

# **Minnesota Metadata Guidelines – Dublin Core**

## **MMG - DC**

### *User Guide For Dublin Core Elements*

**Version 8/8/1999**

**Located in PDF at:**

<http://bridges.state.mn.us/MMG-DCUserGuide.pdf>

**Adapted from *A User Guide for Simple Dublin Core*, draft version 6b**

<http://128.253.121.110/DCGuideliones/UserGuide6b.html>

**Developed by:**

The Foundations Project at the  
Minnesota Department of Natural Resources

Eileen Quam  
Project Coordinator & Metadata Specialist  
Robert Olson

*Funding for the Foundations Project was approved by the Minnesota Legislature,  
ML 1997, Chapter 216, Sec.15, subd. 11(A), as recommended by the  
Legislative Commission on Minnesota Resources from the Minnesota Future Resources Fund.*

# Minnesota Metadata Guidelines

## A User Guide for Dublin Core Elements

### INTRODUCTION

#### 1. What is metadata?

Metadata describes an information resource. The term "meta" derives from the Greek word "denoting a nature of a higher order or more fundamental kind," such as metalanguage or metatheory. "Metadata," then, is data about other data.

A metadata record consists of a set of attributes, or elements, necessary to describe a resource. For example, a metadata system common in libraries -- the library catalog -- contains a set of metadata records with elements that describe a book or other library item: author, title, date of creation or publication, subject coverage, and the call number specifying location of the item on the shelf.

The linkage between a metadata record and the resource it describes may take one of two forms:

- a. the metadata may be embedded in the resource itself; or
- b. elements may be contained in a record separate from the item, as in the case of the library's catalog record.

Examples of embedded metadata that is carried along with the resource itself include the Cataloging In Publication (CIP) data printed on the verso of a book's title page; or the TEI header in an electronic text. Many metadata standards in use today, including the Dublin Core standard, do not prescribe either type of linkage, leaving the decision to each particular implementation.

Although the concept of metadata predates the Internet and the Web, worldwide interest in metadata standards and practices has exploded with the increase in electronic publishing and digital libraries, and the concomitant "information overload" resulting from vast quantities of undifferentiated digital data available online. Anyone who has attempted to find information online using one of today's popular Web search engines has likely experienced the frustration of retrieving hundreds, if not thousands, of "hits" with limited ability to refine or make a more precise search. The wide scale adoption of descriptive standards and practices for electronic resources will improve retrieval of relevant resources from the "Internet commons." As noted by Weibel and Lagoze, two leaders in the field of metadata development:

The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself." (Weibel and Lagoze, 1997)

The Dublin Core addresses this need for "standardized descriptive metadata."

## 2. What is the Dublin Core?

The Dublin Core metadata standard is a simple yet effective element set for describing a wide range of networked resources. The Dublin Core standard comprises fifteen elements, the semantics of which have been established through consensus by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship.

Dublin Core has as its goals the following characteristics:

\$ *Simplicity of creation and maintenance*

The Dublin Core element set has been kept as small and simple as possible to allow a non-specialist to create simple descriptive records for information resources easily and inexpensively, while providing for effective retrieval of those resources in the networked environment.

\$ *Commonly understood semantics*

Discovery of information across the vast commons of the Internet is hindered by differences in terminology and descriptive practices from one field of knowledge to the next. The Dublin Core can help the 'digital tourist' -- a non-specialist searcher -- find his or her way by supporting a common set of elements, the semantics of which are universally understood and supported. For example, scientists concerned with locating articles by a particular author, and art scholars interested in works by a particular artist, can agree on the importance of a "creator" element. Such convergence on a common, if slightly more generic, element set increases the visibility and accessibility of all resources, both within a given discipline and beyond.

\$ *International scope*

Although the specific linguistic challenges of the World Wide Web have not been directly addressed by the Dublin Core development community, the involvement of representatives from almost every continent has ensured that the development of the standard considers the multilingual and multicultural nature of the electronic information universe.

\$ *Extensibility*

While balancing the needs for simplicity in describing digital resources with the need for precise retrieval, Dublin Core developers have recognized the importance of providing a mechanism for extending the DC element set for additional resource discovery needs. It is expected that other communities of metadata experts will create and administer additional metadata sets. Metadata elements from these sets could be linked with Dublin Core metadata to meet the need for extensibility. This model allows different communities to use the DC elements for core descriptive information which will be usable across the Internet, while allowing domain specific additions which make sense within a more limited arena.

### 3. The purpose and scope of this guide

This document is intended to help users create descriptive records for electronic information resources. Creators of these records include catalogers, authors, editors, and web site administrators.

Another important goal of this document is to promote "best practices" for describing resources using the Dublin Core element set. 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.

### USE OF HTML FOR EMBEDDED METADATA

Dublin Core examples are in HTML, the Web's Hypertext Markup Language format and in a generic form (Element="value"). HTML provides an easily understood format for demonstrating Dublin Core's underlying concepts. It is important to note, however, that Dublin Core concepts are equally applicable to virtually any file format, as long as the metadata is in a form suitable for interpretation both by search engines and by human beings.

HTML has two tags that can be used to capture metadata. These are the "<META>" and "<LINK>" tags. When creating metadata that will be embedded in, or appear alongside, an actual document these tags must appear within the HEAD section of the HTML document. For example:

```
<HTML>
<HEAD>
<TITLE>Mating Habits of the Northern Hairy Nosed Wombat</TITLE>
<META NAME= "DC.Creator" CONTENT="Smythe, Pearl">
</HEAD>
<BODY>
<H1>Northern Hairy Nosed Wombats</H1>
<P>
The Northern Hairy Nosed Wombat is an animal native to Australia....</P>
</BODY>
</HTML>
```

Search engine programs understand that the metadata record starts after the "<HEAD>" line and ends before the "</HEAD>" line, and are thus able to extract metadata automatically. The metadata does not appear during normal document formatting or printing, and metadata-aware Web browsers may even be able to exploit it. Several popular current search engines now make use of the HTML <META> tag in Web documents. In HTML, each record element definition begins with "<META" and ends with ">". Within the META tag, two attribute/value pairs (as found in other HTML tags) are used to define the metadata. The first is NAME, the second, CONTENT. These two work together to define the

metadata within the META tag.

Below are some examples of how the META tag might be used in stand-alone and embedded metadata. Note that each metadata definition happens to fit on one line, but in general a definition can span several lines.

