

Network Working Group
Request for Comments: 2413
Category: Informational

S. Weibel
OCLC Online Computer Library Center, Inc.
J. Kunze
University of California, San Francisco
C. Lagoze
Cornell University
M. Wolf
Reuters Limited
September 1998

Dublin Core Metadata for Resource Discovery

1. Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (1998). All Rights Reserved.

2. Abstract

The Dublin Core Metadata Workshop Series began in 1995 with an invitational workshop which brought together librarians, digital library researchers, content experts, and text-markup experts to promote better discovery standards for electronic resources. The Dublin Core is a 15-element set of descriptors that has emerged from this effort in interdisciplinary and international consensus building. This is the first of a set of Informational RFCs describing the Dublin Core. Its purpose is to introduce the Dublin Core and to describe the consensus reached on the semantics of each of the 15 elements.

3. Introduction

Finding relevant information on the World Wide Web has become increasingly problematic due to the explosive growth of networked resources. Current Web indexing evolved rapidly to fill the demand for resource discovery tools, but that indexing, while useful, is a poor substitute for richer varieties of resource description.

An invitational workshop held in March of 1995 brought together librarians, digital library researchers, and text-markup specialists to address the problem of resource discovery for networked resources.

This activity evolved into a series of related workshops and ancillary activities that have become known collectively as the Dublin Core Metadata Workshop Series.

The goals that motivate the Dublin Core effort are:

- Simplicity of creation and maintenance
- Commonly understood semantics
- Conformance to existing and emerging standards
- International scope and applicability
- Extensibility
- Interoperability among collections and indexing systems

These requirements work at cross purposes to some degree, but all are desirable goals. Much of the effort of the Workshop Series has been directed at minimizing the tensions among these goals.

One of the primary deliverables of this effort is a set of elements that are judged by the collective participants of these workshops to be the core elements for cross-disciplinary resource discovery. The term "Dublin Core" applies to this core of descriptive elements.

Early experience with Dublin Core deployment has made clear the need to support qualification of elements for some applications. Thus, a Dublin Core element may be expressed without qualification (as described in this RFC) or with qualifiers that refine its semantics (the subject of future RFCs). For the sake of interoperability, simple indexing and discovery tools should be able to ignore any qualifiers provided, while more advanced, semantically richer tools should be able to use qualifiers to support more specialized or precise discovery.

The broad agreements about syntax and semantics that have emerged from the workshop series will be expressed in a series of Informational RFCs, of which this document is the first.

4. Description of Dublin Core Elements

The following is the reference definition of the Dublin Core Metadata Element Set. Further information about the Dublin Core Metadata Element Set is available at [1]:

http://purl.org/metadata/dublin_core

In the element descriptions below, each element has a descriptive name intended to convey a common semantic understanding of the element, as well as a formal single-word label intended 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 labels given below to avoid conflicts in the event that the metadata is subsequently extracted or converted to a case-sensitive environment, such as XML (Extensible Markup Language) [2].

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 system.

To promote global interoperability, a number of the element descriptions suggest a controlled vocabulary for the respective element values. It is assumed that other controlled vocabularies will be developed for interoperability within certain local domains.

The metadata elements fall into three groups which roughly indicate the class or scope of information stored in them: (1) elements related mainly to the Content of the resource, (2) elements related mainly to the resource when viewed as Intellectual Property, and (3) elements related mainly to the Instantiation of the resource.

| Content | Intellectual Property | Instantiation |
|-------------|-----------------------|---------------|
| ----- | ----- | ----- |
| Title | Creator | Date |
| Subject | Publisher | Format |
| Description | Contributor | Identifier |
| Type | Rights | Language |
| Source | | |
| Relation | | |
| Coverage | | |

4.1. Title Label: "Title"

The name given to the resource, usually by the Creator or Publisher.

4.2. Author or Creator Label: "Creator"

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.

4.3. Subject and Keywords Label: "Subject"

The topic of the resource. Typically, subject will be expressed as keywords or phrases that describe the subject or content of the resource. The use of controlled vocabularies and formal classification schemes is encouraged.

4.4. Description Label: "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.

4.5. Publisher Label: "Publisher"

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

4.6. Other Contributor Label: "Contributor"

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 a Creator element (for example, editor, transcriber, and illustrator).

4.7. Date Label: "Date"

A date associated with the creation or availability of the resource. Recommended best practice is defined in a profile of ISO 8601 [3] that includes (among others) dates of the forms YYYY and YYYY-MM-DD. In this scheme, for example, the date 1994-11-05 corresponds to November 5, 1994.

4.8. Resource Type Label: "Type"

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 currently under development in the workshop series.

4.9. Format Label: "Format"

The data format and, optionally, dimensions (e.g., size, duration) of the resource. The format is used to identify the software and possibly hardware that might be needed to display or operate the

resource. For the sake of interoperability, the format should be selected from an enumerated list that is currently under development in the workshop series.

4.10. Resource Identifier

Label: "Identifier"

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.

4.11. Source

Label: "Source"

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 metadata for the second resource when it is considered important for discovery of the present resource.

4.12. Language

Label: "Language"

The language of the intellectual content of the resource. Recommended best practice is defined in RFC 1766 [4].

4.13. Relation

Label: "Relation"

An identifier of a second resource and its relationship to the present resource. This element is used to express linkages among related resources. For the sake of interoperability, relationships should be selected from an enumerated list that is currently under development in the workshop series.

4.14. Coverage

Label: "Coverage"

The spatial or temporal characteristics of the intellectual content of the resource. Spatial coverage refers to a physical region (e.g., celestial sector) using place names or coordinates (e.g., longitude and latitude). 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). Temporal coverage is typically specified using named time periods (e.g., neolithic) or the same date/time format [3] as recommended for the Date element.

4.15. Rights Management

Label: "Rights"

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.

5. Security Considerations

The Dublin Core element set poses no risk to computers and networks. It poses minimal risk to searchers who obtain incorrect or private information due to careless mapping from rich data descriptions to the simple Dublin Core scheme. No other security concerns are likely to be raised by the element description consensus documented here.

6. References

- [1] Further information about the Dublin Core Metadata Element Set, http://purl.org/metadata/dublin_core
- [2] Extensible Markup Language (XML), <http://www.w3.org/TR/REC-xml>
- [3] Date and Time Formats (based on ISO 8601), W3C Technical Note, <http://www.w3.org/TR/NOTE-datetime>
- [4] Alvestrand, H., "Tags for the Identification of Languages", RFC 1766, March 1995.

7. Authors' Addresses

Stuart L. Weibel
OCLC Online Computer Library Center, Inc.
Office of Research
6565 Frantz Rd.
Dublin, Ohio, 43017, USA

Phone: +1 614-764-6081
Fax: +1 614-764-2344
EMail: weibel@oclc.org

John A. Kunze
Center for Knowledge Management
University of California, San Francisco
530 Parnassus Ave, Box 0840
San Francisco, CA 94143-0840, USA

Phone: +1 510-525-8575
Fax: +1 415-476-4653
EMail: jak@ckm.ucsf.edu

Carl Lagoze
University Library and Department of Computer Science
Cornell University
Ithaca, NY 14853, USA

Phone: +1 607-255-6046
Fax: +1 607-255-4428
EMail: lagoze@cs.cornell.edu

Misha Wolf
Reuters Limited
85 Fleet Street
London EC4P 4AJ, UK

Phone: +44 171-542-6722
Fax: +44 171-542-8314
EMail: misha.wolf@reuters.com

8. Full Copyright Statement

Copyright (C) The Internet Society (1998). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

