**Metadata for Managing Internet Resources**

**Khaiser Nikam**
Department of Library and Information Science
University of Mysore
Mysore
Email: khaisernikam@rediffmail.com

**Abstract**

This article deals with the concept of metadata and its importance in a web based environment. It describes the meaning and functions of metadata. Briefly discusses the guiding principles such as intrinsicality, extensibility, syntax independence, optionality, repeatability and modifyability behind the identification of the metadata element sets. The fifteen Dublin Core element sets and their definition is shown in the form of a table. The relevance of metadata to librarians in the context of rapidly changing web based environment is emphasized along with the necessary skills that the library community has to possess. It also includes a couple of worked out examples.
1 Introduction

The emergence of Internet and its rapid growth in the recent past has influenced the libraries and information centers a great deal. This has resulted in the digital extension of the academic and research libraries. A variety of scholarly information resources such as texts, images and sound now exist in electronic form. They may be accessed from any place and at any point. Among the researchers and scholarly community this information is passed on by word of mouth. But for the others much of this information is very difficult to locate and access.

So, the question arises as to why it is so difficult to find a needed piece of information on the Internet and the World Wide Web? This is because there is lack of standard description about the web resources. Just like how in a physical conventional library, the documents are catalogued, indexed and abstracted in the same manner the varied Internet resources need to be properly indexed and described, so as to facilitate speedy access and precise retrieval. Of course there are a number of well designed search engines like Lycos, which automatically index many of the resources available on the web and maintain up-to-date databases of locations. Indexes are useful in small collection within a given domain. But as and when the space and coverage of the collection increases, indexes pose problems of large retrieval sets and problems of cross-disciplinary semantic drift. Therefore richer records created by content experts need to improve search and retrieval.

Formal standards such as Text Encoding Initiatives (TEI header), MARC cataloguing and government Information Locator Services (GILS) will provide necessary richness, but such records are time consuming to create and maintain and hence may be produced for most important resources.

Hence an alternative solution that promises to mediate these extremes involves the creation of a record that is more informative than an index entry, but is less complete than
a formal cataloguing record with the help of a small amount of human effort such records
could be described especially if the author of the resource could be encouraged to create
the description. If the description followed an established standard only the creation of
the record could require human intervention; automated tools could discover to achieve
consensus on a list of metadata elements that would yield simple description in a wide
range of subject areas, and to lay the ground work for achieving further progress in the
definition of metadata elements that describe electronic information.

However, it is a fact that the Internet has more information than the traditional libraries;
professional indexers and cataloguers can manage using existing methods and systems.
Hence, it was agreed that a reasonable alternate way to obtain usable metadata for
electronic resources is to give authors and information providers a means to describe the
resources themselves.

2 What is Metadata?

Metadata is data about data. Tim Berners – Lee of the W3C describes metadata as
machine understandable information about web resources or other things (Berners, Lee
1998). Metadata is any structured descriptive information about other data that is used to
aid the identification, description, location and management of web resources. The use of
the term metadata predates the web.

The functions of metadata are to provide a means to discover that the data set exists and
how it might be obtained or accessed. To indicate how the resource may be used, by
documenting the content quality, and features of data set. Metadata is intended to use one
of the common search engines to find a simple term, and evaluate the effort actually
needed to find useful information.

Metadata aims to satisfy resource discovery irrespective of subject matter. The
identification of elements for information resource discovery is restricted to DLO’s
(Document like Objects). DLO’s are still considered as the most common resource
sought in the Internet and whatever solutions can be proposed to DLO’s could be extended to other kinds of resources. The DLO’s for instance are electronic versions of newspaper articles or a Dictionary.

In a networked environment DLL’s can be very complex because they consist of text with callouts to images, audio and video clips or other hypertext documents. It is a general belief that the DLO’s are primarily intellectual text and that the metadata required for DLO’s will be similar to the metadata that is need for describing printed text. Therefore the Dublin Core workshop’s goal was to define a core set of metadata elements that would allow authors and information providers to describe their work and facilitate interoperability among resource discovery tools.

3 Guiding principles of Metadata

The Dublin core metadata workshop decided to keep the element sets as small as possible. Their purpose was to see that the meaning of the elements will be understood by most users, and that the elements will be flexible enough for the description of the resources in a wide range of subject areas. The guiding principles are:

- **Intrinsicality**: It refers to properties of the objects such as intellectual contents and physical form. ‘Subject element is intrinsic data. Cost and access considerations are extrinsic data.

- **Extensibility**: This refers to extension mechanisms, which allow the inclusion of intrinsic data for objects that can not be adequately described by a small set of elements. This will allow revisions while maintaining some backward compatibility with the original definition element set.