## 1. Metadata Contained in a Resource

This example is of a metadata record contained in a file alongside the document that it describes. The document is a short poem expressed in HTML.

```
<HTML>
<HEAD>
<TITLE>Song of the Open Road</TITLE>
<meta name="DC.Title" CONTENT="Song of the Open Road">
<meta name="DC.Creator" CONTENT="Nash, Ogden">
<meta name="DC.Type" CONTENT="e/document">
<meta name="DC.Date" CONTENT="1939">
<meta name="DC.Format" CONTENT="text/html">
<meta name="DC.Identifier" CONTENT="http://www.poetry.com/nash/open.html">
</HEAD>
<BODY><PRE>
I think that I shall never see
A billboard lovely as a tree.
Indeed, unless the billboards fall
I'll never see a tree at all.
</PRE></BODY>
</HTML>
```

## 2. Stand-Alone Metadata

Stand-alone metadata can exist in any kind of database. This example describes a photograph in another file that has a location given by a Uniform Resource Locator (URL). The entire record file looks like this:

```
<meta name="DC.Title" CONTENT="Kita Yama (Japan)">
<meta name="DC.Creator" CONTENT="Kertesz, Andre">
<meta name="DC.Date" CONTENT="1968">
<meta name="DC.Type" CONTENT="e/photograph">
<meta name="DC.Format" CONTENT="image/gif">
<meta name="DC.Identifier" CONTENT="http://foo.bar.zaf/kertesz/kyama">
```

## BASIC PRINCIPLES OF DESCRIPTIVE ELEMENTS

The notation (one of several) described in this guide is based on the HTML META tag. If non-ASCII characters are required, use the same conventions as in the body of the document. For example:

```
<meta name="DC.Title" CONTENT="Les biscuits &agrave; la banane">
```

Each element is optional and repeatable. Metadata elements may appear in any order. The ordering of multiple occurrences of the same element (e.g., Creator) may have a significance intended by the provider, but ordering is not guaranteed to be preserved in every user environment.

### **1. Element Parts and Syntax**

As demonstrated above, each descriptive element definition has a NAME attribute and a CONTENT attribute, as in:

```
<meta name="DC.Creator" CONTENT="Browning, Elizabeth">
```

Any metadata element may be omitted or repeated. When repeating elements, it is recommended best practice to list each element definition separately, as in:

```
<meta name="DC.Creator" CONTENT="Marx, Karl">  
<meta name="DC.Creator" CONTENT="Engels, Friedrich">
```

However, it is also valid to express repeated elements using a single NAME attribute with multiple semi-colon delimited values for the CONTENT attribute, as in:

```
<meta name="DC.Creator" CONTENT="Marx, Karl ; Engels, Friedrich">
```

AA Proposed Convention for Embedding Metadata in HTML<sup>®</sup> agreed upon a convention for identifying and grouping metadata schemes in HTML. This convention relies on the use of a prefix to indicate that the elements used are from Dublin Core or another metadata scheme. For increased readability the prefix "DC" should be written in upper case letters and element names should be capitalized. For example:

```
meta name="DC.Title"  
meta name="DC.Creator"
```

## NOT

DC.CREATOR or dc.CREATOR or DC.creator

## 2. Element Content and Controlled Vocabularies

Content data for some elements may be selected from a "controlled vocabulary," which is a limited set of consistently used and carefully defined terms. This can dramatically improve search results because computers are good at matching words character by character but weak at understanding the way people refer to one concept using different words, i.e. synonyms. Without basic terminology control, inconsistent or incorrect metadata can profoundly degrade the quality of search results. For example, without a controlled vocabulary, "candy" and "sweet" might be used to refer to the same concept. Controlled vocabularies may also reduce the likelihood of spelling errors when recording metadata. The Foundations Project is developing an interactive controlled-vocabulary thesaurus to enhance search results.

## THE CORE ELEMENTS

This section lists each Core element by its full name and label. For each element there is a reference description (taken from the RFC) and there are guidelines to assist in creating metadata content, whether it is done "from scratch" or by converting an existing record in another format.

The elements are listed in the order they were developed, but there are other useful ways to group them. In the following table, you can see that some elements relate to the content of the item, some to the item as intellectual property, still others to the particular instantiation, or version, of the item.

<u>Content</u>	<u>Intellectual Property</u>	<u>Instantiation</u>
Coverage	Contributor	Date
Description	Creator	Format
Type	Publisher	Identifier
Relation	Rights	Language
Source		
Subject		
Title		

In the element descriptions below, a formal single-word label is specified to make the syntactic specification of elements simpler for encoding schemes. Although some environments, such as HTML, are not case-sensitive, it is recommended best practice always to adhere to the case conventions in the element names given below to avoid conflicts in the event that the metadata is subsequently converted to a case-sensitive environment, such as XML/RDF. Some information may appear to belong in more

than one metadata element. While there will normally be a clear preferred choice, there is potential semantic overlap between some elements. Consequently, there will occasionally be some judgment required from the person assigning the metadata.

Minimum DC elements to be applied: Title, Subject(s), Description, Date(s) (creation, modified)

Highly desirable elements: Creator, Publisher, Type, Format, Language

## PUNCTUATION

Avoid extraneous punctuation, or ending punctuation in all elements except Description, where a closing period is used. Quote marks within elements, such as Description, must be omitted to avoid confusion with quote marks used around DC elements. Personal names are entered last name first. Corporate names follow Anglo-American Cataloging Rules, 2<sup>nd</sup> edition (AACR2). See Library of Congress Name Authority File for correct form. State names are spelled out; department is abbreviated as Dept.

*Example:*

```
<meta name="DC.Publisher.CorporateName" scheme="AACR2" content="Minnesota Dept. of Natural Resources. Division of Forestry">
```

Punctuation for subject headings subheads is space-dash-dash-space.

*Example:*

```
<meta name="DC.Subject" scheme="LIV-MN" content="Fruit -- Minnesota -- Directories">
```

Order of subheads is based on Legislative Indexing Vocabulary (LIV) and Library of Congress (LC) rules, in that place follows the subject heading, and form follows place. See LIV for standard subdivisions: <http://thomas.loc.gov/liv/subdiv.html>

## CAPITALIZATION

Capitalize Titles and Subject terms according to AACR2 format; that is, capitalize first word and proper names only. Do not alter general meta title tag; the above rule is for Dublin Core titles. Capitalize Description element according to normal rules of writing. If copy and paste has been used, correct capitalization. No elements should be left in all caps except acronyms.

## TITLE

**Element Description:** The name given to the resource by the Creator or Publisher.

### **Guidelines for creation of content:**

If in doubt about what constitutes the title, repeat the Title element and include the variants in second and subsequent Title iterations. If the item is in HTML, view the source document and make sure that the title identified in the title header is also include as a meta title (unless the DC metadata element is to be embedded in the document itself). First title comes from meta title, viewed in source codes, also at top of screen in Netscape. See **Punctuation**, above.

## TITLE - continued

Second title comes from the web page itself; subsequent titles, if necessary, are expansions or portions of previous titles. Skip beginning articles.

### Examples:

```
<meta name="DC.Title" CONTENT="Pilot's Guide to Aircraft Insurance">
```

```
<meta name="DC.Title" CONTENT="Sound of Music">
```

```
<meta name="DC.Title" CONTENT="Green on Greens">
```

```
<meta name="DC.Title" CONTENT="AOPA's Tips on Buying Used Aircraft">
```

### Schemes:

#### Internal

Not part of an external coding system.

