METADATA for ECAI

Helen Jarvis, University of New South Wales, Australia

Abstract

The paper will discuss the meaning of the term "metadata" as it has evolved in the information management world, particularly for application in the World Wide Web environment. Particular attention will be paid to the Dublin Core standard, still under development by international working groups, but gradually formalised at a series of six workshops held since March 1995. Within the Dublin Core fifteen elements are defined, and within each element various options are permitted -- either alternative schemes to be followed, or more detailed sub-elements to allow greater precision. The Dublin Core is proposed for adoption by ECAI as the standard method of describing and codifying electronic resources. The Australian ECAI Working Group (OzECAI) has during 1998 developed a proposed ECAI Metadata Standard, selecting particular schemes and sub-elements from Dublin Core, and adding several elements. Certain elements are proposed as mandatory for resources to be included in ECAI. The paper will also, more briefly, discuss other metadata schemes being developed, and some outstanding areas of discussion.

A new word has recently entered into the daily vocabulary of many of us working in the field of documentation: metadata. The word 'meta' means 'together with', 'after', 'transcending' as in the compounds 'metaphysics', 'metamorphosis' and 'metatarsal'. Metadata is commonly defined as meaning 'data about data' or 'information about data'.

The term metadata has been used in the worlds of data modelling and information systems world since at least the 1960s. And, as a concept (although known by other terms), it has a long history in the worlds of cataloguing and indexing, being represented as entries or records that serve as surrogates for items, carrying both information derived from the items themselves (such as title, author, publisher, physical description) as well as information supplied by cataloguers or indexers (such as subject headings, classification number and accession number).

Metadata has come into the vocabulary of the library and information management world only since the development of the Internet, especially with the World Wide Web in 1994, and the consequent proliferation of electronic data available to the searcher, accompanied by "serious problems associated with Internet retrieval, of which massive recall, coupled with an

equal lack of precision, is arguably the worst". [1]

Different ways have been pursued to assist the identification of desired data in this new environment. The most well known are firstly, the development of search engines with their varying retrieval algorithms, as well as multiple and integrated search tools; and, secondly, the creation of subject directories, gateways and referral tools. [2]

A third approach is to enhance the data, or the "documents" themselves, by provision of metadata to improve discovery and recovery (over what search engines alone can deliver) by:

+ identifying significant terms in and about the document
+ tying them to specific roles (such as creator, title)
+ allowing terms to be associated with non-HTML text (such as images or sound)
+ deriving terms from lower level pages/documents on a site

The Dublin Core

Of particular interest to us in ECAI, as to many others seeking the most appropriate metadata protocol to follow to improve retrieval of information resources on the World Wide Web, is the Dublin Core.

In March 1995 at a workshop held in Dublin, Ohio sponsored by the Online Computer Library Center (OCLC) and the National Center for Supercomputing Applications (NCSA), the first draft of what has become the Dublin Core Metadata Element Set was defined. It is being developed as a generic metadata standard able to be applied both by the many and varied creators of WWW resources themselves, and by third parties, such as librarians and other information managers, who have a long tradition of following standards and formal models and who are aware of the complexities of describing bibliographic and other documentary data.

The tension that exists between the objectives of meeting the needs of these two communities of users has been reflected in two different notions and expressions of the Dublin Core --

---

[1] *The Nordic metadata project : final report* / Juha Hakala [et al.]. Helsinki : Helsinki University Library, July 1998. ISBN 951-45-8248-9. http://linnea.helsinki.fi/meta XXX (Accessed November 1998) 1. Introduction.

[2] For a succinct analysis and comparison of these approaches, see Maureen Henninger, *Don't just surf : effective research strategies for the Internet*. Sydney : UNSW Press, 1997.

simple and complex. [3]

Over the four years that have elapsed since that initial Dublin Core meeting, five more workshops have been held, and a series of Working Groups have been active to varying degrees so as to refine and define the Dublin Core standard, and to develop tools to facilitate its application in various environments.

