Dublin Core: Process and Principles

Shigeo Sugimoto[1], Thomas Baker[2], Stuart L. Weibel[3]

[1] Institute of Library and Information Science, University of Tsukuba, Japan
sugimoto@slis.tsukuba.ac.jp
[2] Institutszentrum Schloss Birlinghoven, Fraunhofer-Gesellschaft, Germany
Thomas.Baker@bi.fhg.de
[3] Dublin Core Metadata Initiative, OCLC Office of Research
weibel@oclc.org

Abstract. The Dublin Core metadata element set has been widely adopted by cultural and scientific institutions, libraries, governments, and businesses to describe resources for discovery on the Internet. This paper provides an overview of its history and underlying principles and describes the activities of Dublin Core Metadata Initiative as an organization.

1. Introduction

In the Web age, metadata is typically defined as "data about data" – a simple definition that embraces a broad range of resources from library catalogues and indexes to thesauri, ratings, reviews, terms and conditions for use. In the Internet, metadata is designed for tasks ranging from resource description and discovery to archiving, trading, content filtering, resource syndication, and information management. This diversity of purpose reflects the variety of information resources available on the Internet, which range from personal Web pages to huge portals for government information, digital libraries, and shopping catalogues. Users of the Internet range from small children to businesses and professionals.

From its origins in the mid-1990s, the Dublin Core has defined itself as a small set of core descriptive attributes by which users can search for information across a broad range of sources [1]. It was recognized from the outset that semantic interoperability across domains implied conceptual simplicity, much in the manner of natural-language pidgins, which use small vocabularies and simple grammars to enable rough comprehension between speakers of different languages. As the Dublin Core became adapted for specialized purposes, however, the focus shifted to methods for qualifying and extending the core vocabulary, and to architectures for encoding Dublin Core descriptions interoperably.

The first, thirteen-element Dublin Core was the result of a workshop held in Dublin, Ohio in 1995, which had been planned as a result of a casual conversation at the Chicago WWW conference in 1994. Since then a series of workshops and conferences has built on this initial consensus, clarifying issues of architecture and extensibility and broadening international participation.

Since the initial "classic" Dublin Core, which stabilized at fifteen elements in 1998, several implementation styles have emerged. "Simple Dublin Core" uses the fifteen

elements in a very broad, generic manner. In addition to such simplicity, the ability to use Dublin Core for more precise and detailed description was identified as an important need. "Qualified Dublin Core", therefore, uses additional terms to specify the meaning of the Core elements within the context of specific domains. Since such complexification seemed contrary to simplicity and interoperability among different domains, the Dublin Core community embraced the notion of a modular architecture and elaborated the so-called Dumb-Down Principle. A process model (centered around a Usage Board that acts as an editorial body) is used to approve new descriptive terms. The Dublin Core has evolved among participants who have differing requirements but can integrate their needs through the approval of modular extensions to the basic core.

2. Dublin Core Metadata Element Set (DCMES)

2.1 Simple Dublin Core and Qualified Dublin Core

"The Dublin Core" has been defined since 1998 as a set of fifteen elements for cross-domain resource discovery. The set of elements is shown in Table 1. By design, any of the fifteen elements is optional and repeatable. This set has been approved as an international standard in Europe (CEN/ISSS CWA 13874) and a national standard in the USA (ANSI/NISO Z39.85).

Table 1. The Fifteen Elements of "Simple Dublin Core"

Identifier	Definition
Title	A name given to the resource.
Creator	An entity primarily responsible for making the content of the resource.
Subject	The topic of the content of the resource.
Description	An account of the content of the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the content of the resource.
Date	A date associated with an event in the life cycle of the resource.
Type	The nature or genre of the content of the resource.
Format	The physical or digital manifestation of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A reference to a resource from which the present resource is derived.
Language	A language of the intellectual content of the resource.
Relation	A reference to a related resource.
Coverage	The extent or scope of the content of the resource.
Rights	Information about rights held in and over the resource.

The use of these fifteen elements for metadata records, with no additional qualifiers, and with only plain-text strings as values, is known as "Simple Dublin Core" [2].

