

Dublin Core and Electronic Information Retrieval

By

Ezra Shiloba Gbaje

**Department of Library and Information Science
Ahmadu Bello University, Zaria, Nigeria**

Abstract

The explosion of information on the Internet has resulted to the inability of professional abstractors, indexers and catalogers to manage effectively the digital resources using the formal bibliographic apparatus. Hence search for a reasonable alternative way to obtain usable metadata for electronic resources, so as to give authors and information providers a means to describe the resources to facilitate their retrieval. Dublin Core (DC) provides a relatively simple and concise method for resource discovery of composite electronic resources over various networks. This article examines the importance and application of Dublin Core metadata in electronic resource discover over various networks. It is expected that the endorsement of Dublin Core by librarian and other information professionals will address the lack of a common categorization scheme for web metadata which makes searching for information on the web difficult.

Introduction

The development of web technology has made it possible for anyone with access to the technology and the skills to publish on the web. This has resulted to an exponential increase in electronic documents with a dynamic nature. Consequently, users of electronic resources are overwhelmed by the number of hits resulting from a search query with limited ability to refine or make a more precise search. This article examines the importance and application of Dublin Core metadata in electronic resource discover over various networks. The wide scale adoption of descriptive standards and practices for electronic resources will improve retrieval of relevant resources in any venue where information retrieval is critical (Hillmann D., 2005).

Dublin Core (DC) named after the site of the first metadata workshop held in 1995 at Dublin, Ohio was developed within an international and interdisciplinary environment through the collaboration of various internet communities. Metadata is described as data about data, It is structured information that describes the attributes of information packages for the purpose of identification, discovery and

sometime management (Taylor 2004). Some metadata can be derived mechanically from the objects while others are intellectual creation. Hence, metadata associated with an information object can be subjective. Metadata schemas are set of elements designed to meet the needs of particular communities. While most schemas focus on descriptive elements to support resource discovery, some elements are use to support administrative and structural purpose. Dublin Core metadata scheme consists of an element set for describing a wide range of networked resources. Lynch (1996), assert that the Dublin Core metadata scheme is designed to support discovery of information objects of interest using search tools and systems and not designed to provide comprehensive support for access, management, use, or assessment of networked information. It was originally developed to be embedded within web pages written in HTML as a static descriptive record. Caplan (2004) also states, that the Dublin Core metadata schema is a general purpose schema for resources description originally intended to facilitate discovery of information object on the web. Authors and information provider are expected to use the schema to describe their work and to facilitate interoperability among resource discovery tool.

The Dublin Core Metadata Initiative (DCMI) is responsible for promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems. DCMI is international in scope and participation and has brought fore issues related to language and multilingual representation of both metadata and metadata schemes. The DCMI has also created a venue for libraries, museums and other cultural heritage institutions to exchange information with governmental organizations, scientific, agencies, web developers, computer scientists and educators. Also, it raises awareness of interoperability issues beyond the closed MARC environment within the library community (Caplan, 2003).

Underlying Assumptions of Dublin Core (DC)

The underlying assumptions and goals of the development of Dunlin Core according to Hillmann (2005) were:

1. **Simplicity of creation and maintenance:** The DC element set is small and simple to allow every information provider to create simple descriptive records for information resources easily and inexpensively, while providing for effective retrieval of those resources in the networked environment.
2. **Commonly understood semantics:** The DC can help inexperience searcher effectively search the internet by supporting a common set of elements, the semantics of which are universally understood and supported. It will enhance information retrieval across the Internet without hindrance to differences in terminology and descriptive practices from various field of knowledge.
3. **International scope:** Versions of DC Element Set are being created in many other languages including Finnish, Norwegian, Thai, Japanese, French, Portuguese, German, Greek, Indonesian, and Spanish.
4. **Extensibility:** DC developers have recognized the importance of providing a mechanism for extending the DC element set for additional resource discovery needs. Metadata elements

from these sets could be used in conjunction with DC metadata to meet the need for interoperability.