#### AACR2

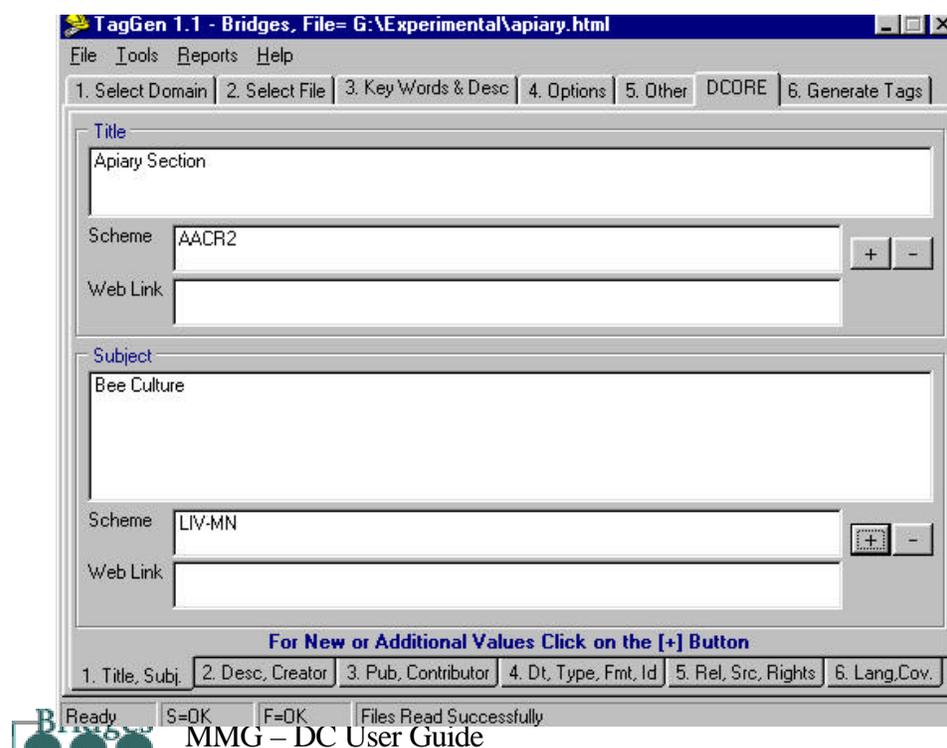
Title is devised according to the Anglo-American Cataloging Rules, 2<sup>nd</sup> edition.

This is really only useful for uniform titles, where title derives from another form of the item.

### Modifiers: none

**Web links:** pointer to web site of title authority record, an online record containing the correct, or uniform, title according to AACR2 as determined by the Library of Congress (LC) or LC-authorized library, in case differing versions exist.

Screen shot of TagGen 1.1 Metadata Generator for DC.Title and DC.Subject



## SUBJECT and KEYWORDS

**Element Description:** The topic of the resource. Typically, subject will be expressed as keywords or phrases that describe the subject or content of the resource. Recommended best practice is to select a value from a controlled vocabulary.

### **Guidelines for creation of content:**

Select subject keywords from either the Title or Description information. If the subject of the item is a person or an organization, use the same form of the name as you would if the person or organization were a Creator, but do not repeat the name in the Creator element.

If using keywords, choose the most significant and unique terms, avoiding those too general to describe a particular item. This element might well include classification data (for example, Library of Congress Classification Numbers or Dewey Decimal numbers) or controlled vocabularies (such as Medical Subject Headings or Art and Architecture Thesaurus descriptors) as well.

For meta keywords, add 1-2 precise subjects from LIV-MN, no subheads unless it's nonsense without them, in which case add one level only, never repeating same subhead for next meta keyword.

*Example:*

```
<meta name="keywords" content="Foundations Project (Minn.), Environmental research">
```

For DC subject element, keep the 1-2 precise subjects from LIV-MN and expand with subheads as necessary.. **Rarely use keywords** . If LIV-MN is insufficient, recommend term to thesaurus editor via e-mail, including URL for page being cataloged. See also **punctuation**, above.

*Examples:*

```
<LINK REL=SCHEMA.DC HREF="http://www.bridges.state.mn.us/servlet/lexico">
```

```
<meta name="DC.Subject" scheme="AACR2" content="Foundations Project (Minn.)">
```

```
<LINK REL=SCHEMA.DC HREF="http://www.bridges.state.mn.us/servlet/lexico">
```

```
<meta name="DC.Subject" scheme="LIV-MN" content="Environmental research -- Minnesota">
```

### **Schemes:**

Keyword

Not part of an external coding system

AACR2

Corporate subject headings created according to Anglo-American Cataloging Rules, 2<sup>nd</sup> edition

AAT

Art and Architecture Thesaurus  
LCC  
Library of Congress Classification  
LCSH  
Library of Congress Subject Headings  
**LIV-MN**  
**Legislative Indexing Vocabulary - Minnesota edition**  
MESH  
National Library of Medicine Subject Headings

**Modifiers:** none

**Web links:** pointer to subject authority

For LIV-MN the web link is:

REL=SCHEMA.DC HREF="<http://www.bridges.state.mn.us/servlet/lexico>"

*Examples:*

```
<meta name="DC.Subject" CONTENT="Aircraft leasing and renting">
```

```
<meta name="DC.Subject" CONTENT="Olympic skiing"
```

```
<meta name="DC.Subject" scheme="LIV-MN" content="Agronomy -- Minnesota">
```

Screen shot of TagGen 1.1 Metadata Generator for DC.Title and DC.Subject available on page 9.

## DESCRIPTION

**Element Description:** A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.

### **Guidelines for creation of content:**

Since the description field is a potentially rich source of indexable vocabulary, care should be taken to provide this element when possible. Some metadata collections could include content description (spectral analysis of a visual resource, for example) that current network systems may not be able to embed. In such a case, this field might contain a link to such a description rather than the description itself.

Descriptive information can be taken from the item itself, if there is no abstract or other structured description available. Normally, if a Description cannot be found either in the introductory or front matter, or the first few paragraphs, it should be set up "on the fly" by the metadata provider. Normally, Description should be limited to 1-2 concise sentences. Include

## DESCRIPTION – continued

important keywords (not necessarily terms from thesaurus). Look at the title of the page as well as the entire page to get the sense of its purpose.

**Schemes:** none

**Modifiers :**

Abstract – use only if page provides a formal abstract or something that fills that role  
Table of Contents

**Web links:** pointer to a more detailed external description

*Example:*

```
<meta name="DC.Description" CONTENT="Illustrated guide to airport markings and lighting signals, with particular reference to SMGCS (Surface Movement Guidance and Control System) for airports with low visibility conditions.">
```

The screenshot shows the TagGen 1.1 application window. The title bar reads "TagGen 1.1 - Bridges, File= G:\Experimental\apiary.html". The menu bar includes "File", "Tools", "Reports", and "Help". Below the menu bar is a navigation bar with tabs: "1. Select Domain", "2. Select File", "3. Key Words & Desc", "4. Options", "5. Other", "DCORE", and "6. Generate Tags". The main area is divided into two sections: "Description" and "Creator".

**Description Section:**

- Text area: "A textual description of the content of the resource, including abstracts in the case of document-like objects o"
- Modifier: "Abstract" (with "+" and "-" buttons)
- Web Link: (empty text box)

**Creator Section:**

- Text area: "The person or organization primarily responsible for creating the intellectual content of the resource."
- Modifier: "Corporate.name" (with "+" and "-" buttons)
- Scheme: "A\ACR2" (with "+" and "-" buttons)
- Web Link: (empty text box)

Below the sections is a blue banner that reads: "For New or Additional Values Click on the [+] Button". At the bottom, there is a status bar with tabs: "1. Title, Subj.", "2. Desc, Creator", "3. Pub, Contributor", "4. Dt, Type, Fmt, Id", "5. Rel, Src, Rights", and "6. Lang, Cov.". The status bar also shows "Ready", "S=OK", "F=OK", and "Files Read Successfully".

Screen Shot of Description/Creator Pages of TagGen

## CREATOR

**Element Description:** The person or organization primarily responsible for creating the intellectual content of the resource. For example, authors in the case of written documents, artists, photographers, or illustrators in the case of visual resources.