"Qualified Dublin Core", in contrast, uses the elements together with qualifiers that increase the richness and precision of description [3][4]. Table 2 shows a list of qualifiers approved by DCMI as "recommended" qualifiers as of September 2002.

The approval process and status are explained in detail in section 3.4. Qualified DC has two types of qualifiers – element refinement and encoding schemes. An element refinement narrows the meaning of an associated element; for example, "Date Created" is a more narrowly defined instance of *Date*, and an *Abstract* is seen as a type of *Description*. An encoding scheme qualifier specifies a name of a vocabulary or a name of data encoding scheme used in encoding of a value of its associated element; for example, LCSH encoding scheme qualifier associated with Subject element specifies that a value of the Subject element is expressed in terms of the Library of Congress Subject Headings (LCSH). Qualified DC does not include qualifiers to express components of a value, such as first and last names.

DCMES is a stable but not a closed set. DCMES evolves in accordance with requirements to express resource properties and value-types which are not expressible using existing ones. Table 2 includes a qualifier associated with *Audience* element, which was approved in 2001. The definition of *Audience* element is "A class of entity for whom the resource is intended or useful". *Audience* element was originally proposed by the working group on educational applications and approved as a recommended element for the global community.

Table 2. DCMI Recommended Qualifiers

Element	Element Refinement	Encoding Scheme
Title	Alternative	
Subject		LCSH, MeSH, DDC, LCC, UDC
Description	Table of Contents, Abstract	
Date	Created, Valid, Available, Issued, Modified	DCMI Period, W3C-DTF
Type		DCMI Type Vocabulary
Format	Extent	
	Medium	IMT
Identifier		URI
Source		URI
Language		ISO 639-2, RFC 1766, RFC 3066
Relation	Is Version Of, Has Version, Is Replaced By, Replaces, Is Required By, Requires, Is Part Of, Has Part, Is Referenced By, References, Is Format Of, Has Format, Conforms To	URI
Coverage	Spatial	DCMI Point, ISO 3166, DCMI Box, TGN
	Temporal	DCMI Period, W3C-DTF
Audience	Mediator	

2.2 Encoding Dublin Core Metadata

DCMI provides documents describing three predominant encoding styles for Dublin Core metadata. The oldest of these styles embeds Dublin Core descriptions in HTML

with special tags [5]. In the example below, META tags are used to hold the descriptive elements and their values.

```
<meta name="DC.Title"
    content="Dublin Core Metadata Initiative Home Page">
<meta name="DC.Language" content="en">
<meta name="DC.Contributor"
      content="Dublin Core Metadata Initiative">
<meta name="DC.Date" content="2001-01-16">
<meta name="DC.Format" content="text/html">
```

In the HTML style, LANG attributes are used to indicate the language of metadata values, as in the following example in German.

```
<meta name="DC.Title" lang="de"
    content="Dublin-Core Metadata-Diskussionen">
```

In XML encoding, element names appear in the tags with a prefix "dc", which is associated to a namespace [6].

```
<?xml version="1.0"?>
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="..." xsi:schemaLocation="..."
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>UKOLN</dc:title>
  <dc:description>UKOLN is a national focus of expertise
   in digital information management...</dc:description>
  <dc:publisher>UKOLN, University of Bath</dc:publisher>
  <dc:identifier>http://www.ukoln.ac.uk/</dc:identifier>
</metadata>
```

