CANADIAN CULTURE ONLINE

Metadata

La version française de ce document est aussi disponible.

CCO Best Practice Guide

Metadata

CCO Requirement – 6.1 Dublin Core Metadata

a) The following six mandatory Dublin Core elements **must** be applied to describe the most important pages of the website, such as the home page, each section's main page, and pages featuring resources for which there is sufficient context and meaning and are also worth listing in a search engine:

- Title
- Creator
- Subject
- Date Created
- Language (where applicable)
- Identifier

b) Metadata **must** be embedded in the `<head>` section of the (X)HTML page.

c) Each page described **must** feature a unique set of metadata content.

Note: Copying and pasting the same metadata content is bad practice as it makes pages compete among themselves to be retrieved. At the very least, no two pages should have the same identifier, title and list of subjects.

d) Metadata content **must** be in the language of the page.

Note: This means that a page in English **must** include metadata content in English, and a page in French, metadata content in French. The syntax (e.g., `<meta name="dc.title" content="" />` is never translated. In cases where a page is bilingual or multilingual, metadata elements Creator, Title, Subject and Language **must** be repeated to reflect all languages; however, in such cases, it is not needed to repeat the Date Created and Identifier elements.

Definitions of mandatory elements

The definitions below are excerpted from the document, "Using Dublin Core", available at http://www.dublincore.org/documents/usageguide/. In all definitions, the word "resource" can be read to mean a Web page or a website.

Comments are meant to clarify the definitions or give useful tips about how to complete the content of an element.

The syntax in the examples below is provided for XHTML. In HTML, the trailing slash ("/") is not required before the closing bracket (">").

Title

Definition: The name given to the resource. Typically, a Title will be a name by which the resource is formally known.

Comment: For an (X)HTML page, the content of the (X)HTML `<title>` element and the Dublin Core Title element should be the same. Make sure your page title is significant and descriptive enough so that a user immediately gets an idea of the page's content. Also, search engines tend to give more weight to the words included in the title when indexing the page.

Syntax: `<meta name="dc.title" content="`**`insert the title of the Web page being described`**`" />`

Creator

Definition: An entity primarily responsible for making the content of the resource. Examples of a Creator include a person, an organization, or a service. Typically the name of the Creator should be used to indicate the entity.

Comment: Personal names should be listed family name first, followed by a comma, and ending with the first name. If a Web page has more than one creator, repeat the element as many times as needed.

Syntax: `<meta name="dc.creator" content="`**`insert the name of the Web page's creator`**`" />`

Subject

Definition: The topic of the content of the resource. Typically, a Subject will be expressed as keywords or key phrases or classification codes that describe the topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Comment: It is advisable to use keywords that are significant and unique, neither too broad nor too general. Think of how your targeted audience would describe the page being searched for. Keywords can come from the page's text, or can be drawn from a formal source, such as the Art and Architecture Thesaurus or Canadian Subject Headings. Multiple keywords can be separated by consistently using commas or semicolons, but not both. A good practice is to include five to seven keywords.

Syntax: `<meta name="dc.subject" content="`**`insert keywords describing the Web page, separated by comma`**`" />`

It is best practice **to limit the number of keywords** to between five and seven. While there is no official limit to the number of words a Dublin Core element can contain, search engines and harvesters tend to interpret longer lists of keywords as spam. Typically, a well-designed Web page will not deal with a wide range of subjects. If a page does address numerous topics, it is normally preferable to break it into several pages, each one dealing with a single issue or idea.

Date created

Definition: Date of creation of the resource.

Comment: The Date element should not be used alone; rather, use it with one of its element refinements. The refinement, Date Created, is mandatory. Other refinements include Date Modified and Date Issued. It is strongly recommended to write the date using the international YYYY-MM-DD format, such as 2006-12-20. If the month or the day is unknown, the value "01" is entered, as in these examples: 2006-12-01 or 2006-01-01.

Syntax: `<meta name="dcterms.created" content="`**`insert the date of creation of the Web page, using the YYYY-MM-DD format`**`" />`

Language

Definition: A language of the intellectual content of the resource.