### **Guidelines for creation of content:**

Creators should be listed separately in the same order that they appear in the publication. Personal names should be listed surname or family name first, followed by forename or given name. When in doubt, give the name as it appears, and do not invert. Follow Library of Congress interpretation of AACR2 as identified in their authority records for form of corporate entity. This includes spelling out state name and abbreviating Dept.

### **Schemes:**

Internal

The information provided is not part of an external coding system

AACR2

Formatting according to the rules of Anglo-American Cataloging Rules, 2<sup>nd</sup> edition

### **Modifiers:**

Corporate name

Corporate entity which created the resource

Personal name

Person who created the resource

**Web links:** pointer to authority record for creator

### *Examples:*

```
<meta name="DC.Creator" CONTENT="Duncan, Phyllis-Anne">
```

```
<meta name="DC.Creator" CONTENT=" Melendez Santiago, Maria Luz ">
```

```
<meta name="DC.Creator" CONTENT=" Maimonides ">
```

BUT:

```
<meta name="DC.Creator" CONTENT="Park Sung Hee">
```

In the case of organizations where there is clearly a hierarchy present, list the parts of the hierarchy from largest to smallest, separated by full stops.

## CREATOR - continued

*Example:*

```
<meta name="DC.Creator" CONTENT="United States. Internal Revenue Service"
```

```
<meta name="DC.Creator" CONTENT="Federal Aviation Administration. Aviation Safety Program.">
```

NOT:

```
<meta name="DC.Creator" CONTENT="Aviation Safety Program of the Federal Aviation Administration">
```

If it is not clear whether there is a hierarchy present, or unclear which is the larger or smaller portion of the body, give the name as it appears in the item.

```
<meta name="DC.Creator" CONTENT="Art Institute of Chicago">
```

```
<meta name="DC.Creator" CONTENT="Association of the Bar of the City of New York"
```

```
<meta name="DC.Creator" CONTENT="Baltimore County Medical Society"
```

If the nature of the responsibility is ambiguous, the recommended practice is to use Publisher for organizations, and Creator for individuals. In cases of lesser responsibility, other than creation, use Contributor.

Screen shot of TagGen 1.1 Metadata Generator for DC.Description and DC.Creator available on page 12.

## PUBLISHER

**Element Description:** The entity responsible for making the resource available in its present form, such as a publishing house, a university department, or a corporate entity.

### **Guidelines for content creation:**

The intent of specifying this field is to identify the entity that provides access to the resource. If the Creator and Publisher are the same, do not repeat the name in the Publisher area. If the nature of the responsibility is ambiguous, the recommended practice is to use Publisher for organizations, and Creator for individuals. In cases of lesser responsibility, other than creation, use Contributor. Follow Library of Congress interpretation of AACR2 as identified in their authority records for form of corporate entity. This includes spelling out state name and abbreviating Dept.

### **Schemes:**

Internal

The information provided is not part of an external coding system

AACR2

Formatting according to the rules of Anglo-American Cataloging Rules, 2<sup>nd</sup> edition

### **Modifiers :**

Corporate name

Corporate entity which created the resource

Personal name

Person who created the resource

**Schemes:** pointer to authority record for creator

### *Examples:*

```
<meta name="DC.Publisher" CONTENT="Moguls Anonymous">
```

```
<meta name="DC.Publisher" CONTENT="University of Miami. Dept. of Economics"
```

```
<meta name="DC.Publisher" CONTENT="Minnesota Dept. of Natural Resources"
```

Screen shot of TagGen 1.1 Metadata Generator for DC.Publisher and DC.Contributor available on page 16.

Screen Shot of Publisher/Contributor Pages of TagGen

## CONTRIBUTOR

**Element Description:** A person or organization not specified in a Creator element who has made significant intellectual contributions to the resource but whose contribution is secondary to any person or organization specified in the Creator element (for example, editor, transcriber, and illustrator).

### **Guideline for content creation:**

The same general guidelines for using names of persons or organizations as Creators apply here. Follow Library of Congress interpretation of AACR2 as identified in their authority records for form of corporate entity. This includes spelling out state name and abbreviating Dept.

### **Schemes:**

Internal

The information provided is not part of an external coding system

AACR2

Formatting according to the rules of Anglo-American Cataloging Rules, 2<sup>nd</sup> ed.

## CONTRIBUTOR - continued

**Modifiers:** see TagGen D.C. tables for full listing

Page designer

Illustrator

**Web links:** pointer to authority record

Screen shot of TagGen 1.1 Metadata Generator for DC.Publisher and DC.Contributor available on page 16.

## DATE

**Element Description:** A date associated with the creation or availability of the resource. Such a date is not to be confused with one belonging to the Coverage element, which would be associated with the resource only insofar as the intellectual content is somehow about that date. Recommended best practice is defined in a profile of ISO 8601 [Date and Time Formats (based on ISO 8601), W3C Technical Note <http://www.w3.org/TR/NOTE-datetime>] that includes (among others) dates of the forms YYYY and YYYY-MM-DD. In this scheme, the date 1994-11-05 corresponds to November 5, 1994.

### **Guidelines for content creation:**

Dates on the page have top priority. If there is both a creation date and last modified date, put both on the record, with appropriate modifiers. If there is only one of these dates, use it only.

*Examples* (shown in html, but chosen from TagGen pull-down menus):

```
<meta name="DC.Date.Creation" scheme="ISO 8601" content="1997-11-20">
```

```
<meta name="DC.Date.Modified" scheme="ISO 8601" content="1998-06-10">
```

If there is a copyright date of a year only, use the Creation modifier with double zeroes for the month and day.

*Example:*

```
<meta name="DC.Date.Creation" scheme="ISO 8601" content="1998-00-00">
```

Use the Current modifier only if there are no dates on the page, putting today's date there. *Example:*

```
<meta name="DC.Date.Current" scheme="ISO 8601" content="1998-12-07">
```

**Schemes:** see TagGen D.C. tables for full listing

ISO 8601 (yyyy-mm-dd) is the preferred MMG-DC form

**Modifiers:** see TagGen D.C. tables for full listing

Creation

Current

Modified

## DATE - continued

**Web Links:** none

Screen shot of TagGen 1.1 Metadata Generator for DC.Date, DC.Type, DC.Format and DC.Identifier available on page 19.

## RESOURCE TYPE

**Element Description:** The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. For the sake of interoperability, type should be selected from an enumerated list that is under development in the workshop series.

### **Guidelines for content creation:**

This element should describe the genre of the content of the resource.

#### **text**

resources in which the content is mainly words for reading; for example: books, letters, dissertations, poems, newspapers, archives of mailing lists

#### **image**

the content is primarily visual in two dimensions and is not text; for example: images, paintings, animations, diagrams

#### **sound**

the content is primarily audio; for example: music, speech, recorded sounds

#### **data**

information encoded in lists, tables, databases, etc., which will often be in a format ready for direct machine processing; for example: spreadsheets, databases, GIS data

See TagGen pull-down menu for complete listing.