At its simplest, encodings in RDF (below) resemble the XML style above [7][8]. As detailed in the draft DCMI recommendations, however, metadata in RDF provides conventional ways to embed diverse types of information into metadata without compromising the Dumb-Down Principle.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description>
    <dc:creator>Karl Mustermann</dc:creator>
    <dc:title>Algebra</dc:title>
    <dc:subject>mathematics</dc:subject>
    <dc:date>2000-01-23</dc:date>
    <dc:language>EN</dc:language>
    <dc:description>An introduction to algebra</dc:description>
  </rdf:Description>
</rdf:RDF>
```

3. Dublin Core Principles

DCMI is a forum for development of metadata standard for resource discovery across domains and is opened to international and multilingual communities. Simplicity, extensibility and semantic interoperability are the fundamental technical characteristics of the development of Dublin Core. This section discusses these key concepts.

3.1 Warwick Framework – Basic Framework for Extensibility

Since the Internet is a very diversified environment, it is useless to assume that a single metadata set will meet the needs of all domains and purposes. It is also impractical to develop metadata sets application by application: the result would be expensive and chaotic, and interoperability would be non-existent. On the other hand, it is desirable for application developers to use established metadata schemas and adopt them in accordance with local requirements. The Warwick Framework, a conceptual model that resulted from the 2nd Workshop in 1996, gave an early expression to the notion of metadata as modular components that may come from more than one metadata schema [9]. In this model, a metadata instance is expressed as a container which contains one or more packages, each of which is expressed in a given metadata schema. The Resource Description Framework (RDF), the development of which began in 1998, provided a practical realization of many of the ideas of the Warwick Framework [10].

The Warwick Framework is important as a model for modular metadata on the Internet. No single metadata schema is sufficient to all applications. Rather, it is necessary to adopt appropriate elements from various schemas in accordance with the functional requirements of an application. The role of Dublin Core in Warwick Framework is to provide a core set of metadata elements for resource discovery in any application domains. In other words, Dublin Core can work as a common schema to find resources across many domains.

3.2 The Dumb-Down Principle as a Basis for Interoperability

The Dumb-Down principle gives a guideline for qualification. The Dumb-Down principle suggests that a value of a qualified element has to be consistent as a value of the element without any qualification. For example, assume the following qualified values;

1. (Element Refinement) Date Accepted: "2002-12-11",
2. (Encoding Scheme) Language: "en" encoded in RFC 1766, and
3. (Value Structure) Creator: {name: "Sugimoto, Shigeo", affiliation: "University of Tsukuba", contact: "sugimoto@slis.tsukuba.ac.jp"}

Then, assume that qualification in the above examples, *Accepted*, *RFC 1766* and the component names of the value structure (i.e., *name*, *affiliation* and *contact*) are removed. The values of example 1 and 2, "2002-12-11" and "en" are consistent with

their elements after the removal. However, the value of example 3 {"Sugimoto, Shigeo", "University of Tsukuba", "sugimoto@slis.tsukuba.ac.jp"} causes problems since the second and third values are not valid values of *Creator*.

Dumbing-down is a crucial function for metadata interoperability in the global community; local communities can extend their schemas in accordance with their requirements whereas they are also encouraged to keep their metadata interoperable with other metadata communities.

3.3 Application Profiles

Dublin Core Metadata defines the vocabulary of metadata, i.e., terms and their meanings, but in general does not specify the encoding or syntax characteristics. An exception is the feature included in Simple DC that is "Any of the 15 elements is optional and repeatable." Local applications, however, may have domain specific requirements appropriate to a given domain or application:
  - Title, Creator and Description might be mandated but others are optional,
  - Use only Title, Creator, Description, Date and Language elements,
  - Use the 15 elements of Simple DC and some elements from other metadata
    sets such as the IEEE Learning Object Metadata, and so forth.
These requirements can be defined independently of the definition of vocabulary. Description of this application specific syntactic feature is called an *application profile*. Any application can have its own application profile, which specifies a set of metadata vocabulary terms used in the application as well as syntactic or structural features of the particular application. The vocabulary terms could be borrowed from one or more source schemas. More importantly, the application profile could be used to define a mapping between the application's scheme to global scheme(s), which is crucial for interoperability.

3.4 A Process Model for Maintaining the Standard

