Information Management Resource Kit

Module on Management of Electronic Documents

UNIT 3. METADATA STANDARDS AND SUBJECT INDEXING

LESSON 2. METADATA STANDARDS FOR THE WEB: BACKGROUND

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.
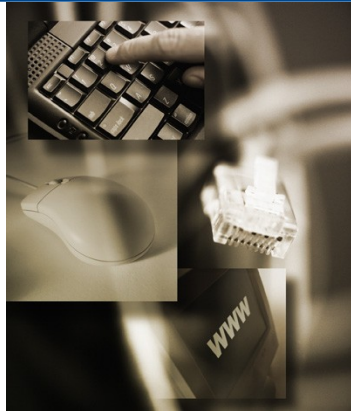
© FAO, 2003

Objectives

At the end of this lesson you will be able to:

• understand the structure of the **Dublin Core** Metadata Standard, and

• describe a network resource using **Dublin Core elements**.

Metadata and Dublin Core

Because of the rapid growth in number of **digital resources**, information professionals quickly understood the need for **metadata schemes** that could facilitate easier search and retrieval of the resources and whose application would be simpler than those schemes, like **MARC 21**, traditionally used to describe print resources.

One of these schemes, the internationally supported Dublin Core (DC) metadata standard, consisting of fifteen elements, was introduced in 1995, by the National Center for Supercomputing Applications (NCSA) and the Online Computer Library Center (OCLC), and made **Standard Z39.85** in 2001, by the National Information Standards Organization (NISO).

The Dublin Core (DC) Metadata Element Set consists of **fifteen elements** that can be used to describe the content of the information resource, the information resource itself (its manifestation), and to identify the individual responsible for creating the resource.

Elements of DC Metadata Set

| Content	Manifestation	Responsibility
Title	Format	Creator
Subject	Date	Contributor
Description	Language	Publisher
Coverage	Identifier	Rights
Source		
Relation		
Type		

These elements are optional, may be repeated, and may appear in any order. Their application will depend on the specific needs of the information professional and the information resource for which he or she is creating the metadata.
They can be represented in a number of different syntaxes, e.g. HTML and XML.

The Title

Now, you're going to take a web-based resource and use Dublin Core to create a descriptive metadata record for it.

The following image is the home page of the FAO web site: you may think of it as **your information resource**.
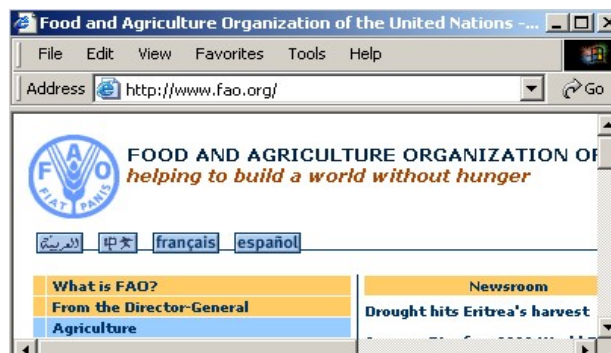
The Title

You will see the **title of the resource**, on the blue **title bar** of the window. Although all DC elements are optional, a resource title assists in accessing a resource, so let's include the title in your Dublin Core description.

Content
Title
Subject
Description
Coverage
Source
Relation
Type

Food and Agriculture Organization of the United Nations

File   Edit   View   Favorites   Tools   Help

Address http://www.fao.org/   Go

Title of the resource

FOOD AND AGRICULTURE ORGANIZATION OF
helping to build a world without hunger

العربية   中文   français   español

What is FAO?
From the Director-General
Agriculture

Newsroom
Drought hits Eritrea's harvest

The Title

For explanatory purposes, we will use **HTML** to encode Dublin Core in our resource description. These are the tags to use:

META — Since you're creating a metadata description of the resource, you will use the HTML tag most common for describing metadata: "META."

NAME — "NAME" refers to the **name of the element** being used.
The value of "NAME" will be one of the Dublin Core elements, stated like this: "DC.[element]"

CONTENT — "**CONTENT**" refers to the **actual value of the element**. The value of "CONTENT" will be taken from the value you've assigned to the element.

For the Title, the HTML tag encoded using Dublin Core would look like this:

<META NAME="DC.Title" CONTENT="Food and Agriculture Organization of the United Nations">

The Title

For example, let's take the title of this resource:

FAO: What it is, what it does - Microsoft Internet Explorer   _ □ ×

This is the record that describes the title: Can you identify what these elements correspond to?

