

Metadata in the University Environment

Kerrie Talmacs
Metadata Co-ordinator
University of New South Wales Library
k.talmacs@unsw.edu.au

Abstract

Universities need to consider metadata as a means of managing and providing access to information resources. A considerable commitment is required to assess standards; train staff and utilise metadata appropriately. The University of New South Wales Library has made a start and has embedded metadata in Library Web pages, created metadata for subject gateways – MetaChem and AVEL, and is generating metadata for the Australian Digital Theses Project.

Introduction

It is becoming essential for universities to consider using metadata to manage, and provide access to, their information resources. University communities depend on excellent, precise information systems to support their teaching, learning and research functions. Librarians, with extensive experience in creating metadata for their own Library collections, are well equipped to play a leading role in assisting resource discovery in the digital age. There is general agreement in libraries that traditional MARC cataloguing cannot be applied to all, or even a substantial part of, Web resources. Universities have started searching for solutions to effective information retrieval, both in terms of their own data gathering, and of leading other inquirers to their Web resources. The UNSW Web Co-ordinator, for example, spoke at a "Cybermarketing Conference" on how to increase the chances of Web sites being found by search engines, and advised that keywords and descriptions are the most commonly recognised metatags. Keywords "can be semantic links between uncommon phrases mentioned in the body of the document, or they can be other words that give context to the body of the document or words that are not mentioned in the body of the document" (Jacka 11). Metadata descriptions can also assist in clarifying the contents of documents for the searcher. Other metadata may include rights information, relationships with other resources and authorship details. Much has been written about the shortcomings of search engines in returning precise results to searchers. But can metadata really fill the gap between search engines and full MARC cataloguing?

The success of metadata will depend on its uptake by Web developers and the support provided by the search engines. Already specially developed search engines such as HotMeta (a product of the Distributed Systems Technology Centre at the University of Queensland) support the standard metadata used widely in Australia which is largely based on the Dublin Core schema. Several of the Australian subject gateways developed co-operatively by university libraries rely on such metadata and supportive search engines. University Web sites are starting to add metadata, sometimes tentatively, and often relying on staff with a huge range of other duties and limited metadata expertise. The other obstacle for Web developers and information managers is the unfinished nature of the metadata schemas which are still being developed and fine tuned to meet the requirements of many disciplines and subject areas. Universities no longer have time to wait until schemas are more "finished". Cataloguing rules for description of, and access to, traditional resources took decades to develop, and have been evolving to keep up with the new information materials, user needs and technology. Metadata standards will also evolve and change, even after adoption. Successful metadata implementation depends on the formation of effective alliances and networks across areas of expertise (e.g. cataloguers, indexers, Web technicians, subject experts, Web site creators), and in the broader metadata arena across sectors (e.g. types of libraries, museums, government organisations, educational institutions).

University of New South Wales

At the University of New South Wales changes such as the formation of the Division of Information Services, the creation of the Electronic Information Resources Group within the Library, and the new positions of University Web Co-ordinator, Library Web Co-ordinator and the Metadata Co-ordinator, have assisted the implementation of metadata. They have provided effective working teams and a suitable mix of skills to face the challenges of Web site development, and provision of access to the University's Web resources. Because of the new structure and positions, the University is well positioned to evaluate metadata, and able to keep up-to-date with developments. Metadata is the key to many internal (e.g. Library Web page creation) and national projects (e.g. Australian Digital Theses).

Management within the Division realise the importance of metadata in assisting access to Web resources and prefer a distributed model of application, ideally involving Web developers and those people with special subject expertise. This model will depend on central co-ordination by some of the staff mentioned above. Their duty statements specifically focus on metadata and they are expected to keep up-to-date and provide guidance as necessary. For example, the current initial attempts to develop a University thesaurus are being co-ordinated by the UNSW Web Co-ordinator, with input from the Metadata Co-ordinator and two cataloguers. There will be an extra burden on UNSW staff in the creation of metadata, just as there has been for them to establish Web sites. Universities and other organisations have had to do these extra tasks without extra resources. It really amounts to: "Can we afford not to do it?" This has created a dilemma in organisations already struggling to handle traditional, established activities. At the same time there is a real need for universities to come to terms with the relevant metadata standards. While Dublin Core is relatively straightforward and simple, it has the disadvantage of not yet being fine-tuned for the educational sphere. IMS (Instructional Management Systems) metadata, on the other hand, while tailored for the educational domain, brings with it a complexity of more elements and sub-elements to learn and apply.