Dublin Core Workshops

1st	Dublin	March 1995	13 elements defined
2nd	Warwick	April 1996	Warwick Framework & syntax
3rd	Dublin	Sept 1996	15 elements & including images
4th	Canberra	March 1997	Canberra Qualifiers
5th	Helsinki	October 1997	DC profiles for RDF & Z39.50
6th	Washington	November 1998	interoperability among DC implementations

Responsibility for the maintenance of the archival record of this process, and for posting the latest version of the standard, as well as numerous discussion papers, links to related tools and user guides is now held by OCLC in Columbus, Ohio.
<http://purl.org/DC>

Another important source of ongoing information about the development of the Dublin Core is the online journal *D-Lib*, which in January 1998 published an article by Harold Thiele reviewing the literature on the Dublin Core from 1995 to 1997[4].

The Dublin Core Metadata Element Set consists of 15 elements, which may be seen as forming three categories[5], as shown below:

---

[3] Warwick S. Cathro, "The Dublin Core : simplicity or complexity", paper presented at the ALIA Information Science Section Metadata Seminar, 30 July 1997, Sydney, State Library of New South Wales; see also Roger Clarke, *Beyond the Dublin Core : rich meta-data and convenience-of-use are compatible after all.* (Version of 11 July 1997)
<http://www.anu.edu.au/people/Roger.Clarke/II/DublinCore.html>
[4] Harold Thiele, "The Dublin Core and Warwick Framework", *D-Lib Magazine,* January 1998.
<http://www.dlib.org/dlib/january98/01thiele.html>>
[5] Stuart Weibel and Juha Hakala, "DC-5: The Helsinki metadata workshop", *D-Lib magazine,* February 1998. <http://www.dlib.org/dlib/february98/02weibel.html)

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

Each Dublin Core (DC) element may be left unqualified, with values provided in free text (in its simple manifestation), or each may be qualified for either or all of three purposes:

+ to refine or narrow the semantics of an element (by declaration of a sub-element or type) – eg. Date.Created
+ to explain or specify a context for the interpretation of an element or sub-element (by declaration of a scheme) – eg. Subject Scheme LCSH
+ to specify the language of the element value or content data (by declaration of a language)

Dublin Core Metadata has been specifically developed for application in two different ways:

+ embedded in the resource itself; and/or
+ maintained as external metadata in an independent repository, where it may be used to describe materials that are themselves not electronic, such as museum objects.

Where embedded in the resource, specific Dublin Core syntaxes have been defined for HTML utilising the META tag as part of the HEAD of an HTML document, eg.
<META NAME=DC.Title" CONTENT="ECAI – Berkeley Home Page">
(with slight differences in syntax between HTML versions 2.0 and 4.0).

One of the most important recent developments in the Dublin Core is its involvement in the Resource Description Framework (RDF)[6], now adopted as a draft by the World Wide Web Consortium, which will permit "the creation of metadata records comprising multiple metadata 'sets' compiled from different cataloguing paradigms" and at different levels (such as a collection and individual items within it). This work is also leading to the development of a formal data model for the Dublin Core.[7]

---

[6] *Resource Description Framework (RDF) model and syntax.*
  <http://www.w3.org/RDF/Group/WD-rtf-syntax>

[7] *ibid*

ECAI Metadata

At the Third ECAI Workshop held in Heidelberg, Germany on 29-30 June 1998, it was decided that all ECAI resources would be required to have metadata applied, and it was also decided to adopt the Dublin Core as the basis for the ECAI Metadata Element Set.

Two meetings of the Australian ECAITech Working Group (OzECAI) have since been held (in August and in November 1998). These meetings have devoted considerable attention to deciding which of the optional forms of Dublin Core should be adopted by ECAI, and it is proposing an ECAI Metadata Element Set is being proposed for adoption at the Fourth ECAI Workshop in Taipei in January 1999.[8] This set prescribes certain elements as mandatory, and it selects various sub-elements and schemes for certain elements.