Comment: A name of a language or a code representing it can be used here. A recommended practice is to use the two- or three-letter codes as defined in the ISO 639 standard, "Codes for the Representation of Names of Languages". Many Canadian Aboriginal languages are included in the three-letter codes section. The codes are available at http://www.loc.gov/standards/iso639-2/. If a Web page features more than one language, repeat the element as many times as needed.

Syntax: `<meta name="dc.language" content="`**`insert the language name or code for the Web page being described`**`" />`

Identifier

Definition: An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string or number conforming to a formal

identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Comment: Local identifiers, as assigned by the Web page creator or a database entry ID, can also be used.

Syntax: `<meta name="dc.identifier" content="`**an identifier can be the URL of the Web page being described**`" />`

Encoding

To facilitate compliance with the requirements, the template below can be directly embedded in the `<head>` element of an (X)HTML page. Simply fill it out with the appropriate metadata that best describes a given Web page.

Figure 1: Sample metadata code for a unilingual page

```
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.dcterms" href="http://purl.org/dc/terms/" />
<meta name="dc.title" content="insert the title of the Web page
being described" />
<meta name="dc.creator" content="insert the name of Web page's
creator" />
<meta name="dc.subject" content="insert keywords describing the Web
page, separated by comma" />
<meta name="dcterms.created" content="insert the date of creation of
the Web page being described, using the YYYY-MM-DD format" />
<meta name="dc.language" content="insert the language name or code
for the Web page being described" />
<meta name="dc.identifier" content="an identifier can be the URL of
the specific Web page being described" />
```

The template above is provided for XHTML. In HTML, the trailing slash ("/") is not required before the closing bracket (">"), as in this example:

`<meta name="dc.creator" content="`**insert the name of the Web page's creator**`">`

Recipients are required to describe the website's most significant pages, such as the welcome page, each section's main page, and pages that offer added value. Each page described must feature a unique set of metadata content.

A page in English must include metadata content in English. A page in French must include metadata content in French. The syntax (for example `<meta name="dc.title" content="" />`) is never translated. In cases where a page is bilingual or multilingual, the Creator, Title, Subject, and Language

metadata elements must be repeated to reflect all languages; however, in such cases, it is not necessary to repeat Date Created and Identifier. The order of the elements does not matter.

Figure 2: Sample metadata code for a bilingual (English and French) page

```
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.dcterms" href="http://purl.org/dc/terms/" />
<meta name="dc.title" content="insert English title of the Web page
being described" />
<meta name="dc.title" content="insert French title of the Web page
being decsribed " />
<meta name="dc.creator" content="insert the name of the Web page's
creator in English" />
<meta name="dc.creator" content="insert the name of the Web page's
creator in French" />
<meta name="dc.subject" content="insert English keywords describing
the Web page, separated by comma" />
<meta name="dc.subject" content="insert French keywords describing
the Web page, separated by comma" />
<meta name="dcterms.created" content="insert the date of creation of
the Web page, using the YYYY-MM-DD format" />
<meta name="dc.language" content="English" />
<meta name="dc.language" content="Français" />
<meta name="dc.identifier" content="an identifier can be the URL of
the specific Web page being described" />
```

Optional Dublin Core elements

The Dublin Core Metadata Element Set (DCMES) features 15 elements and a number of element refinements. Some of them might be useful to further describe Web pages and special sections of a website, such as lesson plans and collections of pictures or audio-visual materials. Below is an overview of selected additional elements.

The definitions are excerpted from the document, "Using Dublin Core", available at http://www.dublincore.org/documents/usageguide/. In all definitions, the word "resource" can be read to mean a Web page or a website.

Comments are meant to clarify the definitions or give useful tips about how to complete an element's content.

The syntax is provided for XHTML. In HTML, the trailing slash ("/") is not required before the closing bracket (">").

Audience

Definition: A class of entity for whom the resource is intended or useful. A class of entity may be determined by the creator or the publisher or by a third party.

Comment: This element is used to describe the intended audience of a Web page. It is particularly useful to describe lesson plans and other learning materials. Values, such as "students", "teachers", "caregivers", or "general public", can be used. It is recommended to draw terms from a controlled source, or to develop a list of words describing a website's target audience and to use it consistently. If a Web page targets more than one audience, repeat the element as many times as needed.