*Examples:*

```
<meta name="DC.Type" CONTENT="image">
```

```
Type="sound"
```

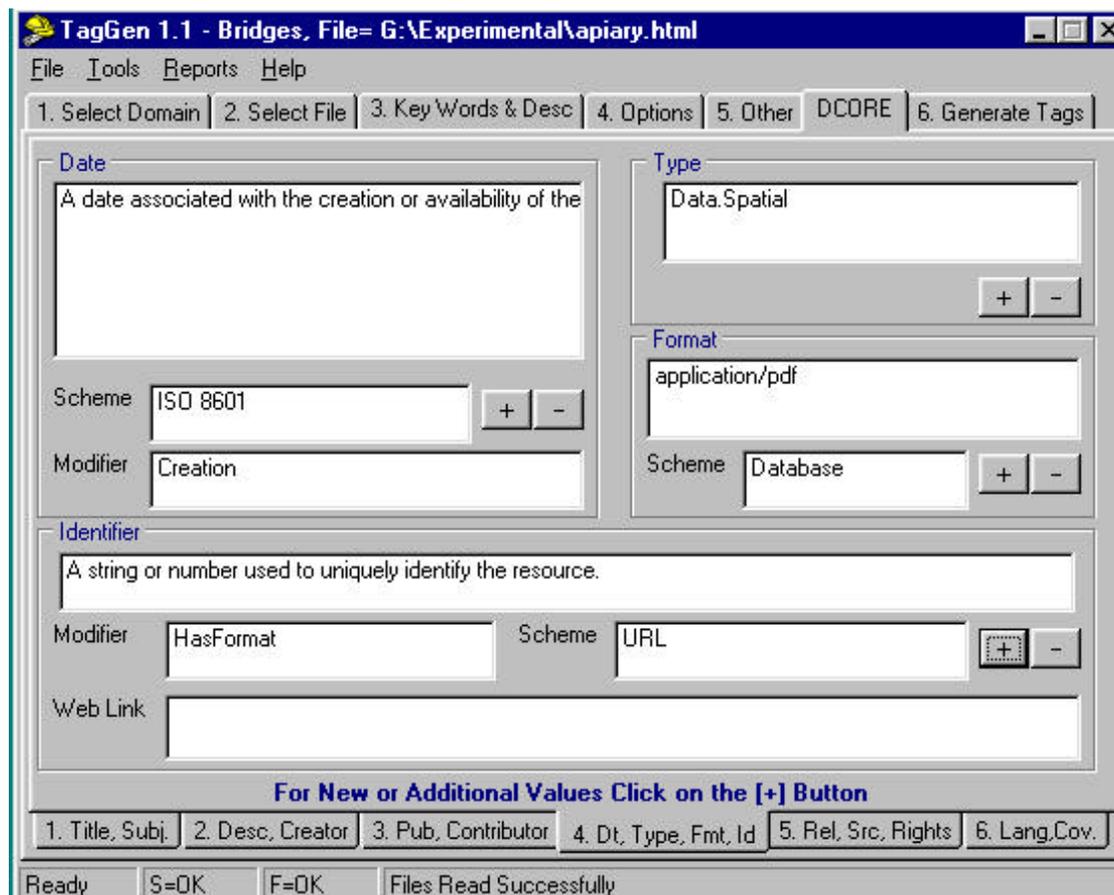
```
Type="text"
```

**Schemes:** none

**Modifiers:** none

**Web links:** none

Screen shot of TagGen 1.1 Metadata Generator for DC.Date, DC.Type, DC.Format and DC.Identifier available on page 19.



Screen Shot of Date/Type/Format/Identifier Pages of TagGen

## FORMAT

**Element Description:** The data format of the resource, used to identify the software and possibly hardware that might be needed to display or operate the resource. For the sake of interoperability, Format should be selected from an enumerated list that is currently under development in the workshop series.

### **Guidelines for content creation:**

Formats, such as text/html, ASCII, Postscript file, executable application, or JPEG image may be included in this area. Assign a Format from Internet Media Types (MIME types). In principle, formats can include physical media such as books, serials, or other non-electronic media.

## FORMAT - continued

Information concerning the size of a resource may be included in the content of the Format element. In resource discovery this might be used as a criterion to select resources of interest, since a user may need to evaluate whether they can make use of the resource within the infrastructure available to them.

**Schemes:** see TagGen D.C. tables for full listing

Freetext

HTML

MIME

PDF

**Modifiers:** none

**Web links:** none

*Examples:*

```
<meta name="DC.Format" CONTENT="image/gif">  
Title="Dublin Core icon"  
Identifier="http://purl.org/metadata/dublin_core/images/dc2.gif"  
Type="image"  
Format="image/gif 4kB"
```

```
<meta name="DC.Subject" CONTENT="Saturn">  
<meta name="DC.Type" CONTENT="image">  
<meta name="DC.Format" CONTENT="image/gif 640 x 512 pixels">  
<meta name="DC.Identifier" CONTENT="http://www.not.iac.es/newwww/photos/images/satnot.gif">
```

```
Title="The Bronco Buster"  
Creator="Frederic Remington"  
Type="physical object"  
Format="bronze 22 in."
```

Screen shot of TagGen 1.1 Metadata Generator for DC.Date, DC.Type, DC.Format and DC.Identifier available on page 19.

## IDENTIFIER

**Element Description:** A string or number used to uniquely identify the resource. Examples for networked resources include URLs and URNs (when implemented). Other globally-unique identifiers, such as International Standard Book Numbers (ISBN) or other formal names are also candidates for this element.

**Schemes:**

ISBN  
ISSN  
PURL  
URL  
URN

**Modifiers :** none

**Web links:** none

Screen shot of TagGen 1.1 Metadata Generator for DC.Date, DC.Type, DC.Format and DC.Identifier available on page 19.

## RELATION

**Element Description:** An identifier of a second resource and its relationship to the present resource. This element permits links between related resources and resource descriptions to be indicated. Examples include an edition of a work (IsVersionOf), a translation of a work (IsBasedOn), a chapter of a book (IsPartOf), and a mechanical transformation of a dataset into an image (IsFormatOf). For the sake of interoperability, relationships should be selected from an enumerated list that is currently under development in the workshop series.

**Schemes:**

ISBN  
ISSN  
FPI

**Modifiers :** see TagGen D.C. tables for full listing

IsPartOf	HasPart
IsVersionOf	HasVersion
IsFormatOf	HasFormat
References	IsReferencedBy
IsBasedOn	IsBasisFor
Requires	IsRequiredBy

**Web links:** pointer to the resource.

Screen shot of TagGen 1.1 Metadata Generator for DC.Relation, DC.Source and DC.Rights available on page 23.

## SOURCE

**Element Description:** Information about a second resource from which the present resource is derived. While it is generally recommended that elements contain information about the present resource only, this element may contain a date, creator, format, identifier, or other metadata for the second resource when it is considered important for the discovery of the present resource: recommended best practice is to use the Relation element instead. For example, it is possible to use a Source date of 1603 in a description of a 1996 film adaptation of a Shakespeare play, but it is preferred instead to use Relation "IsBasedOn" with a reference to a separate resource whose description contains a Date of 1603. Source is not applicable if the present resource is in its original form.

### **Guidelines for content creation:**

In general, include in this area information which does not fit easily into Relation.