The critical success factors for the University include the dedication of relevant staff to the implementation of metadata, the dissemination of expertise, experimentation with the new and evolving standards, the "learning by doing" approach and the support from management who have recognised the potential of metadata. Involvement in metadata projects has been possible through some staff being able to dedicate time to the work and to then spread the knowledge through training – mostly in the context of specific projects. Management have been prepared to commit staff to projects, which have involved exploring new territory. Much of the exploratory work has been in the collaborative environment through national projects and has allowed for the benefits of an increased array of expertise to move things forward. The University has stood by its commitments, largely because of the staff dedicated to this new work and the confidence acquired through taking small initial steps (e.g. starting with adding Library metadata).

Getting Started with Metadata

With a long-established role in assisting the university community to access information, the Library was chosen as the test-bed for inclusion of metadata in Web pages. By the end of 1998 the Library had decided that metadata would be mandatory for new Web pages, and that all current pages would be upgraded to include metadata by the beginning of the 2000 academic year. Standards were settled, training provided, and the librarians creating Web pages set about their new task of adding metadata.

As with any new field of work, it was critical to research the topic thoroughly and to talk with others who had some experience. Two contacts for my work during the first months were Jennie Thornely, the leader of the metadata project at the State Library of Queensland and Debbie Campbell, then the Metadata Co-ordinator at the National Library. The State Library of Queensland documented their early decisions on the Web, and this information tracing their project through the various stages was invaluable. Debbie Campbell has been available as a consultant regarding the MetaWeb products, Dublin Core standard and subject gateway metadata development. The University of Queensland Library has also added very comprehensive metadata to their Web site, and Chris Taylor, Manager, their Information Access and Delivery Service has written an excellent overview of metadata (1999).

Library Web Pages

The main issues in the Library metadata work centred on the choice of a schema and standards including the decision whether to use a thesaurus, selection of a template, adoption of policies/priorities, deciding who would add the metadata, and the creation of a metadata Web resources page. Management decided that metadata would be added by those librarians creating the Web pages, i.e. the experts in the content. They would need to be trained in the use of the selected schema's elements. Dublin Core was a natural choice of schema with the following characteristics in its favour: simplicity of creation and maintenance, commonly understood semantics, international scope and extensibility. Next came the need to settle on mandatory and optional elements, with the assumption that there would be some defaults and automatic inclusions. The Library Web Co-ordinator and the Metadata Co-ordinator proposed nine core elements which were approved by management. The most controversial decision of this process was the choice of natural language keywords in the "Subject" element, with some staff preferring Library of Congress Subject Headings. Natural language keywords have remained our choice to date, with LCSH considered difficult to access (in the distributed work pattern), complicated to use, sometimes inappropriate with its American terminology, and often out of date. Specialised thesauri were encouraged where appropriate. The other six Dublin Core elements are optional, giving the metadata creators some leeway. We also use a few qualifiers. While Dublin Core qualifiers are still being decided, we settled on a rather conservative approach, realising that global changes might still need to be made in the future.