<META NAME="DC Title" CONTENT="FAO: What it is, what it does">

1 [   ]   [   ] 2

Actual value of the element

Name of the DC element

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

Subject, Description and Coverage

In addition to Title, the **Subject**, **Description**, and **Coverage** elements are all used to access a resource. You can think of all three as going hand-in-hand.

Click on each element for a description:

| Subject | Description | Coverage |

The DC Subject element uses **keywords** or **subject phrases** that are usually chosen from a **controlled vocabulary**, a list of terms that are devised prior to the creation of the metadata records and consistently used across the collection (e.g. the FAO AGROVOC thesaurus and the thesaurus of the United States National Agriculture Library).

Example:

<META NAME="DC.Subject" CONTENT="Agriculture">

Subject	Description	Coverage

The DC Description element provides a **free-text description**, such as an abstract, of the resource, and doesn't rely on a controlled vocabulary. The DC Description tag can even be used to provide a table of contents for the resource being described.

Example

`<META NAME="DC.Description" CONTENT="This is the home page for the F.A.O. website">`

Subject	Description	Coverage

The DC Coverage element uses keywords or subject phrases that are usually taken from a specific **type** of controlled vocabulary, for example, one that deals with the **geographic location** or **time period** covered by the resource.
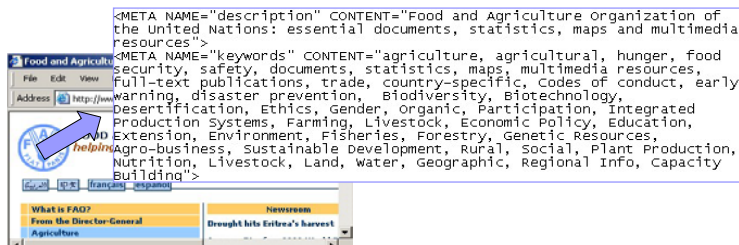
Example

`<META NAME="DC.Coverage" CONTENT="Australia">`

Specifically, when using the DC Coverage element to describe the geographic location or time period covered by a resource, named places or time periods should be used in preference to numeric identifiers, such as sets of coordinates or date ranges.

Subject, Description and Coverage

It can sometimes be difficult to tell **what a resource is about**; with web-based resources, you can always look at the **source** (HTML code) of the web page to check whether descriptive **keywords have been included**. You can see in the source of the FAO web site that a large number of keywords have been used to describe the site.

You can use some of these keywords to write the Subject and Description elements.


```
<META NAME="description" CONTENT="Food and Agriculture Organization of
the United Nations: essential documents, statistics, maps and multimedia
resources">
<META NAME="keywords" CONTENT="agriculture, agricultural, hunger, food
security, safety, documents, statistics, maps, multimedia resources,
full-text publications, trade, country-specific, Codes of conduct, early
warning, disaster prevention,  Biodiversity, Biotechnology,
Desertification, Ethics, Gender, Organic, Participation, Integrated
Production Systems, Farming, Livestock, Economic Policy, Education,
Extension, Environment, Fisheries, Forestry, Genetic Resources,
Agro-business, Sustainable Development, Rural, Social, Plant Production,
Nutrition, Livestock, Land, Water, Geographic, Regional Info, Capacity
Building">
```

Subject, Description and Coverage

Example

```
<META NAME="DC.Subject" CONTENT="Food and Agriculture Organization of the United
Nations">
<META NAME="DC.Subject" CONTENT="Agriculture">
<META NAME="DC.Subject" CONTENT="Agriculture—Economic aspects">
<META NAME="DC.Subject" CONTENT="Sustainable development">
<META NAME="DC.Description" CONTENT="This is the home page for the Food and Agriculture
Organization of the United Nations.">
<META NAME="DC.Coverage" CONTENT="Africa">
<META NAME="DC.Coverage" CONTENT="Asia">
<META NAME="DC.Coverage" CONTENT="Australia">
<META NAME="DC.Coverage" CONTENT="Europe">
<META NAME="DC.Coverage" CONTENT="North America">
<META NAME="DC.Coverage" CONTENT="South America">
```

Comments:

DC Subject: you can choose keywords that are included in your controlled vocabulary.
DC Description: you can describe the FAO web site.
DC Coverage: you could repeat the tag for each country.

Relation and Source

The **Source** and **Relation** elements of Dublin Core are used to relate information resources to one another.