To remain relevant in a rapidly evolving Web environment, Dublin Core must be able to grow and evolve in response to user needs. DCMI has therefore instituted a Usage Board and a process model for reviewing proposals for expanding or clarifying the standard. Primary among these functions is the review of proposals for new elements and qualifiers (generically, "terms"). Requirements for new terms may originate in a particular application community. DCMI working groups crystallize these requirements both on mailing lists and in face-to-face meetings and formulate proposals for presentation to the Usage Board.

The Usage Board evaluates such proposals for their conformance to architectural and grammatical principles. This Dublin Core "grammar" includes a typology of Elements, Element Refinements, and Encoding Schemes along with some general principles, such as the axiom that the values of element refinements should be usable as values of the element refined. Proposed elements and element refinements that conform to Dublin Core principles are taken into the standard with the status of *conforming*. To some proposed terms of proven usefulness for resource discovery

across domains the Board may assign the status of *recommended*. Proposals for encoding schemes are reviewed for accuracy and given the status of *registered*.

Once approved, each new term is assigned a Uniform Resource Identifier using one of the official namespace URIs maintained by DCMI. A "namespace policy" defines limits within which the metadata terms maintained by DCMI can evolve or change over time. According to this policy, editorial changes or updates are allowed, but changes of semantics (meaning) are not; new semantics require the creation of new element.

The Usage Board has met twice in 2001 and once in 2002, defining formal review processes, developing procedures for registering externally maintained encoding schemes, and approving several proposals for new terms. Proposals which are not approved are sent back to working groups with suggestions on how they might be revised and resubmitted. The process has the feel of the review board for a scientific journal or conference, where reviewers may actively engage with authors for the common purpose of improving the end results.

The underlying motivation for the Usage Board is to provide a framework in which metadata requirements that "bubble up" in particular implementation contexts can be shared in wider circles and eventually be incorporated into a standard where they will be declared in a persistent way and maintained in accordance with known principles. This reflects the conviction that metadata usage, analogously to language usage in general, can only partially be steered from the top down, on the model of traditional standardization activity. In the DCMI model, the art of standards development lies in striking balances between innovation from below and qualified review from above or between domain-specific specificity and cross-domain applicability. The Usage Board process aims to guide the formulation and formalization of community standards for particular domains that integrate well into broader frameworks for interoperability.

3.5 Internationalization and Localization

As described above, Dublin Core is intended for resource discovery on the global Internet. There are several issues which have been identified as crucial for the adoption of Dublin Core by the international community where resources and their metadata are created in different languages under different cultures.

DCMI has been soliciting local communities to translate descriptions of DCMI terms and other documents into local languages in order to promote understanding by non-English speaking people. For example, the DCMI registry described below provides translations of DCMES into 23 languages, which have been translated mostly by volunteers.

A local community which shares a local language and/or culture plays a crucial role for the global use of DC metadata. Translation is a part of the efforts to adopt DC metadata to a local language. Only a local community can identify local requirements based on a local language and culture and let the global community know the requirements. DCMI is promoting the formation of regional organizations to support the local activities and development of local communities.

3.6 Metadata Schema Registry for Information Sharing

A Metadata schema registry is an infrastructure function sharing metadata schemas on the network to enhance interoperability of metadata. DCMI is building a metadata schema registry, which provides reference descriptions of DC terms. The reference descriptions are declared using RDF schema to promote readability and exchange by machines and applications. The reference description encoded in RDF schema provides an identifier given to a term which is uniquely identifiable in the Internet, and also a label and a description which could be given not only in English but also in other languages. Thus, the registry associates human understandable labels and descriptions in multiple languages with a unique identifier for machine understandability.