Simple and Qualified Dublin Core standard are the two main level of DC: **Simple Dublin Core:** The simple Dublin Core defines a set of fifteen basic data elements for resource description and retrieval. The fifteen elements correspond to data in traditional catalog records and are divided among three information categories:

Content:

- **Title:** The name given to the resource by the author or publisher.
- **Subject and Keywords:** The topic of the resource work, or keywords that describe the content of the work.
- **Descriptors:** A textual description of the content of the resource.
- **Source:** Objects, either print or electronic, from which this object is derived, if applicable.
- **Language:** Language of the intellectual content.
- **Relation:** Relationship to other objects.
- **Coverage:** The spatial locations and temporal period or jurisdiction.

Intellectual Property:

- **Creator or Author:** The person(s) primarily responsible for the intellectual content of the object.
- **Publisher:** The agent or agency responsible for making the object available in its present form. Generally a publisher, an institution (university department, for example) or a corporate entity.
- **Contributors:** The person(s) other than the author(s) who have made other significant intellectual contributions to the work (for example, editors, transcribers, illustrators, conveyors).
- **Rights management:** Intended to be a link to a copyright notice, a rights-management statement, etc.

Instantiation

- **Date:** The date the work was made available in its present form.
- **Resource Type:** The genre of the object, such as home page, novel, poem, working paper, technical report, essay, dictionary, etc.
- **Format:** The data representation of the object, such as Postscript file or Windows executable file.
- **Resource Identifier:** String or number used to uniquely identify the object.

Qualified Dublin Core: The qualified DC includes Audience, Provenance and Right Holder as well as group of qualifiers which are companion specification to the Dublin Core Metadata Element Set. Vellucci (2000), assert that the ability to add standard qualifiers to the DC elements would improve the scheme, especially in the area of authority control. Standard qualifiers would increase semantic specificity by allowing controlled vocabularies or classification schemes to be identified and used for either description or as a formal browsing structure. Standard qualifiers will also allow authority controlled data to identify unique name, title, work, object, etc. as well as allow definition of a formal substructure so that compound values can be assigned to one element.

Dublin Core is not governed by mandatory rules as each element is optional and may be repeated. Each element value for an object could reasonably be defined by its author, or by a site manager, rather than by a trained specialist cataloger or indexer. There was no syntax defined for use in DC as data elements could be used in many different environments and with any type of database architecture. The implementation however is subject to the syntax of the software used in consistent with the existing web standards.

Despite the simplicity of the DC scheme some of the lapses listed below have been observed.

1. **Overlap in Meaning:** The overlap in meaning in the definition of some elements has led to confusion among implementers about when it is appropriate to use one element rather than another. An example is when to use Creator, Contributor, and Publisher. Also, Creator can be seen as a particular

type of Contributor, and Source. However distinguishing the special role of Creator from other contributors is logical for describing electronic versions of resources created by retrospective conversion project.

2. **Relationship to surrogate:** For documents that are born digital, the creator of the document is the person who authored it. But for a digital version of a painting, the name of the person who scanned the photo is the creator of the image. However some Dublin core implementers feel that the scanning technician is intellectually meaningless and should be recorded, if at all as contributor. Others believe that a Dublin Core record should accurately describe the resource in hand attributing the creator to the scanner and artist as the contributor. This principle, known as "one-to-one", prescribes that if multiple versions of resources exist, each should be described separately and accurately (Caplan, 2004).

Encoding Dublin Core

Dublin Core has been implemented through the use of HTML <META > and <LINK> tags in HTML-based web pages and the use of XML structures as specified by the World Wide Web Consortium's Resource Description Framework (RDF) for XML-based web pages. The DC elements are associated with the objects by incorporating them as part of the web pages either by the authors of the pages, by website managers, or by third party indexers/catalogers. There are currently some web-based DC generator and editor used to create DC <meta> tags for web pages which can be displayed in HTML, XHTML, XML and RDF formats. The followings are two web based Dublin Core generator:

1. **DC-dot:** DC-dot is a Dublin Core generator and editor provided to support XHTML, HTML, XML, RDF. It was developed by Andy Powell, UKOLN, University of Bath. It allows users to enter the URL of a web page. It retrieve the Web page and automatically generate DC metadata either as HTML <meta> tags or as RDF/XML. It is suitable for embedding in the