### **Schemes:**

FPI  
ISSN  
SICI  
Version

**Modifiers:** none

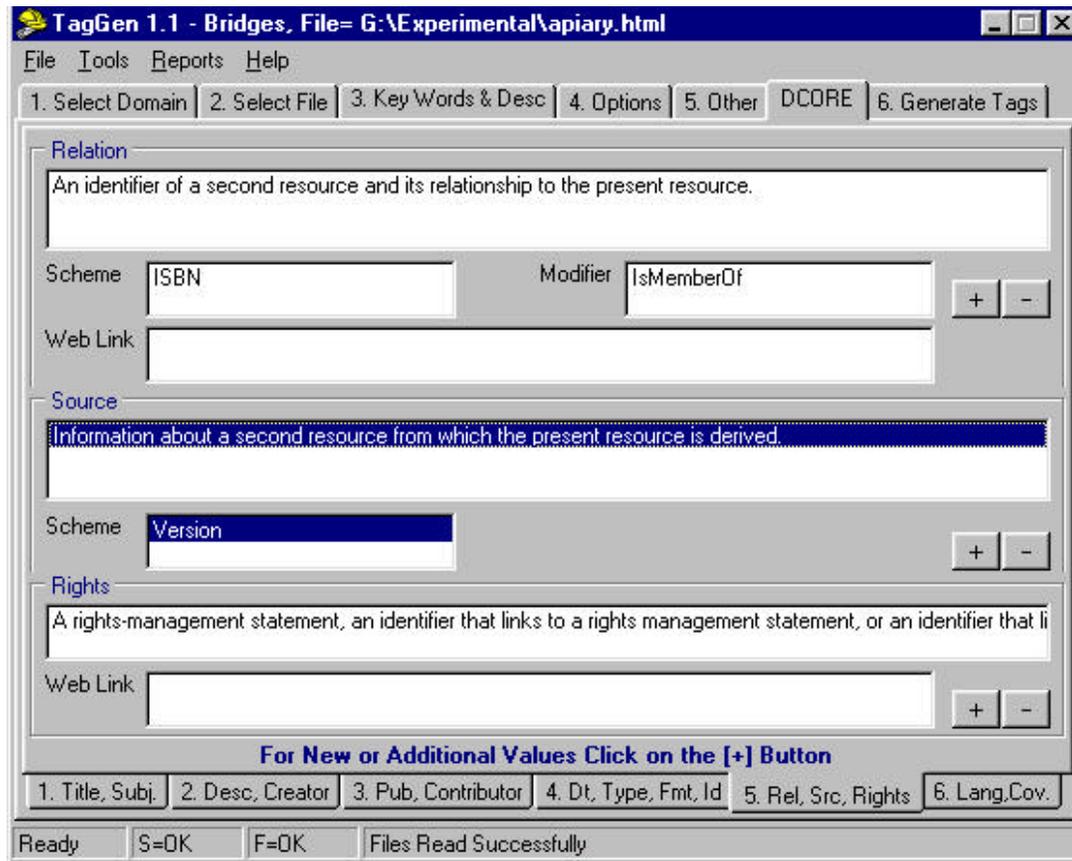
**Web links:** none

### *Example:*

```
<meta name="DC.Source" CONTENT="RC607.A26W574 1996">
```

[where "RC607.A26W574 1996" is the call number of the print version of the resource, from which the present version was scanned]

Screen shot of TagGen 1.1 Metadata Generator for DC.Relation, DC.Source and DC.Rights available on page 23.



Screen Shot of Relation/Source/Rights Pages of TagGen

## RIGHTS MANAGEMENT

**Element Description:** A rights-management statement, an identifier that links to a rights management statement, or an identifier that links to a service providing information about rights management for the resource.

**Guidelines for content creation:**

At present, used only for a URL.

```
<meta name="DC.Rights" CONTENT= "http://cs-tr.cs.cornell.edu/Dienst/Repository/2.0/Terms">
```

**Schemes:** none

**Modifiers:** none

**Web links:** pointer to the rights management scheme

## LANGUAGE

**Element Description:** The language of the intellectual content of the resource. Where practical, the content of this field should coincide with RFC 1766 [Tags for the Identification of Languages, <http://ds.internic.net/rfc/rfc1766.txt>]; examples include en, de, es, fi, fr, ja, th, and zh.

### **Guidelines for content creation:**

Coded or textual information can be represented here. If the content is in more than one language, the element may be repeated.

**Schemes:** see TagGen D.C. tables for full listing

ISO 639-1 is the preferred scheme; English is "en"

**Modifiers :** none

**Web links:** none

*Examples:*

Language=en

Language=fr

OR,

```
<meta name="DC.Language" CONTENT="en;fr">
```

OR,

```
<meta name="DC.Language" CONTENT="Primarily English, with some abstracts also in French.">
```

```
<meta name="DC.Language" CONTENT="en-US">
```

Screen shot of TagGen 1.1 Metadata Generator for DC.Language and DC.Coverage available on page 26.

## COVERAGE

**Element Description:** The spatial or temporal characteristics of the intellectual content of the resource. Spatial coverage refers to a physical region (e.g., celestial sector); use coordinates (e.g. longitude and latitude) or place names that are from a controlled list or are fully spelled out. Temporal coverage refers to what the resource is about rather than when it was created or made available (the latter belonging in the Date element); use the same date/time format (often a range) [Date and Time Formats (based on ISO8601), W3C Technical Note, <http://www.w3.org/TR/NOTE-datetime>] as recommended for the Date element or time periods that are from a controlled list or are fully spelled out.

### **Guidelines for content creation:**

Whether this element is used for spatial or temporal information, care should be taken to provide consistent information that can be interpreted by users. For most simple applications, where place names or coverage dates might be useful, whether the information is numeric or alphabetical may be enough to differentiate. For more complex applications, consideration should be given to additional qualification.

### **Schemes:**

ANSI.X3.30-1985

A date/time range in a format specified in ANSI X3.30-1985 standard. Must be used in conjunction with the modifier "temporal"

LatLong

Latitude / Longitude coordinates for the coverage of the resource. Must be used in conjunction with the modifier 'spatial'