Tools of the Trade

The UNSW Library's application of Dublin Core is spelt out clearly in its metadata template (UNSW Library). The Library Web Co-ordinator found the Nordic Project template (Koch and Borell) the easiest to adapt to our needs, and the Metadata Co-ordinator added element explanations based on current Dublin Core standards. This tool allows for speedy metadata creation (approximately 10 minutes per record), with Web developers concentrating on the metadata content without having to grapple with the syntax. The metadata at the time of creation is automatically sent to the Metadata Co-ordinator via e.mail for perusal and quality control. The standard has been high, and this can be attributed to the suitability of template and Dublin Core's simplicity. Librarians are also very familiar with the fields - similar in many respects to those used in traditional cataloguing. Laboratory training sessions of approximately one to two hours included some background explanation and demonstration and hands-on exercises. This was provided to all interested staff, especially Web page developers. A Library metadata resource page has also been created. A critical issue is how well the metadata can be kept up-to-date as Web pages change and come and go. This is, of course, similar to the maintenance work in any catalogue, but is aggravated by the very nature of Web resources. Needless to say, the metadata will only be as good as its currency.

UNSW Campus

To assess the support for metadata across campus the University Web Co-ordinator and the Metadata Co-ordinator surveyed UNSW Web developers. First indications are that they support metadata in principle and the need for training, but are unsure of their ability to commit resources to the task. Campus metadata application might be enhanced through central standardisation of terms in a UNSW index, allowing staff in the schools and faculties to select from a site thesaurus. Some initial work by two cataloguers has identified several hundred preferred terms from a list of actual terms used to search the University over the last 12 months. Other options include automatic metadata generation using tools such as the MetaWeb Project's generator (1998). The results from the generator are less than optimal and only provide six of the fifteen Dublin Core elements, but such a solution may be better than doing nothing when staff resources are not available. The UNSW search engine - Infoseek's Ultraseek, supports Dublin Core metadata so will be able to take advantage of any added metadata. Whatever happens in the long run, the approach is certain to be a hybrid model.

While the emphasis for most of 1999 has been on how Dublin Core might be used within the University, there has been a growing interest in IMS (Instructional Management Systems) metadata. A visible sign of this at UNSW has been the development of the Universitas 21 Learning Resource Catalogue template established by the director of the UNSW's Educational Development and Technology Unit. The Library is assisting the development of the template by adding some of its publications, and in the longer term may assist academics to create metadata for theirs. This catalogue will list courseware and will be based on IMS metadata (IMS 1999) The latter includes a subset of Dublin Core metadata, with additional elements for the education domain. An agreement between Dublin Core and IMS struck at the DC Frankfurt meeting in October 1999 clarified the co-ordinating and promotional

role for IMS in the area of educational metadata. The implications for practitioners should become clearer during 2000 as implementation of both schemas proceeds.

The application of metadata within the University is thus a co-operative venture, involving librarians from a variety of departments with different skills, the University Web Co-ordinator and a variety of staff in the University faculties. The intention is that the metadata work is carried out in a distributed fashion, with some reliance on central expertise initially and during the introduction of new standards. In most cases, as mentioned above, the work will amount to an additional workload on University staff.

National Projects

The UNSW Library is involved in several projects with other university libraries – MetaChem (a catalogue of chemistry resources), AVEL (Australasian Virtual Engineering Library) and the Australian Digital Theses (ADT) Project. Not only do these projects provide excellent opportunities for collaboration between the university libraries involved, but they allow the participants to jointly come to grips with merging metadata standards. Metadata is an integral part of these databases, providing indexes for searching and resource discovery. The HotMeta search engine, developed out of the MetaWeb project by the Distributed Systems Technology Centre fully supports the metadata schemas based on Dublin Core. The MetaChem and AVEL projects employ subject or cataloguing experts who critically evaluate resources for inclusion. The ADT metadata is, however, automatically generated from submission details according to specifications provided by the Metadata Co-ordinator who received advice from the National Library's Metadata Co-ordinator. The National Library's broader national perspective has proven invaluable because their experience in national projects and the work with the DSTC and other organisations. The DSTC/NLA partnership, demonstrated in the MetaWeb project, and through other work such as the development of A-Core metadata schema and the Australian Government locator service (AGLS) metadata schema, has provided universities and other organisations with strong metadata support.