The Source element contains information about a second resource from which the present resource is derived.
The Relation element indicates that the resource is **related in some way** to another resource.
For example, to connect the main FAO web site to an FAO divisional web site:

```
<META NAME="DC.Relation" CONTENT="http://www.fao.org/ag">
```

Type

The **Type** element describes what the information resource consists of. Values for Type should be selected from a controlled vocabulary such as the working draft **list of Dublin Core Types**:

- Collection
- Dataset
- Event
- Image
- Service
- Software
- Sound
- Text
- Interactive Resource

Many web resources, such as an initial home page, are best described with **more than one type** element, depending on what kinds of information the web resource consists of. The FAO web site can best be described as both an interactive resource and a service:

```
<META NAME="DC.Type" CONTENT="Service">
<META NAME="DC.Type" CONTENT="Interactive Resource">
```

Format

Manifestation

- Format
- Date
- Language
- Identifier

Format is the first element used to provide information about the resource itself.

It provides information about what kind of **hardware or equipment** is **needed to use** or access the resource. Values for Format should be selected from a controlled vocabulary such as the list of Internet Media Types (MIME) defining computer media formats.

The FAO web site consists, in part, of text encoded in html and images encoded in .gif format. It also may have a video encoded in Quicktime. The metadata description would repeat the DC Format tag several times and would look like this:

```
<META NAME="DC.Format" CONTENT="text:html">
<META NAME="DC.Format" CONTENT="image:gif">
<META NAME="DC.Format" CONTENT="video:quicktime">
```

Date, Language and Identifier

In addition to the Format element, Dublin Core uses three other elements, **Date**, **Language** and **Identifier**, to provide information about the resource itself.

Click on each element for a description:

| Date | Language | Identifier |

The Date tag is used to describe when a resource was originally **manifested**, **created**, or **made available**, or when a resource was **changed** in any way. Date values that are defined in a profile of ISO 8601 and follow the YYYY-MM-DD format should be used.

All you know from the FAO website is the year that the site was last updated: 2002.
You could assume that FAO updates its site on a daily basis, and thus use the current month and day for the rest of the date (e.g. 10-09). The DC Date tag would look like this:

```
<META NAME="DC.Date" CONTENT="2002-09-23">
```

| Date | Language | Identifier |

The Language tag is used to describe the language in which the resource was created. The values of the Language should be those defined by RFC 1766 which includes a two-letter language code, followed optionally by a two-letter country code. For example, use 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.
In the case of the FAO web site, it's apparent that the site is available in English, French, and Spanish, so we would add *three* DC Language tags to our metadata record:
```
<META NAME="DC.Language" CONTENT="en">
<META NAME="DC.Language" CONTENT="fr">
<META NAME="DC.Language" CONTENT="sp">
```

| Date | Language | Identifier |

The Identifier tag is used to unambiguously identify an information resource within a given context. This identifier should be a string or number conforming to a formal identification system. Examples of formal identification systems include the URI/URL, the DOI and the ISBN.

For the FAO web site, it is apparent that its' URL is http://www.fao.org, so the DC Identifier tag for the site would look like this:

```
<META NAME="DC.Identifier" CONTENT="http://www.fao.org">
```
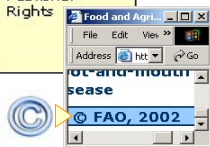
Responsibility

Dublin Core is also used to provide information about those **responsible** for the resource. The **creator**, **contributor**, and **publisher** can all be either individual persons or organizations, but usually the publisher is an organization. The **Rights** element provides information about intellectual and other property rights and copyright.

In our example, FAO is the creator, publisher and also holds property rights. Therefore, it is important to use the same value used in the title: "Food and Agriculture Organization of the United Nations".

Who contributed to the creation of the FAO web site is not clear, so you won't include a DC Contributor tag.

Responsibility

Creator
Contributor
Publisher
Rights

Food and Agri...
File   Edit   Viev
Address   htt   Go
ot-and-mouth
sease
© FAO, 2002

<META NAME="DC.Creator" CONTENT=" Food and Agriculture Organization of the United Nations">
<META NAME="DC.Publisher" CONTENT="Food and Agriculture Organization of the United Nations">
<META NAME="DC.Rights" CONTENT="© Food and Agriculture Organization of the United Nations">

HTML/XML

The same metadata description using Dublin Core can be encoded in **XML**.

In this example the same elements (Format, Date, Language and Identifier) are encoded using the two different syntaxes:

In HTML…

```
<META NAME="DC.Format" CONTENT="text:html">
<META NAME="DC.Format" CONTENT="image:gif">
<META NAME="DC.Format" CONTENT="video:quicktime">
<META NAME="DC.Date" CONTENT="2002-09-23">
<META NAME="DC.Language" CONTENT="en">
<META NAME="DC.Language" CONTENT="fr">
<META NAME="DC.Language" CONTENT="sp">
<META NAME="DC.Identifier" CONTENT="http://www.fao.org">
```

In XML…

```
<dc:format>text:html</dc:format>
<dc:format>image:gif</dc:format>
<dc:format>video:quicktime</dc:format>
<dc:date>2002-09-23</dc:date>
<dc:language>en</dc:language>
<dc:language>fr</dc:language>
<dc:language>sp</dc:language>
<dc:identifier>http://www.fao.org</dc:identifier>
```

Summary

• Dublin Core (DC) metadata standard is a standard for describing **Web** resources.

