location
- ○ Identification
- ○ Resource discovery

Click on the answer of your choice

Purposes of metadata

Let's now have a look at a form containing metadata for a FAO publication.

Can you identify the purpose of the highlighted metadata?

		Location	Identification	Resource Discovery
Title	Global Forest Resources Assessment 2000. Main report	☐	☐	☐
Year	2002			
ISBN	9251046425	☐	☐	☐
Job Number	Y1997/E			
Call Number	GF3 H75 2002	☐	☐	☐
Language	En			
Number of pages	512			

Select the correct answer for each item.

When you have finished, click on the confirm button.

Metadata models

Now I know what I can do with metadata! There is only one thing that is not clear to me: sometimes metadata are part of the document and sometimes they are not…

There are two different **metadata models**:

• the catalogue/separated metadata model; and

• the embedded metadata model.

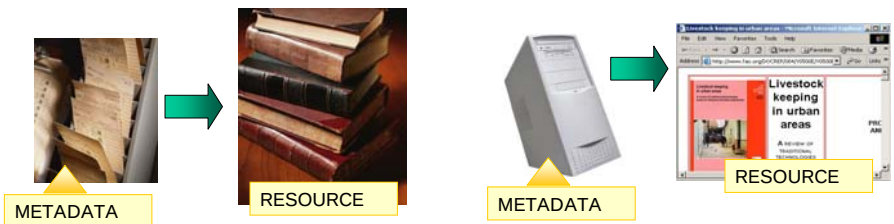
Now let's see what the difference is…

Metadata models

THE CATALOGUE / SEPARATED METADATA MODEL

Metadata for both print and digital resources, but more typically **print** resources, can be kept **separate from the resource** which they describes.

The metadata are kept in a **catalogue or database** and connected to the resource via some type of locating metadata such as a library holding or URL.

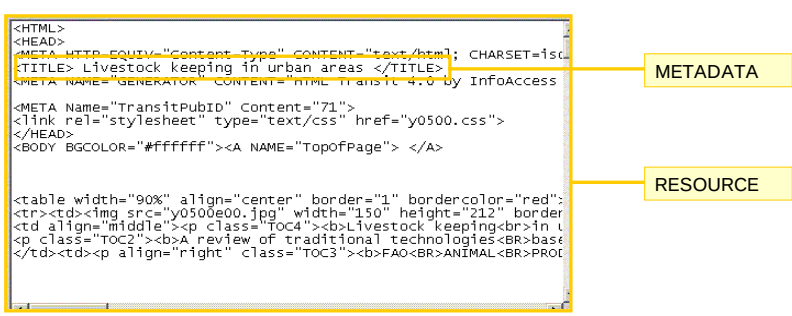


The metadata describe the information resource but are not actually part of it.

Metadata models

THE EMBEDDED METADATA MODEL

Digital information **resources** can **include the metadata** as part of the information they consist of.
For example, a web site will have descriptive meta-tags describing the creator and subject content of the site. These tags are part of the same file as the actual content of the web site.



Metadata applications

Metadata can be used for different purposes. One of the first and best known applications of metadata is in the description of **bibliographic** entities such as books.

As information resources began to be made available digitally, it soon became clear that complex metadata standards were inefficient for dealing with the explosion of **digital information**, so less complex metadata standards started to be developed.

The number of subjects represented by digital information greatly increased, and metadata standards specific to particular subjects were developed, such as those for **geospatial** data and **educational** resources.

Let's now have a look at some common metadata standards to help you understand which one you best suits your needs...

Metadata for bibliographic resources

The International Standard for Bibliographic Description (**ISBD**) is an internationally recognized standard that is used to provide metadata descriptions for **bibliographic resources**. It was developed in order to standardize a number of **rule sets** for bibliographic descriptions from throughout the world. It is used to describe monographic resources, rare books, serials, cartographic materials, printed music, electronic resources, etc..
ISBD provides guidelines for **where** on a bibliographic resource **a bibliographic metadata record should take its specific elements from**, i.e., the title of the resource, the individual or organization responsible for its creation, the place where it was created, the physical description, etc.

049 __ |**a** CGOO
245 00 |**a** Human development and the environment : |**b** challenges for the United Nations in the new millennium / |**c** edited by Hans van Ginkel ... [et al.].
260 __ |**a** Tokyo ; |**a** New York : |**b** United Nations University Press, |**c** c2002.
300 __ |**a** ix, 313 p. : |**b** ill., maps ; |**c** 24 cm.
440 _0 |**a** UNU millennium series
500 __ |**a** "UNUP-1069"--T.p. verso.

ISBD also prescribes the correct format for these pieces of information, i.e., how they should be capitalized, punctuated, and abbreviated in the metadata record.

Metadata for bibliographic resources

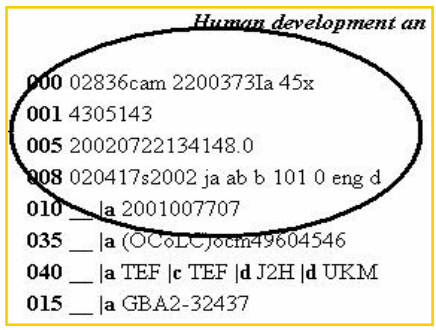
ISBD was originally used in creating the look of paper library catalogues.