MetaChem

MetaChem, lead by the UNSW Library, is a national gateway to selected chemistry resources, with an emphasis on Australian material. Metadata creators in the participating universities were trained in the MetaChem elements - Dublin Core with three additional elements, and the use of the DSTC editors -Reg and Reggie. Over 600 records have been added to date, and these form the MetaChem database. Searchers enter keywords, with the option to search all metadata elements or a selected element. Searchers can also browse a resource's metadata by entering the URL, and can thus view the full description to help them evaluate resources. DSTC has further enhanced the searching interface with refinements such as related subjects and more specific topics appearing in the left-hand column beside search results.

As with all subject gateways, the focus is on selected resources. Users' expectations will only be met if selection guidelines are adhered to and standards are maintained. For example, URLs need to be kept up-to-date and metadata needs to be current. Issues for MetaChem include: long-term maintenance and quality control, the

possible enhancement of the search interface to include a browse option, and the adoption of a MetaChem thesaurus

AVEL

The University of Queensland Library in partnership with four other university libraries, the Institution of Engineers, Australia (IEAust), the Centre for Mining Technology & Equipment (CMTE) and the DSTC, established the Australasian Virtual Engineering Library (AVEL) in 1999. AVEL, a subject gateway for engineering and information technology (IT) resources, works in collaboration with Edinburgh Engineering Virtual Library (EEVL) and the National Library of Australia. Unlike MetaChem, AVEL has its own thesaurus – an amalgam of segments of two existing thesauri and browse and search options. Like MetaChem it uses DSTC's editor, Reg, the HotMeta search engine and the metadata is an augmented Dublin Core. For the UNSW contribution to the project this year, I trained a reference librarian who has special expertise in engineering resources to create metadata. It appears that this pattern of a subject specialist being trained by a cataloguer will prevail in the initial phase of getting metadata work underway. The AVEL Web site outlines the aims of the gateway as a pointer to quality relevant Web-based materials. Engineering professionals will be able to easily index and publish their work on the Web, gaining world- wide exposure. The benefits of AVEL will be: “ Improved sharing of information between industry and university researchers, being a partner in a global network of WWW gateways to engineering and information technology (IT) resources, and building national, regional and global R & D networks between universities, industry and research establishments” (AVEL 1999).

This gateway project is proceeding smoothly with an active project leader and a tight-knit virtual team. The AVEL Team is currently investigating the technical requirements and infrastructure for the provision of new services and the maintenance of the database. In addition to this a business plan is being drafted to explore opportunities for future partnerships.

Australian Digital Theses Project

Another co-operative university library venture headed by the UNSW Library is the Australian Digital Theses Project (1998). The project is modelled on the US Networked Digital Library of Theses and Dissertations led by the Virginia Polytechnic Institute and State University. The UNSW Library is working in partnership with six other university libraries to trial the electronic deposit of theses volunteered by students. This is taking place in parallel to the traditional submission of printed theses, with the eventual long-term aim of having all theses submitted electronically, thereby improving access to, and transfer of, research information. The shorter-term objective of the project itself is to establish a distributed database of the selected theses at the participating institutions. The Dublin Core metadata forms the index for searching for the theses and is generated automatically. Although this is a limited application of metadata - nine elements, with only a small number of DC qualifiers, a useful index is very easily and efficiently created to enable access to a wealth of research information. The automatically generated metadata could possibly be enhanced in the future by other metadata e.g. from full MARC cataloguing. For example, Library of Congress Subject Headings from catalogue records could provide

additional terms for subject searching (currently keywords provided by students are used). The ADT project could also be linked in the future to the subject gateways for the purpose of sharing of metadata. Relevant theses might be identified through nominated classification number ranges (e.g. Dewey, LC). The projects might, therefore, become more connected in the future.

