Metadata for Electronic Documents Using the Dublin Core

Till Hänisch
University of Cooperative Education, Heidenheim, Germany

INTRODUCTION

The Dublin Core Element Set was developed at the OCLC/ NCSA Metadata Workshop in Dublin (Ohio), 1995 (hence the name). It is maintained by the Dublin Core Metadata Initiative (DCMI).

The Dublin Core Element Set defines attributes, so called elements, which can be used for the description of electronic documents (Table 1).

Some of these elements can be qualified further. For example the "Date" element can be refined to "Created", "Issued", "Valid", "Available" or "Modified". Some elements can be refined by specifying the format or vocabulary used, for example, a qualifier can specify the scheme used to encode a date.

The Dublin Core standard itself does not specify a format (for example a XML schema) for the elements, usually the metadata are expressed either in HTML-documents (Figure 1) or in XML.

BACKGROUND

The definition of the semantics of Dublin Core elements is very flexible. Although, the DCMI recommends best practices for some elements, there is no definitive standard for their content. For example, the "Language" element should be specified according to RFC 3066 and ISO 639, which is a quite formal definition; but for the "Source" element it is recommended to "identify the referenced resource by means of a string or number conforming to a formal identification system" (DCMI, 2003), which gives each implementation a considerable amount of freedom regarding the system to be used. The advantage of these imprecise definitions is the resulting flexibility: Each implementation can use its own vocabulary specific to the actual application. The disadvantage is the lack of interoperability between applications which reduces the benefit of using a common standard.

The Dublin Core standard itself does not specify which elements are required and which are optional (all elements are optional, so an empty set of elements conforms to the Dublin Core). The selection and the concrete semantics of the elements to be used for a given application may be clarified by a so-called profile, for example, the DC-Library Application profile (DCMI, 2002) clarifies the use of the Dublin Core in library related applications. But profiles exist only for a small number of possible applications, so typically each implementation has to define its own semantics.

The Dublin Core element set mainly provides information about the content (for example "Title", "Description", etc.) and "Rights" related topics (for example "Cre-

Table 1. Dublin Core element set (DCMI, 2003)

Title	A name given to the resource
Creator	An entity primarily responsible for making the content of the resource
Subject	A 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 of an event in the lifecycle 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

M

Figure 1. Example of HTML encoded Dublin Core elements

```
<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="german" content="Meine homepage" />
<meta name="DC.creator" content="Till Haenisch" />
<meta name="DC.subject" lang="german" content="Vorlesungen; dlmeta />
<meta name="DC.date" scheme="DCTERMS.W3CDTF" content="2004-02-23" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.format" content="text/html" />
```

ator", "Publisher", etc.) of a resource. There is no (standard) way to include application specific information, for example, the timecodes of the different scenes in a video or the encoding quality of a MP3 file. This leads to a great variety of extensions, either specific for a single application or a class of applications. For example, Einhorn (2003) describes an extension for capturing metadata about presentations. DLmeta (see Abele, 2002) describes an extensible model for multimedia applications. These extensions enable applications to add necessary information to the metadata set, but since there is no common standard, this may lead to incompatibility.

Extending the Dublin Core results in a more complex set of metadata. A recent study (Ward, 2003) shows that even the unqualified elements are often used not completely: "two elements out of fifteen […] make up half the element usage in over half of the DPs [Data Providers]" (Ward, 2003). Nearly everybody (who uses Dublin Core) uses the "Title" and "Creator" elements. Some other elements (for example "Relation" and "Coverage") are rarely used. Maybe the original Dublin Core has too many elements (is too complex) for a lot of applications.

FUTURE TRENDS

There are two ways to solve the problems described above: Extend the element set or reduce it. Models like DLmeta (DLmeta, 2000) try to extend the Dublin Core to allow a more detailed description of electronic documents and their media specific properties and provide a certain degree of flexibility for specialized applications. Other models like the ABC-Model (Legoze, 2001) try to use a more abstract approach to allow the description of arbitrary objects and complex relationships, for example, in museum catalogs. The higher precision of the description when using these models may result in higher cost for the creation of the metadata and application development.

A different approach is to use a minimal common standard which is extensible according to the applications needs. One of the main reasons for using metadata is the ability to locate and retrieve resources. A minimal usable description should at least support this task.

Kunze (2001) suggests the use of a set of only 4 elements (Table 2) to describe the basic properties of a resource. If necessary, these elements could be furthermore qualified and extended.

One important reason for a common metadata standard is the interchange of metadata. A local application like a digital library may use a propriety metadata model unless this information should be shared with others, for example, to enable cross domain resource discovery. One solution to this problem is that every application uses the same metadata, another solution is a common minimal subset used (only) for the exchange of information. The Open Archives Initiative (Lagoze, 2001) uses this approach. Metadata is exchanged between data providers using Dublin Core but every provider may deliver additional XML-formatted metadata.

CONCLUSION

The trend to use a more complex metadata set resembles the way, library cataloguing techniques have developed from simple schemes to today's complex standards like MARC. The large number of electronic documents to be described requires efficient techniques for their description. The currently used Dublin Core standard has several deficiencies, but it is not clear if a more complex or a simpler standard should be used. Combinations of a small metadata kernel used for data interchange with optional application specific elements may result in the advantages of both directions.

REFERENCES

Abele, H., Frings, T., Hänisch, T., Aschoff, C. & Winter, C. (2002). DLmeta – noch nie waren Standards so wertvoll wie heute. *Praxis der Informationsverarbeitung und Kommunikation*, *25*(1), 39-43

Byrne, D. J. (1998). MARC Manual: understanding and using MARC (2nd ed.). Englewood, Colo: Libraries Unlimited.

1 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/metadata-electronic-documents-using-dublin/14539

Related Content

Thematic Analysis of User Experience of Contact Tracing Applications for COVID-19 Using Twitter Data
Mariam A. Alterkait, AlJawhara Owaid Almutarieand Manal Y. Alduaij (2024). *Journal of Cases on Information Technology (pp. 1-22).*
www.irma-international.org/article/thematic-analysis-of-user-experience-of-contact-tracing-applications-for-covid-19-using-twitter-data/335946

IS Project Management Contemporary Research Challenges
Maggie McPherson (2005). *Encyclopedia of Information Science and Technology, First Edition (pp. 1673-1678).*
www.irma-international.org/chapter/project-management-contemporary-research-challenges/14494

Endoscopic Imaging Results: Web-Based Solution for Video Diffusion with Real-Time Assistance
Joel Braga, Isabel Laranjo, Carla Rolanda, Luís Lopes, Jorge Correia-Pintoand Victor Alves (2014). *Journal of Information Technology Research (pp. 27-40).*
www.irma-international.org/article/endoscopic-imaging-results/111250

A New Algorithm for Subset Matching Problem Based on Set-String Transformation
Yangjun Chen (2009). *Encyclopedia of Information Communication Technology (pp. 607-615).*
www.irma-international.org/chapter/new-algorithm-subset-matching-problem/13412

Heuristics in Medical Data Mining
Susan E. George (2005). *Encyclopedia of Information Science and Technology, First Edition (pp. 1322-1326).*
www.irma-international.org/chapter/heuristics-medical-data-mining/14432