



The University Of Melbourne

Metadata Paper for the Information Futures Commission

AUTHOR: AILSA DOTT

A PAPER FROM INFORMATION SERVICES

TABLE OF CONTENTS

| | |
|---|---|
| Introduction | 2 |
| Current management of digital objects at University of Melbourne... | 2 |
| What should metadata do for us? | 3 |
| Provide access | 3 |
| Provide interoperability..... | 3 |
| Manage legal obligations..... | 3 |
| Manage collections..... | 4 |
| Ensure migration/interoperability of data is possible | 4 |
| Information Services staff | 5 |
| Some Resources | 6 |
| Document Control | 7 |





Introduction

Metadata is often described as data about data; its purpose is to help the search and discovery process. Assigning relevant keywords, descriptors and subject terms to the data objects enhances their 'findability'. But metadata also plays a key role in preservation, rights management and storage. Preservation, rights and technical metadata are used to manage digital objects to control access, record what kind of files are held, amount of storage required, helper applications required, and so on.

Metadata standards are important. We may want to open our collections to other harvesting institutions (e.g. Picture Australia, the Australian Digital Theses Project), and we in turn might want to harvest from other collections, so it is important we use recognised standards, e.g. OAI, Dublin Core, etc. Exposure via Google is enhanced via adherence to standards.

The world of metadata is still fluid and the possibilities are open. A number of metadata schema exist, some generic, such as Dublin Core, some designed for specific applications such as EAD (Encoded Archival Description) for archival applications, VRA (Visual Resources Association) for images.

Attempts are being made to tie these strands together using the WC3 protocol to allow greater interoperability.

The Council of Australian University Librarians conducted a survey of members in 2001 to determine the use of metadata (<http://www.caul.edu.au/surveys/Metadata.doc>). The results showed a belief that metadata should be applied but there was no consensus on how or what. Looking at University digital repositories today there is still a deal of customisation going on as each institution struggles to provide a language relevant to its users (mostly within the broader frameworks of Dublin Core).

Current management of digital objects at the University of Melbourne

DigiTool is the software currently used to manage the digital resources held by the University of Melbourne Library (the Digital Repository). The default scheme currently in use is Dublin Core. When specialist collections are moved into the repository additional schema may be needed to manage these. We have also developed some University of Melbourne specific descriptors to enhance the fields available in Dublin Core.

The Digital Repository currently contains and provides access to items in a variety of formats including text, images, audio and video.

The Repository will grow with the addition of new collections such as the University of Melbourne Archives Image Catalogue (UMAIC), the University of Melbourne Eprints Repository, newly digitised items from collections such as maps, Readings Online and audio and video materials captured for teaching and research. The Learning Management System has the ability to directly link to resources within the Repository.

In the Repository we apply descriptive metadata, this is effectively cataloguing the item. Appropriately applied it will provide useful search access points, for example, keyword, author and title. Descriptive metadata can provide a description of relationships between objects. Administrative metadata manages



the access rights and relationships between collections. This is applied automatically by the DigiTool software once the rights managements for a particular collection have been established.

What should metadata do for us?

Provide access

Good descriptions of objects will be essential to finding them – terminology used must resonate with users of the objects. Generic controlled vocabularies, such as Library of Congress Subject Headings can no longer keep up with developments in knowledge, especially in fields such as science and medicine which change rapidly. User driven terminology (folksonomies) may be more meaningful. Word or tag clouds are an example of user driven input. There are potential problems with this approach as, for example, a lack of control in use of synonyms can lead to less than optimal indexing and therefore searching. The Themis self-deposit process for research publications will enable creators of the publication to describe the content, so there will be greater interactive description and user input.

Provide interoperability

Harvesting between collection/repositories of digital objects shares and makes more widely available the information contained within. There are in Australia examples such as Picture Australia or the Australian Digital Thesis Project.

Provide access via federated searching/interconnections across the University – large amounts of information exist across the University outside the bounds for the Repository. Some may eventually be incorporated, some never will. There are web pages for faculties or individuals, collections outside the Repository, faculty data sets, and cultural collections. [WHAT POINT IS THIS LAST SENTENCE MAKING?]

Beyond this the Semantic Web beckons:

“The Semantic Web is a web of data. There is lots of data we all use every day, and its not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar? Why not? Because we don't have a web of data. Because data is controlled by applications, and each application keeps it to itself. The Semantic Web is about two things. It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.”
(WC3 Semantic Web, <<http://www.w3.org/2001/sw>>.)

Manage legal obligations

Despite the advent of more freely available information repositories of digital objects there is still an obligation to protect the intellectual property of the creators of the object. As noted in the *Scholarly Information in a Digital Age* Consultation Paper:



“The digital world has changed what we can do with information. Copyright and other intellectual property rights are about what we may do with information. Copyright is a form of property that exists only because the law says so. It exists to maximise good by balancing the incentive to create (owners’ or creators’ rights) with the public good of information being disseminated and used (users’ rights).”

O'Brien, Ms Linda and Brodsky, Mr Mark and Ruwoldt, Ms Margaret L and Newton, Ms Sally (2008). *Scholarly Information in a Digital Age: Choices for the University of Melbourne*. Technical Report, Information Services, The University of Melbourne, p. 13.

Manage collections

In the same way as print collections must be managed, digital collections must be managed. There are issues of restricted access, server space, life cycles and relationships between objects. Application of appropriate metadata makes this all possible.

Ensure migration/interoperability of data is possible

The digital world is fluid. Data is going to move from an old to a new repository, from an old to a new metadata schema, between collections of digital objects for example. The use that is made of recognised standards and the less customisation that occurs, however attractive that may be in the short term, will mean this can be accomplished more easily.

Metadata can assist in searching and discovery across these information sources.

Consistency and maximum functionality will not be achieved without mandated use of appropriate standards and the application of protocols such as WC3.



Information Services staff

Provide the structures and frameworks to allow the storage and management of collection of digital objects, provide advice on standards to ensure interoperability and migration of data where necessary, assist in the provision of policy frameworks, no longer necessarily provide description of content (the library catalogue model), no longer necessarily arbitrate on quality of information.

The creation of metadata is expensive (as is the creation of traditional cataloguing); consider outsourcing of metadata creation (as is done with creation of cataloguing), automatic generation of metadata (already in place with much administrative metadata), and user generated metadata.

Ask what comes after metadata?



SOME RESOURCES

<http://www.metadata.net>

<http://libraries.mit.edu/guides/subjects/metadata/standards.html>

<http://www.nla.gov.au/nla/staffpaper/2003/boston1.html>

http://www.naa.gov.au/images/agls_manual_v2_tcm2-914.pdf

<http://www.infodiv.unimelb.edu.au/cosi/dam/dtlmetadata.html#shared>



Document Control

| | |
|------------------|--|
| Title | Metatdata Paper for Information Futures Commission |
| Short Title | Metatdata paper for IFC |
| Project Ref | |
| Author | Ailsa Dott |
| Status | Working |
| Version | V2 |
| Version Date | 07/05/08 |
| Origination Date | 28/04/08 |
| Change History | * BW placed in template for JE |