EdNA Metadata

The UNSW Metadata Co-ordinator is the CAUL representative on the EdNA (Education Network Australia) Metadata Reference Group. The group is co-ordinated by Education.Au, with representatives from each key EdNA stakeholder group. It has been reconvened to progress changes to the EdNA metadata standard (EdNA 1999), taking into account the ability of Dublin Core to accommodate educational metadata requirements, the changes to the AGLS (Australian Government Locator Service) metadata standard (AGLS 1999), and the impact that IMS Instructional Management Systems (IMS) activity is likely to have in Australia. A significant step in the development of metadata for use with educational resources was the formation of the Dublin Core working group for educational materials. Co-chaired by EdNA's Jon Mason and Stuart Sutton from the GEM project (Gateway to Educational Materials), the group will develop suitable qualifiers and extensions to Dublin Core with a final draft of DC educational metadata due in April 2000.

IMS Metadata

Instructional Management Systems (IMS) metadata development culminated in the approval of a final IMS metadata specification in August 1999. This consists of the IEEE Learning Object Metadata Scheme, the IMS Learning Resource Metadata XML Binding Specification and the IMS Learning Resource Metadata Best Practice and Implementation Guide. Australian input has focussed on the Australian interest in Dublin Core and the need for its relationship with IMS metadata to be woven into the Best Practice and Implementation Guide. A partial accommodation of DC in the outcome was secured at the August 1999 IMS meeting, with some recognition of the importance of DC to Australia. The relationship continued to be defined at the DC Frankfurt meeting in October (mentioned above), whereby the schemas will be developed in harmony with each other.

Recordkeeping Metadata

Metadata is an integral part of the process of managing and preserving records in the electronic environment. "Essentially the purposes of recordkeeping metadata are to:

- identify
- authenticate
- ensure the persistence of content, structure, and content
- administer terms and conditions of access and disposal
- document use history, including recordkeeping processes

- enable discovery, retrieval and delivery to authorised users
- restrict unauthorised use and
- enable interoperability with related metadata standards”
(Cumming 3-4)

The Australian SPIRT (Strategic Partnership with Industry – Research and Training) Project which aims to develop a framework for standardising recordkeeping and archival metadata, supports the need for interoperability with generic standards such as Dublin Core and metadata initiatives regarding information locator systems such as the AGLS (Australian Government Locator Service). The project subscribes to the continuum of records where records are active components of business processes. “The recordkeeping perspective links the dynamic world of business activity to the passive world of information resource”. (SPIRT 3). Metadata is applicable to documents throughout their entire life span not just at one moment in time. The project, based at Monash University, involves collaborators from other key recordkeeping organisations and universities, including the University of New South Wales.

The National Archives of Australia published “Recordkeeping Metadata Standard for Commonwealth Agencies” in May 1999 to “help agencies to identify, authenticate, describe and manage their electronic records in a systematic and consistent way to meet business, accountability and archival requirements” (NAA 1). The standard has been developed to take into account the AGLS and SPIRT standards, and there is overlap amongst the three. The NAA standard is in line with AS 4390.

The UNSW is in the process of purchasing a new electronic recordkeeping system. Metadata will be a vital component and the relevant standards mentioned above will need to be evaluated for their suitability. It is fortunate that such progress has already been made on Australian recordkeeping standards.

Possible Implications for Other Organisations

It may be possible to draw conclusions from the UNSW experience for other universities and organisations interested in implementing metadata. Suggestions are as follows:-

- Free up staff to learn the standards and select appropriate ones for the organisation
- Adhere to standards until implications of adding extensions/qualifiers are understood
- Take small steps at the beginning
- Use available expertise e.g. of cataloguers for training others
- Keep applications as simple as possible

- Be prepared to experiment if necessary
- Get management on side if they're not already