   The registry will play an important role for the long-term maintenance of the reference descriptions, which is a crucial but challenging issue. Every DCMI term has its approval status and human readable label and descriptions, which could change over time. Every term could have translations which may be appended and modified over time [11]. Maintenance of local or domain specific schemas is also an important and challenging issue because the community maintaining the registry has to maintain consistency with other registries such as the central DCMI registry.

4. Dublin Core Metadata Initiative (DCMI)

4.1 Dublin Core Metadata Initiative and History in Brief

The DCMI is built on a community of individuals from many different backgrounds and disciplines located in organizations and institutions all over the world. The mission of the DCMI is to make it easier to find resources using the Internet through the following activities:
  - Developing metadata standards for discovery across domains ;
  - Defining frameworks for the interoperation of metadata sets;
  - Facilitating the development of community or discipline-specific metadata sets that work within the frameworks of cross-domain discovery and metadata interoperability.
   The major structural components of DCMI as of 2002 are the Directorate, Board of Trustees, Advisory Board, Usage Board, Working Groups and Interest Groups.
  - The Dublin Core Directorate consists of an Executive Director and a Managing Director to supervise the management and coordination of Working Group activities and assist in the development and refinement of techniques promoting metadata interoperability. The directorate also oversees the development of the Web site and related infrastructure.
  - The Board of Trustees advises the Directorate on strategic issues and allocation of financial resources, contributes to the promotion of the Initiative through liaisons with the public and private sectors and assists in securing support for the Initiative. The trustees were chosen to provide strategic leadership and support to the organization, and were selected for their

leadership and professional abilities in the public, private, and educational sectors. Board members come from six countries on four continents.

- The Dublin Core Advisory Board is comprised of all chairs of DCMI Working Groups and Interest Groups and invited experts. The Advisory Board gives advice to the DCMI Directorate on all technical and strategic issues that occur during the operation of the DCMI. It has a dual role in the DCMI: an internal role to assist in and advise on the developments that take place within DCMI, and an external role to liaise with the stakeholder community and other global metadata initiatives.
- The mission of the DCMI Usage Board is to ensure an orderly evolution of metadata vocabularies. The Usage Board evaluates proposed vocabulary terms (or changes to existing terms) in light of grammatical principle, semantic clarity, usefulness, and overlap with existing terms. The Usage Committee strives for consensus, justifying its decisions and interpretations in terms both of principle and of empirical practice.
- Working Groups and Interest Groups are formed and dissolved as necessitated by the work at hand and the availability of expertise to accomplish such work. Working and Interest groups will be comprised of volunteers with the interest, expertise, and time to contribute to the solution of problems.

The workshop series and the mailing lists are the major forums for discussion of the development of Dublin Core metadata. Table 3 shows the locations and primary hosts of the Dublin Core workshop series since 1995. There was an active discussion on qualifiers at the 4th Workshop in Canberra, Australia in 1997. Simple DC was fixed at the 5th Workshop in Helsinki in 1997. The Dumb-down principle proposed at the 6th Workshop in Washington DC clarified the qualifier types. Development of the fundamental concept of Dublin Core was completed by the 6th Workshop, and maintenance and evolution of Dublin Core were recognized as an important topic since the 7th Workshop in Frankfurt in 1999. Organizational issue for sustainability of DCMI became one of the key issues since this workshop. The 8th Workshop in Ottawa was the first meeting which included sessions for posters and demos to report implementation experiences and new technologies. At the 9th Workshop the presentation session was extended and the whole event was organized as an international conference, which was named DC-2001 [12].

Table 3. Dublin Core Workshop Series

	Year	Location	Primary Host
1	1995	Dublin, Ohio, USA	OCLC
2	1996	Warwick, UK	UKOLN
3	1996	Dublin, Ohio, USA	OCLC
4	1997	Canberra, Australia	National Library of Australia
5	1997	Helsinki, Finland	National Library of Finland
6	1998	Washington DC, USA	Library of Congress
7	1999	Frankfurt, Germany	Die Deutsche Bibliothek
8	2000	Ottawa, Canada	National Library of Canada
9	2001	Tokyo, Japan	National Institute of Informatics
10	2002	Florence, Italy	Biblioteca Nazionale Centrale Firenze

4.2 DCMI and Other Metadata Initiatives