- **Syntax Independence**: Syntactic bindings are avoided because it is too early to propose formal definitions and because the Dublin core is intended to be eventually used in a range of disciplines and application program.

- **Optionality**: All elements are optional. Dublin core may eventually be applied to objects for which some elements have no meaning (who is the author of a satellite image).
Repeatability: all elements in Dublin core are repeatable. Ex: Multiple author elements would be used when a resource has multiple authors.

Modifiability: an optional qualifier can modify each element in the Dublin core. If no qualifier is present, the element has its common sense meaning; otherwise, the definition of the element is modified by the value of the qualifier. Ex: Subject (scheme = LCHS) indicates that the subject terms are taken from the Library of congress Subject Headings.

4 The Dublin Core workshop

The Dublin Core workshop, which was conducted at Dublin, Ohio in 1995, decided to identify simple set of elements for describing networked resources. The Dublin core metadata set is a simple semantic metadata. It is the consensus of an international group of cross-disciplinary experts, including librarians. The aim of this informal international meet was to design simple metadata sets to facilitate information discovery in a networked environment. The element sets and their definition is shown in Table I.

Table I

DUBLIN CORE DATA ELEMENTS

<table>
<thead>
<tr>
<th>Element Sets</th>
<th>Definition of Element Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>The name given to the resource by the creator or publisher.</td>
</tr>
<tr>
<td>Creator</td>
<td>The person(s) or organization(s) primarily responsible for the intellectual content of the resource.</td>
</tr>
<tr>
<td>Subject</td>
<td>The type of the resource or key words or phrases that describe the subject or content of the resource.</td>
</tr>
<tr>
<td>Descriptor</td>
<td>A textual description of the content of the resource, including abstracts in the case of a document like objects or content descriptors in case of visual resource.</td>
</tr>
<tr>
<td>Publisher</td>
<td>The entity responsible for making the resource available in its present form. Ex: Publisher.</td>
</tr>
</tbody>
</table>
5 Metadata in libraries

Librarians have a lot to look forward if metadata gets smart. The electronic library program in the UK and the European Union’s DESIRE project for research information location and retrieval and developing automated web based cataloguing and indexing systems: subject information gateways are based on quality controlled resource catalogues – containing full description of resources. As archives community are seriously considering metadata to ensure the integrity and longevity of records, a similar approach might be useful for digital preservation in libraries, particularly for networked documents.

As digital resources became more established as information resources, metadata specialists in libraries will have an increasing role in maintaining intellectual access: enhancing collections through digitized preservation of materials, and being a consultant for and with libraries on electronic collection issues. To respond to the demand for
electronic resources, libraries will need personnel with strong computing skills, library knowledge, ad innovative ness and flexible to challenge constantly changing needs and environments.

6 DC in HTML

EXAMPLES I

<html><head>
<title>UKOLN Home page</title>
<meta name="DC.Title" content="UKOLN: office for Library and Information Networking">
<meta name="DC.Subject" content="national centre, network information support, library community, awareness, research, information services, public library networking, bibliographic management, distributed library systems, metadata, resource discovery, conferences, lectures, workshops">
<meta name="DC.Description" content="UKOLN is a national center for support in network information management in the library and information communities. It provides awareness, research and information services">
<meta name="DC.Creator" content="UKOLN services group">
</head>

EXAMPLE 2

<html><head>
<title>Mysore University Library</title>
<meta name="DC.Creator" content="Stephen P. Robinson">
<meta name="DC.Subject" content="Organization Behaviour">
<meta name="DC.Publisher" content="Prntice Hall">
<meta name="DC.Date" content="1998">
<meta name="DC.Description" content="This book deals with various aspects of organization behaviour relating specifically to individuals and organization systems">
<meta name="DC.Language" content="English">
</head>

7 Conclusion

Effective management of networked information will definitely depend on the proper management of metadata. There is a great demand for metadata, as it is not only the key to discovery but it is also essential for interoperability of across protocol domains.
8 References

1. **DESIRE** Project.
   
   http://www.ukoln.ac.uk/metadata/desire/overview/rev_pre.htm


3. **Hodgon (Katrina).** Metadata: Foundations, Potentials and Application.
   
   http://www.gpu.srv.ualberta.ca/-khl1/Metadata.html


   
   http://www.dlib.org/dlib/july96/lagoze/o7lagoze.html

6. **Edward (Gaynor).** From MARC to Markup: SGML and Online Library Systems.
   
   http://www.lib.virginia.edu/specol/scdc/article/alcts_brief.html