• DC consists of **fifteen elements**, that can be used to describe the **content** of the information resource, the information **resource itself** and to identify who is **responsible** for creating the resource.

• These elements are **optional** , may be repeated, and may appear in **any order**.

• They can be represented in a number of **different syntaxes**, e.g. HTML and XML.

Exercise

The following exercise will help you test your understanding of the concepts that were covered in the lesson and will provide you with feedback.
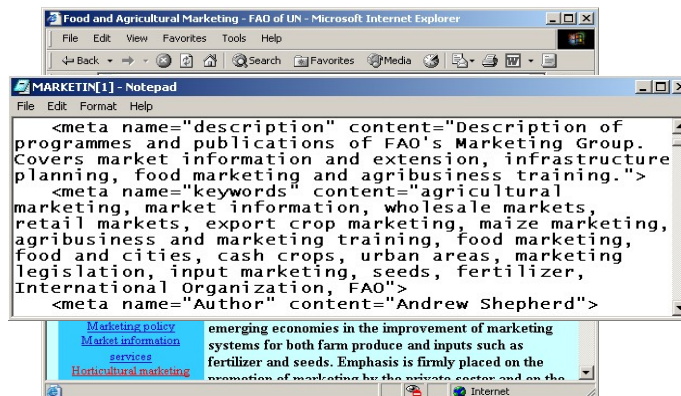
Good luck!

Now it's your turn to use the DC elements!

This web page, within the FAO website, is about Agricultural Marketing. In the following screens we will provide you with some information, that will allow you to describe this web page using the DC elements.

By clicking on **Source** from the **View** menu option, you can access the web page source code.

Exercise

This information can also be useful to you:

This section covers the following geographic areas: Africa, Asia, Australia, Europe, North and South America.

The languages used are the five FAO official languages: English, French, Spanish, Arabic and Chinese.

The web page includes text, images and interactive resources.

The format used in the page is **html** for the text and .**gif** for the images.

The publisher is the FAO Agricultural Department; the publication date and copyright information are not available.

Exercise

Now try to assign the correct value to the DC elements in this template. Remember that all these elements are optional, repeatable and may appear in any order.

We entered the first value to get you started.

| Web Page |
| Source |
| Other info |

DC.Title = Food and Agricultural Marketing – FAO of UN
DC.Subject=
DC.Description=
DC.Coverage =
DC.Source=
DC.Relation =
DC.Type=
DC.Date=
DC.Format=
DC.Date=
DC.Language=
DC.Identifier=
DC.Creator=
DC.Contributor=
DC.Publisher=

| View Answer |

*Type the text in the box. When you are finished, click on **View Answer**.*

DAML+OIL: http://www.daml.org/
Dublin Core Metadata Initiative: http://www.dublincore.org/about/organization/
Dublin Core Type list: http://www.dublincore.org/documents/dcmi-type-vocabulary/
Dublin Core Usage Guide: http://www.dublincore.org/documents/usageguide/
International Organization for Standardization (ISO): http://www.iso.ch/
International Standard Book Number (ISBN): http://www.isbn.org/
Internet Media Types (MIME) list:
http://www.isi.edu/in-notes/iana/assignments/media-types/media-types

Machine Readable Cataloging (MARC) 21: http://www.loc.gov/marc/
National Center for Supercomputing Applications (NCSA): http://www.ncsa.uiuc.edu
National Information Standards Organization (NISO): http://www.niso.org/
NISO Z39.85: http://www.niso.org/standards/resources/Z39-85.pdf
Ogbuji, Uche. "The Languages of the Semantic Web." *New Architect*, June 2002
Online Computer Library Center (OCLC): http://www.oclc.org/
Ontologies: http://www-ksl.stanford.edu/kst/ontology-sources.html
Request for Comment (RFC) at the Internet Engineering Task Force (IETF):
 http://www.ietf.org/rfc.html
Resource Description Framework (RDF): http://www.w3.org/RDF/
RFC 1766: http://www.ietf.org/rfc/rfc1766.txt
Extensible Markup Language (XML): http://www.w3.org/XML/
XML Topic Maps: http://www.topicmaps.org/xtm/1.0/