<head>...</head> section of the page. The generated metadata can be edited using the form provided and converted to various other formats such as USMARC, SOIF, IAFA/ROADS, TEI headers, GILS, IMS or RDF if required. DC-Dot is available from <http://www.ukoln.ac.uk/metadata/dcdot/>

2. **Dublin Core Metadata Template:** In order to ensure good support for the creation of DC creation, Nordic Metadata Project created the Dublin Core metadata Template. Authors or information providers are to fill in the DC elements and the necessary <meta> tags are generated, previewed as well as copy and pasted within the page's <HEAD> </HEAD> field of web page. An example of a DC metadata template can be accessed at <http://www.library.kr.ua/dc/dceditunie.html>.

Using the DC-Dot, the following <meta> tag was generated for the website <http://www.sharondelight.com> which is display using HTML. Option was provided to either edited or add more Dublin Core elements. In the example below, META tags are used to hold the descriptive elements and their values.

```
<link rel="schema.DC"
href="http://purl.org/dc/elements/1.1/" />
```

```
<link rel="schema.DCTERMS"
href="http://purl.org/dc/terms/" />
```

```
<meta name="DC.title" lang="English"
content="SharonDelight Information Consult" />
```

```
<meta name="DC.creator" content="Ezra Shiloba
Gbaje" />
```

```
<meta name="DC.subject" lang="English"
content="open access; MLS Portfolio; MLS Program
Outcomes; Publications; Virtual Library Project;
Systems Consultant; sdconsults@yahoo.com;
Resume" />
```

```
<meta name="DC.description" lang="English"
content="Personal and Business website for my
consultancy department" />
```

```
<meta name="DC.publisher" content="yahoo.com"
/>
```

```
<meta name="DC.contributor" content="Ezra
Shiloba Gbaje, Helen E. Gbaje" />
```

```
<meta name="DC.date"
scheme="DCTERMS.W3CDTF" content="2006-10-
31" />
```

```
<meta name="DC.type"
scheme="DCTERMS.DCMIType" content="Text" />
```

```
<meta name="DC.format" content="text/html" />
```

```
<meta name="DC.format" content="8882 bytes" />
```

```
<meta name="DC.identifier"
scheme="DCTERMS.URI"
content="http://www.sharondelight.com/index.htm"
/>
```

```
<meta name="DC.language"
scheme="DCTERMS.URI" content="English" />
```

```
<meta name="DC.rights"
scheme="DCTERMS.URI" content="Ezra Shiloba
Gbaje" />
```

Capan (2004), asserts that representing DC semantics in HTML works well in environments where web pages are indexed by search engines configured to take advantage of <meta> tags. However, HTML cannot represent complex constructions where sets of repeated elements need to be grouped to be meaningful. Also, for applications that require metadata records, as opposed to metadata embedded in documents, XML tends to be the preferred exchange syntax.

An example of Dublin Core <meta> element generated with DC-Dot and displayed in XML format is provided. In XML encoding, element names appear in the tags with a prefix "dc", which is associated to a namespace

```
<?xml version="1.0"?>
```

```
<metadata
xmlns="http://www.ukoln.ac.uk/metadata/dcdot/"
">
```

```
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
```

```
xsi:schemaLocation="http://www.ukoln.ac.uk/m
etadata/dcdot/
http://www.ukoln.ac.uk/metadata/dcdot/dcdot.xs
d"
```

```
xmlns:dc="http://purl.org/dc/elements/1.1/">
```

```
<dc:title xml:lang="English" >
  SharonDelight Information Consult
</dc:title>
```

```
<dc:creator>
  Ezra Shiloba Gbaje
</dc:creator>
```

```
<dc:subject xml:lang="English" >
  open access; MLS Portfolio; MLS Program
  Outcomes; Publications; Virtual Library
  Project; Systems Consultant;
  sdconsults@yahoo.com; Resume
</dc:subject>
```

```
<dc:description xml:lang="English" >
  Personal and Business website for my
  consultancy department
</dc:description>
```

```
<dc:publisher> yahoo.com </dc:publisher>
```

```
<dc:contributor> Ezra Shiloba Gbaje, Helen E.
Gbaje </dc:contributor>
```

```
<dc:date> 2006-10-31 </dc:date>
```

```
<dc:type> Text </dc:type>
```

```
<dc:format> text/html </dc:format>
```

```
<dc:format> 8882 bytes </dc:format>
```

```
<dc:identifier>
http://www.sharondelight.com/index.htm
</dc:identifier>
```

```
<dc:language> English </dc:language>
```

```
<dc:rights> Ezra Shiloba Gbaje
</dc:rights>
```

```
</metadata>
```