DCMI and IEEE-LOM. At DC-8 in Ottawa in October 2000, DCMI and representatives of the IEEE-Learning Object Metadata (LOM) working group concluded a memorandum of understanding indicating areas of possible convergence on principles and encoding approaches that have the potential to increase interoperability between the two communities. A subsequent meeting in Ottawa in August 2001 identified specific work items. A prominent deliverable from this activity is the recently published "Metadata Principles and Practicalities," an expression of agreement among leaders in the Dublin Core community and the e-learning community concerning basic principles of metadata [13]. This consensus should value to metadata practitioners in these respective communities as well as among metadata practitioners in general.

Dublin Core and Open Archives Initiative (OAI). The Dublin Core Metadata Initiative and the Open Archives Initiative are actively cooperating on metadata issues. Unqualified DC metadata is the default metadata set used in the OAI Protocol for Metadata Harvesting for the purposes of promoting cross-domain interoperability. Other domain-specific sets are encouraged as well, as envisaged in the modular metadata framework that both communities have been striving for. The OAI-DC schema has been developed for use with the OAI Protocol, and has been discussed at length in the DC-Architecture working group. It is expected that the schema will be of use for other applications as well, and will be hosted on the DCMI Website and maintained by representatives of both groups. This development is an important landmark in the development of Web-based metadata services, reflecting as it does the convergence of community consensus and the development of enabling infrastructure to support that consensus.

W3C Semantic Web Activity. The launch of the Semantic Web activity by the W3C recognizes the increasing importance of supporting the infrastructure for defining, registering, and referencing structured vocabularies and ontologies on the Web. The Dublin Core is an important part of this infrastructure, and the DCMI community has played a major role in laying the foundations for this work. A joint project between DCMI staff and W3C staff now under development will help illustrate the value of combining technologies such as the Resource Description Framework of the W3C with the Dublin Core to advance semantic interoperability on the Web.

The joint project between DCMI and W3C staff will unify access to a substantial amount of data from different sectors in different countries using RDF schema declarations as described above. Participants will be recruited from the government sector, museums, business, trans-governmental organizations, and education. The resulting database will comprise a testbed accessible to researchers and designers to demonstrate and experiment with an operational cross-disciplinary store. It will provide a tutorial by example on a schema-based approach to enhancing cross-domain interoperability.

5. Conclusion

Dublin Core has gained wide acceptance and many metadata applications have been developed based on Dublin Core. In these seven years, Dublin Core has evolved; its underlying concepts have been clarified and the community model for maintaining Dublin Core has been accepted. DCMI has been promoting cooperation with other metadata communities, which will greatly enhance semantic interoperability of metadata. On the other hand, there is a lot of work left for future; for example, long-term maintenance of DC metadata in the multi-language community, development of regional communities, and further broadening of the uptake of DC metadata by other communities.

References

1. Dublin Core Metadata Initiative (DCMI) Home Page, http://dublincore.org/ (2001)
2. Dublin Core Metadata Element Set, Version 1.1: Reference Description, http://dublincore.org/documents/1999/07/02/dces/ (1999)
3. Dublin Core Qualifiers, http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/ (2000)
4. Baker, T.: A Grammar of Dublin Core, D-Lib Magazine 6 (2000) http://www.dlib.org/dlib/october00/baker/10baker.html
5. Kunze, J.: Encoding Dublin Core Metadata in HTML, http://www.ietf.org/rfc/rfc2731.txt (1999)
6. Powell A., Johnston, P.: Guidelines for Implementing Dublin Core in XML, http://dublincore.org/documents/2002/09/09/dc-xml-guidelines/ (2002)
7. Beckett, D., et al.: Expressing Simple Dublin Core in RDF/XML, http://dublincore.org/documents/2001/11/28/dcmes-xml/ (2001)
8. Kokkelink, S. Schwaenzl, R.: Expressing Qualified Dublin Core in RDF/XML, http://dublincore.org/documents/2002/04/14/dcq-rdf-xml/ (2002)
9. Lagoze, C.: The Warwick Framework, D-Lib Magazine 2 (1996) http://www.dlib.org/dlib/july96/lagoze/07lagoze.html
10. Resource Description Framework (RDF), http://www.w3.org/RDF/
11. Sugimoto, S. et al.: Versioning the Dublin Core across Multiple Languages and over Time, Proceedings of International Symposium and Applications and Internet 2001 Workshop (2001) 35-45
12. Sugimoto, S., et al.: Report from International Conference on Dublin Core and Metadata Applications 2001, Proceedings of 68th IFLA General Conference and Council, http://www.ifla.org/IV/ifla68/papers/073-151e.pdf (2002)
13. Weibel S. L., et al.: Metadata Principles and Practicalities, D-Lib Magazine 8 (2002) http://www.dlib.org/dlib/april02/weibel/04weibel.html