

State of Minnesota

Best Practice Guidelines for Web Metadata

Presented to the Legislative Commission on Minnesota Resources and the
Minnesota Office of Technology by

Colleen Mlecoch, Foundations Project Manager
and

Eileen Quam, Foundations Project Coordinator

eileen.quam@dnr.state.mn.us

<http://bridges.state.mn.us/bestprac/intro.pdf>



Minnesota Department of Natural Resources
St. Paul, Minnesota

June 2000

Revised April 2002

Funding for this project was from the Minnesota Future Resources Fund, as recommended by the
Legislative Commission on Minnesota Resources (LCMR).
(ML 1997, Ch.216, Sec. 15 Subd. 11(a) and ML 1999, Ch 231, Sec. 16 Subd. 25)

State of Minnesota
Best Practices for Web Metadata

Table of Contents

- I. [Introduction and Overview](#)
 - A. Foundations Project Summary
 - B. Metadata Overview
 - C. Recommended Metadata Standards and Software

- II. [Minnesota Metadata Guidelines Training Manual](#)
 - A.. User Guide for Dublin Core Metadata
 - B. Dublin Core Glossary
 - C. TagGen Download Information
 - D. TagGen Basic Instruction Guide
 - E. TagGen Pointer File Creation
 - E. Visual Help Sheets

- III. [Bibliographies](#)

- IV Appendices
 - A. [Information Needs Assessment Study](#)
 - B. Foundations Project Usability Studies
 - 1. [Dublin Core Metadata and Controlled Vocabulary](#)
 - 2. [Bridges User Interface](#)
 - C. [Background Reports](#)
 - 1. Thesauri
 - 2. Dublin Core
 - 3. GILS
 - 4. XML/RDF

Context, Purpose and Scope

This document is the result of a Legislative Commission on Minnesota Resources (LCMR) project, “Foundations for Integrated Access to Environmental Information.” The purpose of the project was to research, evaluate, select, and implement tools and techniques to improve the discovery of and access to Minnesota environmental and natural resource data and information.

The suite of tools and techniques selected and implemented include: the Inktomi/Ultraseek search engine, the Dublin Core Metadata standard, the Legislative Indexing Vocabulary (LIV), and TagGen -- short for metatag generator -- an application developed by Hiawatha Island Software (www.hisoftware.com) that provides a means to easily create and embed metadata in HTML documents. These tools are described in more detail in this document.

The purpose of these guidelines is to provide a training and reference manual to help users create descriptive records for electronic information resources. Creators of these records include catalogers, authors, editors, and web site administrators. Another important purpose is to promote best practices for describing resources using the Dublin Core metadata standard. Consistency in creating metadata is an important key to achieving complete retrieval and intelligible display across disparate sources of descriptive records. Inconsistent metadata effectively hides desired records, resulting in uneven, unpredictable, or incomplete search results.



For more information about *Foundations*,
please visit the project web site:
<http://bridges.state.mn.us>

or contact:

Colleen Mlecoch, Chief Information Officer	651.296.1305
Eileen Quam, Information Architect	651.297.2341
Minnesota Department of Natural Resources	

Funding for the Foundations Project was approved by the Minnesota Legislature, ML 1997, Chapter 216, Section 15, Subd. 11(a) Public Access to Natural Resource Data, as recommended by the Legislative Commission on Minnesota Resources from the Future Resources Fund.

Metadata

- Is a concept that predates the web, having been coined in the 1960s to describe datasets effectively.
- Is generally defined as data about data.
- Provides basic information to describe the content of an information object such as the author of a work, the date of creation, links to any related works, etc. (A familiar form of metadata is the OCLC MARC record used to describe materials of various types and formats found in libraries.)
- Is used to facilitate information discovery and retrieval in a networked environment.