As library catalogues began to be made **available electronically**, ISBD continued to be used as the prescription for the description of electronic resources.

But a need arose for a scheme that would **allow databases** of records of bibliographic descriptions **to communicate with one another**.

MARC 21 was developed to do this; it takes the rules prescribed in ISBD and formats them in such a way that machines can communicate them. It provides **fields for storing** a bibliographic resource's title, the individual or organization responsible, etc.

In the example on the right, you can see a sample of non-bibliographic related MARC 21 data. These data allow different bibliographic systems to transfer this record between one another.



Metadata for geospatial resources

In recent years, MARC21 and ISBD have been adapted for use in describing a variety of digital resources. However, since they were originally designed for describing bibliographic resources they are not fully adaptable for describing other kinds of resources, such as **geospatial resources**. Instead of being solely bibliographic in nature with a title, author, or subject terms, geospatial resources are primarily related to **geography** and **location**.

Therefore, a variety of specific metadata standards were developed for geospatial resources. In order to harmonize these various standards, the **International Organisation for Standardisation (ISO)** is developing a suite of standards for geographic information known as ISO 19115.

Metadata for geospatial resources

An example of geospatial metadata is the **Content Standard for Digital Geospatial Metadata** established by the Federal Geographic Data Committee (**FGDC**), in 1998.

If we look at the FGDC standard, we can see that some elements are similar to those used for bibliographic resources, while others are specific to geographic information.

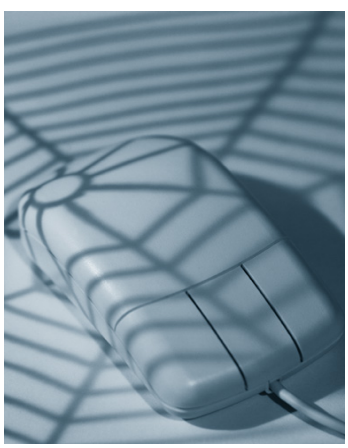
In the example below, two pieces of the same FGDC metadata record are shown. Which one describes geospatial information?

○ *Identification_Information:*
Citation:
Citation_Information:
Originator: U.S. Department of Commerce.
Publication_Date: 1998
Title: Roads, New York County
Publication_Information:
Publication_Place: Washington, DC
Publisher: Bureau of the Census

○ *Spatial_Domain:*
Bounding_Coordinates:
West_Bounding_Coordinate: -74.047
East_Bounding_Coordinate: -73.907
North_Bounding_Coordinate: 40.882
South_Bounding_Coordinate: 40.680
...
...
Longitude_of_Central_Meridian: 75
Latitude_of_Projection_Origin: 0

Click on the answer of your choice

Metadata for networked resources

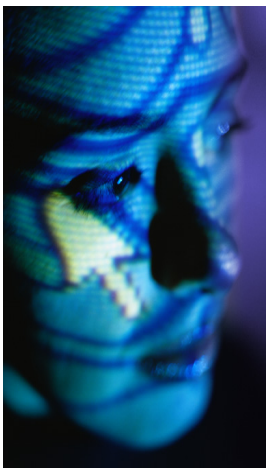


Dublin Core is a metadata standard consisting of fifteen information elements which was introduced in 1995, for describing **networked resources**.

It includes typical bibliographic elements such as title, creator, publisher, etc., but also elements that are related more to networked resources, e.g. **type** and **format** of resource, **relationships** between various resources and **intellectual property rights**.

The Basic Dublin Core, which is used to describe any network resource, has been expanded through the use of qualifiers and other elements to describe specific types of resources.

Metadata for educational resources



In order to describe educational resources, specific elements are required in addition to the common bibliographic metadata. For example, while an educational resource has an author, place and date of publication, it may also have been created for a particular educational audience ranging from grade-school students to students in institutions of higher education. Educational metadata standards have been developed to capture these elements.

The following are two standards developed for this purpose...

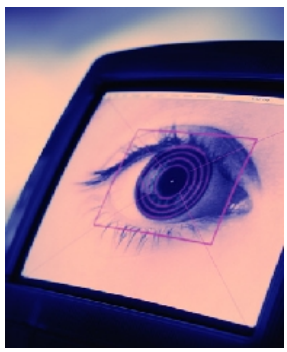
Metadata for educational resources

The **Learning Object Metadata** (**LOM**) standard has an entire section devoted to educational characteristics of an information resource. These include: interactivity type, learning resource type, interactivity level, semantic density, intended end user role, context, typical age range, difficulty, and typical learning time. The LOM standard is used by the **IMS Global Learning Consortium** and **ARIADNE**, a research and technology development project "focused on the development of tools and methodologies for producing, managing and reusing computer-based pedagogical elements and telematics supported training curricula."