Library Application

In the last few years as noted by Hillmann (2005) there have been an increase in the application of DC metadata where it is used to describe resources held, owned or produced by companies, governments and international organizations to support portal services or internal knowledge management. In the library Lynch (1997), identified two different basic applications of Dublin Core to include, allowing libraries to provide their patrons with consistent views of both library and non-library databases and allowing the description of new resources that cannot be cost effectively supported through traditional cataloging approach. Highlighting the application of DC in library Caplan (2003), asserts that it Core has been used in the library as subject gateways or portals, where the description of a resource appearing in the web gateway is generated from a database of brief Dublin Core Information. It is also used for describing electronic texts and images created in retrospective digitization projects for large number of items. For certain materials such as photographs or newspaper articles, where AACR2 rules may be difficult to apply and integrated into the library catalog, Dublin Core-based schemes can be use to identify data elements and guidelines that are meaningful and also create some standardization to use. Dublin Core serves as a least common denominator to which more complex schemes can be mapped, so that searching can take place over a consistent set of data element. It can also serve as a least common denominator in which metadata stored in the local databases in different schemes can be converted into, for use in a union catalog, Internet search engine index, or in other external database containing contributions from multiple sources.

Several digitization projects have been taking advantage of the elements of DC to create their base records. Greenstone and Dspace provide are example of digital software that use qualified DC templates for describing digital materials put into these systems, which enables users search or browse for content relevant to their needs. Another very common use of Dublin Core is for Open Archive Initiative Protocol for Metadata Harvesting ([OAI-PMH](#)) which is an [application-independent interoperability framework](#). The OAI-PMH protocol requires that a simple Dublin Core record be present for

every item represented in an OAI data provider, which can be supplemented with records in other formats. All libraries wishing to become Open Archive Initiative data providers must be able to produce simple Dublin Core for the records they wish to share via the protocol.

Conclusion

Dublin Core has gained wide acceptance within the digitization software developers and many metadata applications have been developed based on Dublin Core. It is expected that web pages will increasingly incorporate Dublin Core data elements as part of the web pages and web indexing programs will be upgraded to capture this metadata and incorporate it into their web databases which will eventually increase their retrieval of web pages for users. As Nigeria libraries move towards digitalization, awareness about DC will help librarian and information professionals contribute to the success of the digitization process. The adoption and use of Dublin Core to describe electronic documents for digital library software is an indication of the acceptability of Dublin Core in the library community.

References

Caplan P (2003). *Metadata Fundamentals for All Librarians*. Chicago American Library Association.

Hillmann D. (2005) Using Dublin Core. Retrieved on 07/20/2005 from <http://dublincore.org/documents/usageguide/>

Lynch C. (1997) The Dublin Core Descriptive Metadata Program: Strategic Implications for Libraries and Networked Information Access. Retrieved on 07/18/2005 from <http://www.arl.org/newsltr/196/dublin.html>

Milstead J and Feldman S.(1999) METADATA: cataloging by any other name. Retrieved on 07/25/2005, from: <http://www.cbuc.es/5digital/1.pdf>

Nordic Dublin Core Metadata Template Retrieved on 07/22/2005 from <http://www.lub.lu.se/cgi-bin/nmdc.pl>

Powell A. (2002). Dublin Core metadata editor. Retrieved on 07/25/2005, from <http://www.ukoln.ac.uk/metadata/dcdot/>

Taylor G. A (2004). *The Organization of Information*. (2nd ed). Westport CT: Libraries Unlimited, Inc.

Vellucci L. S. (2000). Metadata and Authority Control, *Library Resources & Technical Services* 44(1) 33-43