Metadata is important for information retrieval

- It assists the effective discovery and retrieval of networked information resources. Metadata facilitates more sophisticated and comprehensive searching of information as the metadata elements and structures are designed to analyze content of the data in depth.
- It manages large amounts of data with low network bandwidth. Metadata addresses the issue of indexing large quantities of data of various types without requiring enormous network bandwidth. What gets indexed is the representational data rather than the information object itself.
- It integrates heterogeneous information resources. Information resources exist in different formats with different features in heterogeneous databases. Standard metadata description permits the comparing, sharing, integrating and reusing of various types of data in a distributed network environment. It is an important approach for finding information in heterogeneous databases.
- It controls restricted access information. Metadata can manage restricted-access information and services to users, e.g. billing, filtering and rating, privacy, and security. Metadata serves a gatekeeper function, an indispensable feature electronic commerce.

DUBLIN CORE METADATA TEMPLATE

```
<HTML>
  <HEAD>
    <TITLE></TITLE>
    <meta name="DC.Title" content="">
    <meta name="DC.Subject" content="">
    <meta name="DC.Description" content="">
    <meta name="DC.Creator" content="">
    <meta name="DC.Publisher" content="">
    <meta name="DC.Contributor" content="">
    <meta name="DC.Date.Creation" scheme="ISO 8601" content="yyyy-mm-dd">
    <meta name="DC.Date.Modified" scheme="ISO 8601" content="">
    <meta name="DC.Type" content="Text">
    <meta name="DC.Format" scheme="HTML" content="text/html">
    <meta name="DC.Identifier" content="">
    <meta name="DC.Relation" content="">
    <meta name="DC.Source" content="">
    <meta name="DC.Rights" content="">
    <meta name="DC.Language" scheme="ISO639-1" content="en">
    <meta name="DC.Coverage" content="">
    <LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements">
  </HEAD>
  <BODY> ... [document body begins]
</HTML>
```

FOUNDATIONS PROJECT D.C. METADATA EXAMPLE

```
<HTML>
  <HEAD>
    <TITLE>Minnesota Berry Growers Directory</TITLE>
    <!-- Start Dublin Core - Do Not Modify Tags in This Block -->
    <!-- Dublin Core Meta Tags generated by TagGen - The Meta Tag Management System -->
    <meta name="DC.Title" scheme="AACR2" content="Minnesota berry growers directory">
    <meta name="DC.Subject" scheme="LIV-MN" content="Fruit trade -- Minnesota">
    <LINK REL=SCHEMA.DC HREF="http://www.bridges.state.mn.us/servlet/lexico">
    <meta name="DC.Subject" scheme="Keyword" content="Berries -- Minnesota">
    <meta name="DC.Description" content="Directory of berry growers in Minnesota by region: Twin
    Cities, northeast, northwest, southeast, southwest.">
    <meta name="DC.Creator.CorporateName" scheme="AACR2" content="Minnesota Dept. of
    Agriculture">
    <meta name="DC.Publisher.CorporateName" scheme="AACR2" content="Minnesota Dept. of
    Agriculture">
    <meta name="DC.Date.Creation" scheme="ISO 8601" content="1998-00-00">
    <meta name="DC.Type" content="Text">
    <meta name="DC.Format" scheme="HTML" content="text/html">
    <meta name="DC.Identifier" scheme="URL"
    content="www.mda.state.mn.us/docs/mktg/98berry2.htm">
    <meta name="DC.Language" scheme="ISO639-1" content="en">
    <LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements">
    <!-- End Dublin Core - Do Not Modify This Block -->
    <!-- Dublin Core Meta Tags generated by TagGen 1.1 - The Meta Tag Management System -->
  </HEAD>
<BODY>
```

Recommended Standards

- Metadata – Dublin Core (ANSI/NISO standard Z39.85)
- Controlled vocabulary/subjects – Legislative Indexing Vocabulary, Minnesota edition (LIV-MN)
- Name authority – Anglo-American Cataloging Rules/Library of Congress (AACR2/LC)
- Geographic areas – AACR2/LC
- Dates – ISO 8601
- Language – ISO 639-1
- Punctuation and capitalization – see User Guide for Dublin Core Metadata

Recommended Software

- Search engine – Inktomi (formerly Ultraseek)
- Browseable topic categories – Inktomi/Ultraseek Content Classification Engine (CCE)
- Thesaurus – Lexico
- Metatag generation – TagGen Dublin Core edition