Conclusion

It is clear, then, that metadata is becoming an essential component of university information systems. While the goals and philosophies of the projects mentioned above vary, there is considerable overlap in their metadata requirements. Sometimes automatic generation is possible, as in the case of the Australian Digital Theses Project, while in most other cases such as Library Web resources metadata or the metadata for the subject gateways there is substantial human involvement. The latter involves training in metadata elements, and the use of appropriate tools. This requires considerable involvement and commitment to metadata – to remain aware of the standards, to keep up-to-date with their development, and to form networks of experts to bring theory to reality. It seems that Dublin Core metadata does, and will, provide a very firm basis, with its international acceptance, extensibility, simplicity of creation, and commonly understood semantics. Yet even Dublin Core is still not stable beyond the basic 15 elements. The qualifiers will need to be confirmed if metadata editors are to work with any certainty. This process will take some time, but it is hoped that the year 2000 will bring with it some resolution of key issues. Standards such as Dublin Core, AGLS, EdNA, IMS and A-Core need to be harmonised, and there is already much progress in this area. Work at the University of New South Wales has relied heavily on consultation with others working in the field and constant vigilance regarding developments. It is work that is suited to collaboration, and this is facilitated by easy global communication. Even as standards are being formed and developed, it is imperative that universities make a commitment to metadata and prepare themselves for their involvement as the need arises. The University of New South Wales has found the Library metadata project a suitable starting point. Librarians already have a lot of experience with metadata elements, and are well placed to inform others working on new projects. Critical to the process is the gaining of confidence through getting some first results, and realising that metadata is not really new after all. It's just that the standards need to vary with different applications.

Bibliography

- Australian Digital These Project. (1998). *Australian Digital Theses Project* [Online]. Available: <http://www.library.unsw.edu.au/thesis/thesis.html> [1999, September 7]
- AVEL. (1999). *Australian Virtual Engineering Library (AVEL)* [Online]. Available: <http://avel.library.uq.edu.au> [1999, September 7]
- Australian Government Locator Service. (1999). *The Australian Government Locator Service (AGLS) Manual for Users* [Online]. Available: http://www.naa.gov.au/govserv/agls/user_manual/cover.htm [1999, December 8]
- Consortium for the Computer Interchange of Museum Information. (CIMI) (1999). *Guide to Best Practice: Dublin Core* [Online]. Available: http://www.cimi.org/documents/meta_bestprac_final_ann.html [1999, December 9]
- Cumming, Kate. (1999). *Maintaining the Viability of our Electronic Heritage: Recordkeeping Metadata and the SPIRT Metadata Project*. Unpublished.
- EdNA (Education Network Australia) (1999). *EdNA Metadata Homepage* [Online]. Available: <http://www.edna.edu.au/EdNA> [1999, December 8]
- Instructional Management Systems (1999). *IMS Meta-data Specification* [Online]. Available: <http://www.imsproject.org/metadata/index.html> [1999, December 8]
- Jacka, Rod. (1998). *Effectively Listing a Web-Site to Increase its Chances of Being Found* [Online]. Available: <http://www.webcoord.unsw.edu.au/searchengines/searcheng2.html> [1999, September 7]
- Koch, Traugott and Borell, Mattias. (1998). *Dublin Core Metadata Template* [Online]. Available: <http://www.lub.lu.se/cgi-bin/nmdc.pl> [1999, September 7]
- MetaWeb Project. (1998). *MetaWeb – the Metadata Tools and Services Project* [Online]. Available: <http://www.dstc.edu.au/Research/Projects/metaweb/> [1999, September 7]
- National Archives of Australia. (1999). *Recordkeeping Metadata Standard for Commonwealth Agencies* [Online]. Available: <http://www.naa.gov.au/govserv/techpub/rkms/intro.htm> [1999, September 17]
- Strategic Partnership with Industry – Research & Training (SPIRT). (1999). *Recordkeeping Metadata* [Online]. Available: <http://www.sims.monash.edu.au/rcrg/research/spirt/index.html> [1999, September 17]

Taylor, Chris. (1999). *An Introduction to metadata* [Online]. Available: <http://www.library.uq.edu.au/iad/ctmeta4.html> [1999, September 7]

University of New South Wales Library. (1998). *Dublin Core Metadata Template* [Online]. Available: <http://www.library.unsw.edu.au/~eirg/metatemp.html> [1999, September 7]