Dublin Core educational metadata schema, developed by the **DCMI Education Working Group**, proposes two elements, "audience" and "standards," in addition to the basic fifteen elements of Dublin Core. The group has also proposed the adaptation of three elements from the LOM standard: interactivity type, interactivity level, and typical learning time.

Future directions in Metadata

World Wide Web is a great innovation, but it only allows humans to exchange information. On the other hand, the **Semantic Web** also enables machines to exchange information: it will explain **why** or **how** the pieces of information are connected, and it will be able to distinguish between the users' searches **based on their context**.



Semantic Web will consist of three architectural components:

• **semantic** considers what concepts actually mean;

• **structure** considers how concepts relate to one another and how they are organized; **Resource Description Framework (RDF)** will be used to manage the structure of the Semantic Web; and

• **syntax** will consider how these concepts and relations are communicated; eXtensible Markup Language (XML) will be used to manage the syntax of the Semantic Web.

Future directions in Metadata

RDF was designed as an enabling technology to support the structure and organization of resources. RDF provides a "model of statements made about resources." The RDF model consists of three object types:

Resources

All things being described by RDF expressions are called **resources**. A resource may be an entire Web page, part of a Web page, or a whole collection of pages such as an entire Web site. A resource may also be an object that is not directly accessible on the Web; e.g. a printed book.

Properties

A **property** is a specific aspect, characteristic, attribute, or relation used to describe a resource.

Statements

A **statement** consists of three parts called, respectively, the subject, the predicate, and the object.

The RDF data model provides an abstract, conceptual framework for defining and using metadata. A **concrete syntax** is also needed for the purposes of creating and exchanging this metadata. This specification of RDF uses the **XML** encoding as its interchange syntax.

This information has been taken from this site: http://www.w3.org/TR/REC-rdf-syntax/.

Summary

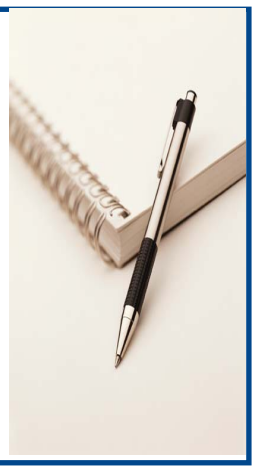
• Metadata are **structured data** that provide a **short summary about any information resource**, print or electronic.

• Metadata can provide data about the **content,** or the **context** of an information resource.

• Metadata facilitate the **location, identification, and discovery** of information resources.

• There are two different **metadata models**: the **catalogue/separated metadata** model, and the **embedded metadata** model.

• There are metadata standards specific to **geospatial** data, **web-based** resources, **print** resources and **educational** resources.



Exercises

The following six exercises will help you test your understanding of the concepts that were covered in the lesson and provide you with feedback.

Good luck!

Exercise 1

Metadata is...

- ○ information about data.
- ○ document mark-up.

Click on the answer of your choice

Exercise 2

Match each of the following types of metadata to the corresponding purpose.

ISBN

Library holdings

Keyword

Location

Identification

Resource discovery

Click each option, drag it and drop it in the corresponding box.
When you have finished, click on the confirm button.

Exercise 3

Metadata for print resources usually include:

- ○ URLs
- ○ PURLs
- ○ Subject Terms

Click on the answer of your choice

Exercise 4

Catalogue metadata for print resources connects the resource and its location via:

- ○ Subject Terms
- ○ Holdings or Classification Information
- ○ Keywords

Click on the answer of your choice

Exercise 5

Which of the following is an example of the embedded model?

- ○ Web page containing its own metadata in the source.
- ○ Database of your local video shop.
- ○ Card catalogue of the university library.
- ○ Web page with a link to a file containing its metadata.

Click on the answer of your choice

Exercise 6

Match each of the following metadata applications with the type of data they usually describe.

MARC 21	describes	1 geospatial data	2 web-based resources	
Ariadne				
ISO 19115		3 print resources	4 educational resources	
Dublin Core				

Click each option, drag it and drop it in the corresponding box.
When you have finished, click on the confirm button.

If you want to know more...

ANZLIC: http://www.anzlic.org.au/

ANZLIC metadata page: http://www.anzlic.org.au/asdi/metaelem.htm

ARIADNE: http://www.ariadne-eu.org/

DCMI Education Working Group: http://dublincore.org/groups/education/

Dublin Core Metadata Initiative (DCMI): http://dublincore.org/

Federal Geographic Data Committee (FGDC): http://www.fgdc.gov/

FGDC metadata page: http://www.fgdc.gov/metadata/metadata.html

IFLA's family of ISBDs: http://www.ifla.org/VI/3/nd1/isbdlist.htm

IMS Global Learning Consortium: http://www.imsproject.org/

International Federation of of Library Associations and Institutions (IFLA): http://www.ifla.org/

International Organisation for Standardization (ISO): www.iso.ch/

ISO's Technical Committee 211 on geographic information: http://www.isotc211.org/

Library of Congress: http://www.loc.gov/

Library of Congress' MARC standards page: http://www.loc.gov/marc/

Semantic Web: http://www.w3.org/2001/sw/