OSGB

Specifies an Ordinance Survey National Grid Reference

### **Modifiers :**

Spatial

Temporal

**Web links:** none

### *Examples:*

Coverage=1995-1996

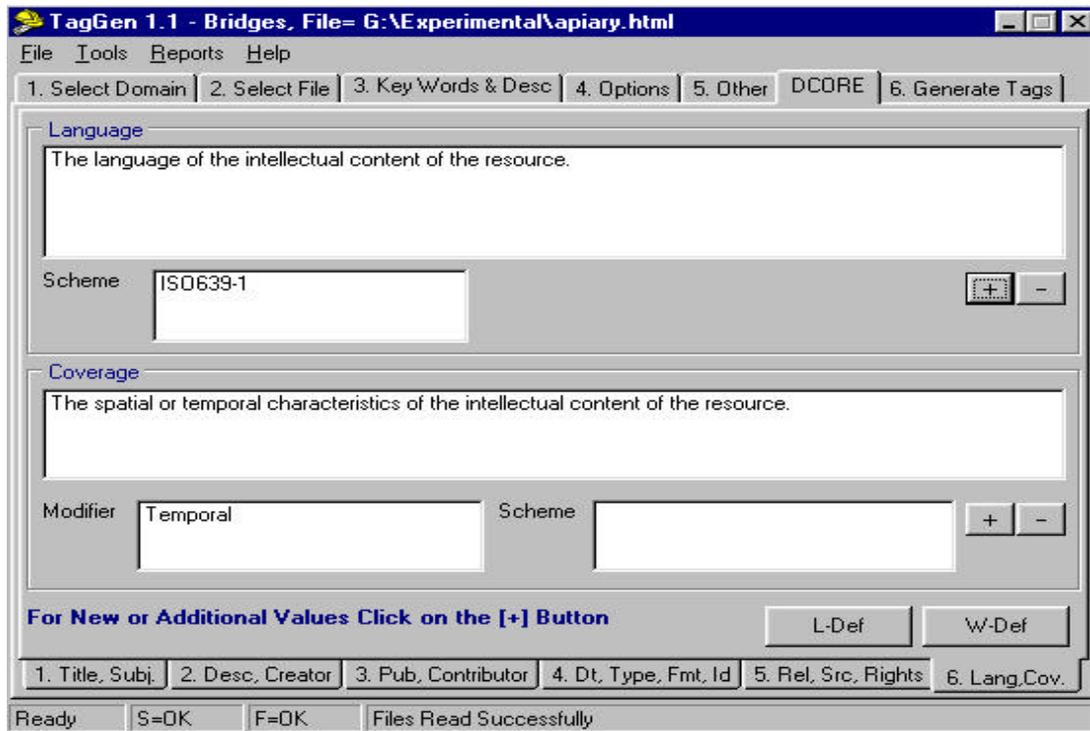
Coverage=Boston, MA

OR,

<meta name="DC.Coverage" CONTENT="17th century">

<meta name="DC.Coverage" CONTENT="Upstate New York">

Screen shot of TagGen 1.1 Metadata Generator for DC.Language and DC.Coverage available on page 26.



Screen Shot of Language/Coverage Pages of TagGen

# Dublin Core Glossary

Adapted from the *Glossary for A User Guide for Simple Dublin Core: Draft*, a collaborative effort between Gail Clement & Pete Winn.

For complete information on qualified Dublin Core elements, see the *Minnesota Metadata Guidelines – Dublin Core: User Guide for Dublin Core Elements*, located at: <http://bridges.state.mn.us/MMG-DCUserGuide.PDF>

## **Anglo-American Cataloguing Rules (AACR2)**

The dominant bibliographic standard regulating descriptive cataloging in the English-speaking world. AACR2 represents a set of rules for the standard description of all materials which a library holds or to which it has access.

## **American Standard Code for Information Interchange (ASCII)**

A scheme that provides standard numeric values to represent letters, numbers, punctuation marks, and other characters. The use of standard values allows computers and computer programs to exchange data.

## **Author**

(Creator) The Dublin Core element used to designate the person(s) or organization(s) primarily responsible for the intellectual content.

## **Best practice**

The set of processes and function designs that best fit a given set of circumstances.

## **Case-sensitive**

Lower and upper case letters are not treated as being the same; e.g. 'a' is not the same as 'A.'

## **Controlled vocabulary**

A prescribed set of consistently used and carefully defined terms.

## **Contributor**

The Dublin Core element used to designate Person(s) or organization(s) in addition to those specified in the CREATOR element who have made significant intellectual contributions to the resource.

**Coverage**

The Dublin Core element used to designate spatial and/or temporal characteristics of the intellectual content of the resource.

**Creator**

(Author) The Dublin Core element used to designate the person(s) or organization(s) primarily responsible for intellectual content.

**Crosswalk**

A table that maps the relationships and equivalencies between two or more metadata formats.

**Date**

The Dublin Core element used to designate the date the resource was made available in its present form.

**Description**

The Dublin Core element used to designate a textual description of the content of the resource.

**Digital tourist**

An inexperienced computer user; in the context of resource discovery, an inexperienced searcher.

**Discovery software**

A computer application designed to simplify, assist and expedite the process of finding information resources.

**Document Type Definition (DTD)**

A description the components of a specific document or class of documents. A DTD description includes:

- The containers or elements that make up the document, e.g. paragraphs, headings, list items, figures, tables etc.
- The logical structure of the document, e.g. chapters containing sections etc.
- Additional information associated with elements (known as attributes), e.g. identifiers, date, stamps etc.

**Document-like object**

An entity the resembles a document from the standpoint that it is substantially text-based and shares other properties of a document; e.g.: electronic mail messages or spreadsheets.

DTD. *See* **Document Type Definition**

**Dublin Core**

The Dublin Core is a 15-element metadata element set intended to facilitate discovery of

electronic resources. The Dublin Core has been in development since 1995 through a series of focussed invitational workshops that gather experts from the library world, the networking and digital library research communities, and a variety of content specialties.

### **Electronic information resource**

An information resource that is maintained in electronic, or computerized format, and may be accessed, searched and retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM)

### **Embedded metadata**

Metadata that is created, maintained and stored within the object it describes; the opposite of stand-alone metadata.

### **Encoding scheme**

A pre-defined way for converting information into code or machine intelligible language.

### **Extensible**

Having the potential to be expanded or stretched in scope, area or size. In the case of Dublin Core, the ability to extend a core set of metadata with additional elements.

### **Extensible Markup Language (XML)**

A subset of Standard Generalized Markup Language (SGML), a widely used international text processing standard. XML is being designed to bring the power and flexibility of generic SGML to the Web, while maintaining interoperability with full SGML and HTML. For more information, *see* <http://www.w3.org/XML/>.

### **Format**

The Dublin Core element used to designate the data representation of the resource.

GIF. *See* **Graphics Interchange Format**

### **Government Information Locator Service (GILS)**

GILS embraces open standards to implement interoperable searching across diverse, decentralized information 'locators' to return references to all kinds of electronic and non-electronic information resources. Locators are implemented as common semantics for characterizing information resources, i.e. common metadata semantics.

### **Graphics Interchange Format (GIF)**

The dominant graphics format on the Web.

### **Granularity**

The degree to which something is composed of small pieces.

**Hypertext Markup Language (HTML)**

The standard text-formatting language for documents on the World Wide Web. HTML text files contain content that is rendered on a computer screen and markup, or tags, which can be used to tell the computer how to format that content. HTML tags can also be used to encode metadata and to tell the computer how to respond to certain user actions, such as a mouse click. For more information, *see* <http://www.w3.org/MarkUp/>.

**Identifier**

The Dublin Core element used to designate a string or number that uniquely identifies the resource.

IMT. *See* **Internet Media Type**

**Indexing program**

Computer software used to order things; frequently used to refer to software which alphabetizes some or all of the terms in one or more electronic documents.

**Information resource**

Any entity, electronic or otherwise, capable of conveying or supporting intelligence or knowledge; e.g. a book, a letter, a picture, a sculpture, a database, a person.

**Instantiation**

An identifiable occurrence or occasion of something; in the case of Dublin Core, a specific occurrence of an information resource.

**Internet Media Type (IMT)**

A set of terms that describe types of resources on the Internet.

**Joint Photographic Experts Group (JPEG)**

A standard for compressing digital images. The advantage of JPEG is that it uses compression to make graphics files smaller, making them faster to transfer and view over the World Wide Web. The disadvantage is some loss of image quality due to data loss during compression.

Keywords. *See* **Subject**

**Language**

The Dublin Core element used to designate the Language(s) of the intellectual content of the resource.