It is this proposed ECAI Metadata Element Set that is being used as the basis for the ECAI Metadata Repository utilising the Time-Map Metadata Manager.[9] The Metadata Tutorial held in Taipei on 15 January 1999, used the proposed ECAI Metadata Element Set and Time-Map Metadata Manager as the operational model for instruction.

Other relevant metadata standards and projects

In addition to Dublin Core, a number of other metadata standards and projects have relevance to the interests of ECAI, and therefore deserve our attention, either merely to broaden our understanding of what is encompassed by metadata or, more pressingly, to be considered for inclusion within our ECAI Metadata Element Set. OzECAI looked at a number of these in its meetings and online discussions during 1998.

Of particular interest to ECAI, given the decision that all ECAI resources are to be georeferenced[10] are the specific metadata standards that have been developed for handling geographic data. As long ago as 1994,[11] all US government agencies were mandated to follow

---

[8] The document "Proposed ECAI Metadata Elements" was prepared by Nereida Cross and Helen Jarvis from the School of Information, Library and Archive Studies at the University of New South Wales, and revised in collaboration with other OzECAI members.

[9] The TMM has been developed by Ian Johnson at the Archaeological Computing Laboratory at the University of Sydney.

[10] Third ECAI Workshop, Heidelberg, June 1998.

[11] US President Executive Order 12906 "Coordinating geographic data acquisition and access : the national spatial data infrastructure".

the Content Standard for Geospatial Metadata[12], developed by the Federal Geographic Data Committee (FGDC) from 1992, and now forming the basis for the draft ISO Metadata Standard for Geographic Information[13]. This FGDC Content Standard has defined a huge number of elements to cope with an extraordinary amount of detail. These 220 elements are then grouped into ten compound elements.

The Australian and New Zealand Land Information Council (ANZLIC) consciously decided to be "less ambitious" and yet consistent with the FGDC Content Standard. The ANZLIC Core Metadata Elements[14] define 30 elements in nine categories.

While both the FGDC and ANZLIC standards were seriously considered by OzECAI for recommendation as the ECAI standard due to their detailed exposition of elements relating specifically to geospatial data, on closer consideration we felt that the needs of ECAI could actually be met adequately within the Dublin Core Coverage element, using its Time and Place sub-elements with (if considered necessary) some local ECAI additions. It was felt that the Dublin Core more adequately catered for the many non-geographic elements needed by ECAI. Further, the very simplicity of the Dublin Core approach, and its widespread utilisation in the library, and museum communities, meant that it would be more practical for all the ECAI participants to use.

OzECAI also carefully considered another metadata standard developed in the US government archive community, known as Government Information Locator Service (GILS)[15], highly elaborated in the areas of rights and provenance but, in our opinion, not superior to the Dublin Core, when viewed overall, at least from our perspective. Similar conclusions were reached by the Australian government archives community, as can be seen by its adoption of Dublin Core as the basis of its Australian Government Locator Service (AGLS)[16], with the

---

[12] *Content Standard for Geospatial Metadata*, Washington, DC : Federal Geographic Data Committee, 1994. <http://www.fgdc.gov/metadata/contstan.html>

[13] ISO 15046-15 prepared by ISO Technical Committee on Geographic Information (ISO TC211).

[14] *Core Metadata Elements*, Canberra : Australian and New Zealand Land Information Council, 1995. <http://www.anzlic.org.au/metaelem.htm>

[15] *Application profile for the Government Information Locator Service (GILS).* Version 2, 24 Nov 1997 <http://info.er.usgs.gov/gils/prof_v2.html>

[16] *The Australian Government Locator Service (AGLS) manual for users.* [Canberra] : Office of Government Information Technology [and] National Archives of Australia, 1998. (Version 1998-07-27) <http://www.naa.gov.au/govserv/agls>

inclusion of two further elements, namely Function Descriptor and Availability.

