

Case Study

Submitted to Dublin Core Metadata Initiative, Global Corporate Circle
by Sarah A. Rice, Senior Information Architect, Seneb Consulting, USA

www.seneb.com.

May, 2004

Introduction

This metadata schema was developed for a public-facing enterprise-class web site initially involving a few thousand web pages and documents. The project included a site redesign and initial implementation of a content management system. The metadata schema was used to support findability of content on the web site through searching and browsing, as well as the CMS implementation.

Before this project, the only metadata elements used on the company's web site were title, keywords and description. Keywords and description were kept as optional metadata elements in the new schema as a marketing strategy for search engine optimization.

Methodology

The metadata schema was constructed from a thorough analysis of each of the company's content types, an analysis of the new design system and page layout for the site redesign, and the current understanding of the CMS technology that would manage, store and publish the content. It was intended to support web site changes in the future as well as current needs. Standards were adopted so the schema could adapt to changes which include merging with other web sites or technologies, frequent and significant changes within the top levels of the taxonomy, and ongoing re-organization of all levels of content.

Standards

It was decided to use existing metadata standards wherever possible. Adhering to a standard allowed for interoperability among the company's web sites and with other software applications that interact with the company web site. Dublin Core was the standard that was adopted for this project.

About the Author

Sarah Rice is an Information Architect trained in Library and Information Science. She has 8+ years experience organizing information in complex networked environments for companies in Silicon Valley. She is currently the co-chair of the DCMI Global Corporate Circle.

Acknowledgements

Thank you to Avi Rappoport of Search Tools Consulting (www.searchtools.com) and Peter Morville of Semantic Studios (www.semanticstudios.com) for assistance in developing this metadata schema.

Metadata Schema

This document details most elements used for this project. Some elements used to implement CMS were not included.

Title Elements

3 different metadata elements were captured by content contributors (page title, subtitle and short title). From these 3 elements, a number of different types of titles were extrapolated by the system for different metadata needs (Browser title, external search title, internal search title, page title, and internal CMS title).

Elements captured by content contributors

Page Title

Description: This describes the name given to a particular web page or content type. When associated with a web page, this element became the title that was shown on the web page within the CMS template.

Format: Free text field.

Required

Notation: dDoc title (metadata element already provided by commercial content management software)

Subtitle

Description: A secondary name given to a web page or content type.

Always appears in conjunction with the page title.

Format: Free text field.

Optional

Short Title

Description: An abbreviated version of the Page Title for a web page or content type. Only used for specific content types.

Format: Free text field.

Optional

Generated by system:

External Search Title

Description: This item describes the name given to a particular web page or content type. This title displays on browser chrome and for search results of external search engines (such as Google).

Input method: system generates this element using “title” and “subtitle” free text fields input by content contributors.

Required

Notation: title

Value Notation: <company name> <title> <subtitle>

Internal Search Title

Description: Similar to external search title, except this element is used to display search results for company's own web site search engine.

Input method: system generates this element using "title" and "subtitle" free text fields input by content contributors.

Required

Notation: dc.title

Value Notation: <title> <subtitle> <company name>

Internal CMS Title

Description: This item describes the name given to a particular web page or content type. This title is displayed in search results within the content management system.

Input method: system generates this element using "title" and "subtitle" free text fields input by content contributors.

Required

Owner

Description: The company division and job title of person ultimately responsible for content.

Input method: choose from controlled list.

Required

Notation: dc.publisher

Value Notation: <division name> <job title>

Subject

Description: describes subject matter of the content. Used in conjunction with the company's taxonomy, including the official list of products, services and solutions offered by the company.¹

Input method: choose from controlled list.

Required

Notation: dc.subject

Value Notation: for a product page, the value would have been displayed as:

<division name> <product line name> <product family name> <product name>.²

¹ "Subject" was used instead of creating a "product" element. Using "subject" allowed for the inclusion of all aspects of the company's taxonomy such as services, finance and customer support. Encoding all levels of the taxonomy within a subject element's value provided a more robust schema to serve the findability needs of the company's content.

² Note: the entire hierarchy of the company's taxonomy was captured within the value notation of the subject element.

Keywords

Description: words used to describe subject matter of content, including synonyms.³

Input method: free text field.

Optional

Notation: keywords

Description

Description: Phrase or sentence format describing subject matter of content.³

Input method: free text field

Optional

Notation: description

Creation Date

Description: Date that content type was created or first input into system.

Input method: automatically generated by system

Required

Notation: dc.date.created

Value Notation: YYYY-MM-DD

Last Updated Date

Description: Date that content type was most recently changed.

Input method: automatically generated by system

Required

Notation: dc.date

Value Notation: YYYY-MM-DD

Display Date

Description: Date used for display (usually on web page). Could be different from creation or release date.

Input method: text input by content contributor

Optional

Notation: dc.date.alternative

Value Notation: varies depending on individual need.

Release Date

Description: Date that content type is published to web site for viewing by user community.

³ Left over from legacy metadata used previously in company. These elements were kept as optional metadata elements in the new schema as a marketing strategy for search engine optimization.

Input method: default suggested by system; override possibility by content contributor

Required

Notation: dc.date.issued

Value Notation: YYYY-MM-DD

Expiration Date

Description: Date that content should be removed from web site/public view.

Content contributor can choose to have this happen automatically or have system generate notice to be done manually.

Input method: default suggested by system; override possibility by content contributor

Required

Notation: dc.date.expires

Value Notation: YYYY-MM-DD

Content Type

Description: describes the type of content being put into the system. Examples might include press release, white paper, data sheet.

Input method: choose from controlled list.

Required

Notation: dc.type

Language

Description: denotes the language of the resource's content

Input method: default suggested from controlled list.

Required

Notation: dc.language

Value notation: taken from <http://www.loc.gov/standards/iso639-2/englangn.html>

Country

Description: denotes a particular geographical audience, or published within the company's country-specific web site.

Input method: default suggested from controlled list.

Required

Notation: dc.publisher.country

Value Notation: taken from http://userpage.chemie.fu-berlin.de/diverse/doc/ISO_3166.html

Audience

Description: a class or group of users that a particular resource would be targeted for.

Input method: default suggested from controlled list.

Required

Notation: dc.audience

Notes and Reflections:

After completing the project and reviewing the schema for submission as a case study to DCMI, I noticed a couple of things:

1. I was hoping to find a way to more clearly explain the titles. Based on the need of the project balanced with the need to simplify what content contributors were required to enter, gathering 3 elements and having the system generate all the needed title tags seemed the best way to go, but I haven't been able to find a way to clearly articulate that in a document.
2. the Country element: I realize now that I might have used the dc.coverage tag instead of dc.publisher. I'm sure many conversations could be had around the best practice and what long-term implications might be. However, this was my first implementation of Dublin Core, so I cannot claim perfection or wisdom as to how it should be done.