Syntax: `<meta name="dcterms.audience" content="`**`insert a keyword describing the target audience of the Web page being described`**`" />`

Description

Definition: An account of the content of the resource. Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

Comment: Significant words are usually used in a description, thus making the content of this element highly interesting for search engines. It is a good practice to have the same content in the Dublin Core Description element and the (X)HTML <meta> "description" element.

Syntax: `<meta name="dc.description" content="`**`insert a description of the content of the Web page being described`**`" />`

Format

Definition: The physical or digital manifestation of the resource. Typically, Format may include the media-type or dimensions of the resource. Examples of dimensions include size and duration. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource.

Comment: The recommended practice is to draw a term from a controlled source, such as the list of Internet Media Types that describes file formats. The list is available at http://www.iana.org/assignments/media-types/. Common formats used on a Web page include html, jpeg, png, mpeg, etc. Format can also be used to express the size or duration (for example, 400 x 600 pixels; 4 kB; 10m23s). If a Web page features more than one format, repeat the element as many times as needed.

Syntax: `<meta name="dc.format" content="`**`insert the file format for the content of the Web page being described`**`" />`

Rights

Definition: Information about rights held in and over the resource. Typically a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), copyright, and various property rights. If the rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

Comment: The Rights element may include either a textual statement, or a URL pointing to a rights statement, or a combination when both a brief statement and a more lengthy one are available.

Syntax: `<meta name="dc.rights" content="`**`insert a textual copyright statement about the page being described, or a URL leading to it`**`" />`

Type

Definition: The nature or genre of the content of the resource. Type includes terms describing general categories, functions, genres, or aggregation levels for content.

Comments: A Web page can feature content such as still images, text, lesson plans, events, moving images, sounds, interactive resources, etc. More types can be found in the "Dublin Core Metadata Initiative (DCMI) Type Vocabulary", available at http://www.dublincore.org/documents/dcmi-type-vocabulary/. If a page features more than one type, repeat the element as many times as needed.

Syntax: `<meta name="dc.type" content="`**`insert the type of resource available on the Web page being described`**`" />`

Figure 3: Sample metadata code with mandatory and optional Dublin Core elements

```
<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.dcterms" href="http://purl.org/dc/terms/" />
<meta name="dc.title" content="insert the title of the Web page
being described" />
<meta name="dc.creator" content="insert the name of Web page's
creator" />
<meta name="dc.subject" content="insert keywords describing the Web
page, separated by comma" />
<meta name="dcterms.created" content="insert the date of creation of
the Web page being described, using the YYYY-MM-DD format" />
<meta name="dc.language" content="insert the language name or code
for the Web page being described" />
<meta name="dc.identifier" content="an identifier can be the URL of
the specific Web page being described" />
<meta name="dcterms.audience" content="insert a keyword describing
the target audience of the Web page being described" />
```

```
<meta name="dc.description" content="insert a description of the
content of the Web page being described" />
<meta name="dc.format" content="insert the file format for the
content of the Web page being described" />
<meta name="dc.rights" content="insert a textual copyright statement
about the Web page being described, or a URL leading to it" />
<meta name="dc.type" content="insert the type of resource available
on the Web page being described" />
```

Other Ways of Using Dublin Core Metadata

Dublin Core metadata elements can also be used to describe large collections of pictures or sound, or to manage website content in content management systems (CMS). Databases are usually used behind the scene. Fields in those databases can be called after the Dublin Core elements, or can be automatically translated into (X)HTML code embedded as shown in the examples provided above. As the content in those databases is not directly exposed to search engines, it is a good practice to take full advantages of a database metadata fields by using them in the internal search engine of a Web site..

References

DCMI Metadata Terms. http://www.dublincore.org/documents/dcmi-terms/.

Expressing Dublin Core in XHTML and HTML. http://www.dublincore.org/documents/dcq-html/.

Government of Canada Metadata Implementation Guidelines for Web Resource Discovery, 5th Edition, October 2006. http://publiservice.tbs-sct.gc.ca/im-gi/mwg-gtm/ts-sf/docs/2006/metaweb/metawebtb_e.asp. (Note: Although these guidelines specifically apply to federal organizations, the guidance found can be useful to CCO fund recipients.)

Using Dublin Core. http://www.dublincore.org/documents/usageguide/.