**MARC (USMARC)**

National standard for library holdings' database descriptions. Stores a given holding's metadata in one record with a flat structure. This record comprises four components: a leader, a record directory,

control files and variable fields. An implementation of ANSI/NISO Z39.2 the American National Standard for Bibliographic Information interchange. The USMARC format documents the designations and content designators for the fields that are to be carried in records structured according to Z39.2.

### **META tag**

The process of applying metadata to an information resource; the HTML element used to demarcate metadata: <META>, </META>.

### **Metadata**

Descriptive information about an information resource. In the case of Dublin Core, information that expresses the knowledge content, intellectual property and/or instantiation characteristics of an information resource.

### **Metadata record**

A syntactically correct representation of the descriptive information (metadata) for an information resource. In the case of Dublin Core, a representation of the Dublin Core elements that has been defined for the resource. The majority of metadata records and record fragments in this document are presented in HTML syntax.

### **Metadata registry**

A publicly accessible system that records the semantics, structure and interchange formats of any type of metadata. A formal authority, or agency, maintains and manages the development and evolution of a metadata registry. The authority is responsible for policies pertaining to registry contents and operation.

### **Multipurpose Internet Mail Extensions (MIME)**

The standard for attaching files to Internet e-mail messages. Attached files may be text, graphics, spreadsheets, documents, sound files, etc.

### **Networked resource**

An object which is available electronically via a network.

### **Online Computer Library Center (OCLC)**

The major source of cataloging data for libraries around the world; located in Dublin, Ohio, US.

### **Publisher**

The Dublin Core element used to designate the entity responsible for making the resource available in its present form.

### **Qualifier**

Something that describes or characterizes an object. In the case of Dublin Core, attributes

refine or characterize interpretation of an element's content.

RDF. *See* **Resource Description Framework**.

### **Relation**

The Dublin Core element used to designate the identifier of a second resource and its relationship to the first resource.

### **Request for Comment (RFC)**

A Request for Comment (RFC) is a note about the Internet. The note may discuss any aspect of computing and computer communication. All specification documents for the Internet are published as RFCs. For more information, *see* <http://www.isi.edu/rfc-editor/>.

### **Resource Description Framework (RDF)**

The basic language for writing metadata; a foundation which provides a robust flexible architecture for processing metadata on the Internet. RDF will retain the capability to exchange metadata between application communities, while allowing each community to define and use the metadata that best serves their needs. For more information *see* <http://www.w3.org/RDF/>.

### **Resource discovery**

The process through which one obtains knowledge of an information resource.

Resource Type. *See* **Type**.

Resource Description. *See* **Description**.

Resource Identifier. *See* **Identifier**

RFC. *See* **Request for Comment**

### **Rights**

The Dublin Core element used to provide a link to a copyright notice, to a rights-management statement, or to a service that would provide information about terms of access to the resource.

Rights Management. *See* **Rights**

### **ROADS**

An UK funded project whose aim is to develop discovery software for Internet resources.

### **Scheme**

A systematic, orderly design or combination of elements. In the case of the HTML META tag attribute, SCHEME is any recognized coding system used to interpret the meaning of an element.

**Search engine**

Utility capable of returning references to relevant information resources in response to a query.

**Semantics**

Significance or meaning. In the case of Dublin Core, the significance or intended meaning of individual metadata elements and their components.

SGML. *See* **Standard Generalized Markup Language**

**Software agent**

A computer program that carries out tasks on behalf of another entity. Frequently used to reference a program which searches the Internet for information meeting the specified requirements of an individual user.

**Source**

The Dublin Core element used to designate information about a second resource from which this resource is derived.

**Standard Generalized Markup Language (SGML)**

A non-proprietary language/enabling technology for describing information. Information in SGML is structured like a database, supporting rendering in and conversion between different formats. Both XML and later versions of HTML instances of SGML. For more information *see* <http://www.w3.org/SGML/>.

**Stand-alone metadata**

Metadata that is created, maintained and stored independently of the object it describes. The opposite of embedded metadata.

**Subject**

(Keywords) The Dublin Core element used to designate the topic of the resource, or keywords or phrases that describe the subject or content of the resource.

**Surrogate content**

Metadata as a substitute for an actual resource.

**Syntax**

The form and structure with which elements are combined. In the case of Dublin Core, the form and structure of how metadata elements and their components are combined to form a metadata record.

TEI. *See* **Text Encoding Initiative**

**Temporal**

Limited by or in regard to time.

### **Text Encoding Initiative (TEI)**

An international project to develop guidelines for the preparation and interchange of electronic texts for scholarly research as well as a broad range of other language industry uses. The TEI DTD is an SGML Document Type Definition for encoding literary works. For more information, see <http://www-tei.uic.edu/orgs/tei/info/teij16.html>.

### **Title**

The Dublin Core element used to designate the name given to the resource.

### **Type**

The Dublin Core element used to designate the category of the resource.

ULAN. See **Union Lists of Artists' Names**

### **Unicode**

A registered trademark of Unicode, Inc. Unicode refers to a universal encoding scheme designed to allow interchange, processing and display of the world's principal languages, as well as many historic and archaic scripts. Unicode supports and fosters a multilingual computing world community by allowing computers using one language to "talk" to computers using a different language.

### **Unicode Transformation Format, 8-bit (UTF-8)**

Unicode Transformation Format, 8-bit. UTF-8 is a temporary form of Unicode that is well suited for routing data through systems which are not designed for Unicode, such as some email servers and web clients. UTF-8 is an attractive way of storing multilingual data on the Internet, without requiring full Unicode compliance.

### **Uniform Resource Identifier (URI)**

The syntax for all names/addresses that refer to resources on the World Wide Web. For information about Internet addressing, see <http://www.w3.org/Addressing/Addressing.html>.

### **Uniform Resource Locator (URL)**

A technique for indicating the name and location of Internet resources. The URL specifies the name and type of the resource, as well as the computer, device and directory where the resource may be found. The URL for Dublin Core is [http://purl.oclc.org/metadata.dublin\\_core](http://purl.oclc.org/metadata.dublin_core). For information about Internet addressing, see <http://www.w3.org/Addressing/Addressing.html>.

### **Uniform Resource Name (URN)**

A URI (name and address of an object on the Internet) that has some assurance of persistence beyond that normally associated with an Internet domain or host name. For information about Internet addressing, see <http://www.w3.org/Addressing/Addressing.html>.

**Union Lists of Artists' Names (ULAN)**

Union Lists of Artists' Names (Getty Information Institute 1997).

URI. *See* **Uniform Resource Identifier**

URL. *See* **Uniform Resource Locator**

URN. *See* **Uniform Resource Name**

USMARC. *See* **MARC**

UTF-8. *See* **Unicode Transformation Format, 8-bit.**

**Verso**

A left-hand page.

**Warwick Framework**

An architecture for the interchange of metadata packages, or "containers"; designed to satisfying the need for competing, overlapping, and complementary metadata models. For more information, *see* <http://www.dlib.org/dlib/july96/07weibel.html>.

**World Wide Web (WWW)**

The panoply of Internet resources (text, graphics, audio, video, etc.) that are accessible via a web browser.

**World Wide Web Consortium (W3C)**

The W3C is an international industry consortium founded in October 1994 to lead the World Wide Web to its full potential by developing common protocols that promote its evolution and ensure its interoperability. For additional information *see* <http://www.w3.org/Consortium/>

XML. *See* **Extensible Markup Language.**

**Z39.50**

Transfer protocol for bibliographic information in a networked environment.