In selecting the Dublin Core as the basis for our metadata element set, ECAI is following the lead of several international evaluation and comparison studies. Project DESIRE, coordinated by the UK Office of Library Networking, in 1996 published *A review of metadata: a survey of current resource description formats*, which considered 22 different formats (including FGDC, GILS, MARC, Text Encoding Initiative Independent Headers, CIMI (Computer Interchange of Museum Information), Encoding Archival Description (EAD), and IAFA/whois++ Templates. Their report concluded that Dublin Core was a good choice. This point was developed further by the BIBLINK project working to establish relations between national bibliographic agencies and publishers of electronic material. Its metadata comparison report concludes:

"If we are looking towards use of a core then the Dublin Core element set is an obvious choice. There is international involvement in the consensus building, and project participants could influence the development of the format. It is a format that small publishers and web publishers could use without incurring significant overhead."[17]

The Nordic metadata project, which was the first international project to embark upon building tools based on the Dublin Core, issued its final report in July 1998, stating: "The experience from the Nordic metadata has shown that the applications and documentation required for Dublin Core production and use can be developed very effectively in international co-operation, and then adapted according to the local needs for national and regional projects."[18]

Other metadata applications

Aside from resource discovery, a number of other potential applications of metadata have been identified and developed by various special interest groups. These include the flagging of content and quality according to various criteria -- especially violence or pornography -- used widely in site filtering software; of optimum browser/viewer settings for the data in hand; and of verification and authentication information, sometimes known as the data' s "digital signature".

A summary of the range of metadata applications and differing protocols developed to cater

---

[17] "A review of metadata : a survey of current resource description formats", UK Office of Library Networking, 1996. <http://www.ukoln.ac.uk/metadata/desire/overview>

[18] *Nordic metadata,* op cit, 1.2.1 "Evaluation of existing metadata formats"

for them, is given in the following table:[19]

Need	Examples of Protocol
Resource discovery	Dublin Core, GILS (Government Information Locator Service)
Filtering	PICS (Platform for Internet Content Selection)
End user navigation	Web Collections, MCF (Meta Content Format)
Legal	Dsig (Digital signatures)
Multiple formats	TCN (Transparent Content Negotiation)
Authoring	WebDAV

As these various metadata protocols have been developed, new coalitions of interest and working groups have been formed, bringing together the spheres of expertise and interest of what might be called the curatorial and documentation professions (librarians, archivists and museum curators) with those involved in network information management. In the past these two spheres have had little cross-fertilisation and, although they have shared many underlying concepts and have often embarked upon similar undertakings, their language and approach have often differed markedly, so that this new cooperation can be seen as a merging of different cultures.[20]

We have seen this, too, within ECAI where scholars from many different disciplines have come together to share their expertise and perspectives with the intention of establishing something quite distinctive, and certainly greater than the sum of the many parts. The experience of deciding to adopt metadata, and then examining a range of possible standards to see what best suits our needs has engaged geographers, archaeologists, and information and documentation specialists, each with their own subject and area studies background as well as technical expertise. And the metadata process is but one facet of the cross-disciplinary work that characterises ECAI.

Acknowledgements

Professor Lewis Lancaster should be recognised for his approach in charting ECAI' s mission, and his constant encouragement and enthusiasm for the various activities and initiatives that

---

[19] Adapted from Brian Kelly, "Potential Applications of Metadata," as cited in Neil McLean, "Metadata : the search for a new order", Paper presented at the ALIA Information Science Section Metadata Seminar, 30 July 1997, Sydney, State Library of New South Wales.

[20] McLean, op cit.

have been carried out in different parts of the world over the past several years. Without his vision, most of us would not be here today.

Acknowledgement should also be made to Academia Sinica and the various institutions in which we work -- it is fair to say that ECAI may not be the highest priority for any of them, yet they have provided us with the infrastructure and time to engage in this significant international collaborative research project.

I would like to conclude this paper by thanking the other members of the OzECAI group for their input to our deliberations on metadata for ECAI, especially Nereida Cross who was primarily responsible for preparing the ECAI Metadata Manual and the tables of comparison and "crosswalks" between the various metadata standards used in our deliberations, and also Peter Moloney, Research Assistant in the School of Information, Library and Archive Studies of the University of New South Wales for hunting and gathering metadata documentation from the Internet.