The Metadata Coalface for Digital Repositories

Adrian Burton
Australian Partnership for Sustainable Repositories
adrian.burton@apsr.edu.au
www.apsr.edu.au

Chris Blackall
Australian Partnership for Sustainable Repositories
chris.blackall@apsr.edu.au

Scott Yeadon
Australian Partnership for Sustainable Repositories
scott.yeadon@apsr.edu.au

Abstract

In this paper we examine a range of metadata-related issues facing the developers and maintainers of digital repositories in Australia. We discuss metadata developments in the areas of digital preservation, repository interoperability, and collection-level discovery services in the context of a range of innovative repository projects designed to improve metadata creation, management and sharing within the Australian higher education and research sector.

Introduction

It is not unusual to overhear negative comments about metadata, particularly at library technology conferences such as the one this paper is written for. Admittedly, listening to conference presentations about metadata can be about as exciting as watching paint dry. Nevertheless, for those of us working at the digital repository coalface, metadata is a critically important topic. Indeed, to mangle the metaphor used for the title of this paper, we argue that metadata is to digital libraries what coal is to electrical power generation, that is, the lights won't shine without it.

Perhaps the coalface is not an apt metaphor to use with the threat of global warming looming. Nevertheless, we adopt the everyday expression of the coalface to help us focus on how metadata-rich applications, such as digital repositories, are transforming the working lives of higher education and research communities in Australia.

This paper focuses on three areas in which metadata-rich applications are making an impact: digital preservation, repository interoperability, and collection-level discovery services. More specifically, we discuss developments in the areas of digital preservation, repository interoperability, and collection-level discovery services in the context of a range of innovative repository projects through the Australian Partnership for Sustainable Repositories (APSR) in 2006-7 to improve metadata creation, management and sharing within the Australian higher education and research sector.

These areas generally relate to extending the metadata capabilities of digital repositories, but also overlap with trends in the so-called Web 2.0 space. Here, metadata-rich applications and services are becoming an integral part of the Web 2.0 applications that are infiltrating our everyday lives. Ironically, it is Google, Amazon, Yahoo, Facebook and the like that are setting benchmarks for the future of scholarly communications.

How will repositories respond to this challenge? We envisage a world where the barriers to sharing and exchanging digital scholarship and information are radically lowered. The world we are undoubtedly moving toward is one of Web-based 'mash-ups'; that is, networked software applications that can combine data in real-time from multiple service providers in ways that are user-friendly, yet powerful. To be effective in this space, it is imperative that repositories become first-class service providers. This involves collecting, curating and preserving good metadata. Hence, metadata is not simply about technical requirements specific to digital repositories; rather, it forms the basis of an emerging information infrastructure for scholarly communications that has far-reaching consequences.

Before examining these developments in detail, we briefly describe the purpose and function of digital repositories for readers unfamiliar with the topic. We then explain what we mean by metadata, and why it is important.

Digital Repositories

Digital repositories are networked software applications primarily used for storing, managing and disseminating digital resources (e.g. digital publications, theses, data sets and so on) (Crow 2002). Digital repositories differ from conventional content management systems because they include technologies to ensure that digital resources are preserved for long-term access and use. Although digital repositories were initially developed for scholarly communications, they are currently being implemented more widely; for example, by museums to facilitate online access to cultural heritage resources, and government agencies to mediate long-term access to documents and other data.

In practical terms, implementing a digital repository nowadays can be as simple as downloading free open-source software and installing it onto a networked computer. Establishing a stable repository for everyday institutional use is an altogether harder proposition however (Barton and Waters 2004). The most popular open source repository applications are DSpace, Fedora and E-prints (OSI 2004). There are some commercial repository software providers, but none have gained the same level of popularity as the open source repositories mentioned. The important point to note here is that a digital repository is essentially a relational database that stores and keeps track of metadata records for files stored in a mass-data storage facility. The underlying technology is relatively straightforward whereas the institutional context of use is typically complex.

Metadata

Within the digital repository community, metadata literally refers to 'data about data'. More typically, it refers to meaningful information *about* digital 'objects' (or files) stored in a digital repository. Repositories gather and maintain three types of metadata; namely: descriptive, administrative and structural metadata (Lee, Clifton, and Langley 2006).

Descriptive metadata generally takes the form of metadata 'records' stored in a repository database. These database records comprise human-readable descriptive information about digital resources or files, such as an author's name, a subject title, abstract, keywords, date of creation and so on. Descriptive metadata is comparable to bibliographic records in a conventional library database; indeed, descriptive metadata records and bibliographic records are routinely exchanged between library management systems and repositories. The point of capturing and maintaining good descriptive metadata is that is makes digital objects easier for end-users to discover using repository search engines.

Administrative metadata is the technical information about the physical data file, or 'bitstream'. This includes information about file formats (e.g. JPEG of TIFF for digital images), level of data compression and so on. The point of capturing and storing administrative metadata is that is essential for the long-term preservation of digital files, particularly when they are migrated from one format to another, or one repository to another, throughout their lifecycle.

Structural metadata captures the logical and physical relationships of files stored in repositories. Structural metadata captures the relationships and hyperlinks between complex digital files; for example, a set of linked HTML pages with embedded digital images. The point of capturing structural metadata is to ensure that repository software can store and then reconstruct the original design, as determined by the producer. Capturing and storing structural metadata is one of the greatest technical challenges facing repository developers and maintainers. Consequently, it is the area requiring the greatest technical development and community input.

Taken together, descriptive, administrative and structural metadata are at the heart of all digital repositories; therefore, efficiently managing and preserving this type of metadata is an essential task for digital repositories, and the professionals who develop and maintain them.

Metadata Challenges

While acknowledging the importance of metadata to scholarly communications in the era of digital repositories, we also appreciate that creating and maintaining 'good' descriptive, administrative and structural metadata is a major challenge for many scholars and their host institutions (Henty 2007). By good metadata, we mean metadata records that are accurate and comprehensive (from a human-readable perspective) and that conform to the relevant technical metadata standards.

The challenges here are simultaneously institutional, economic and technical.

Firstly, the institutional and economic challenges relate to the up-front and hidden costs of creating and maintaining good metadata, particularly if this work has to be done by qualified library cataloguers. This point should not come as a surprise to the library community, which understands the true costs of creating bibliographic records for their automated library management systems. Given the amount of digital resources being generated by the scholarly community it would make good economic sense to reward and educate them to provide good metadata in the first instance, but this strategy has its own unique set of challenges. The academic community, generally speaking, do not see tangible benefits in creating good metadata for their digital resources, nor are they motivated to seek the technology and skills that would help them. Indeed, they are more likely to see metadata creation as a waste of time rather than an opportunity to improve the impact and reach of their research. These attitudes are slowly changing as traditional scholarly communications rapidly migrates to the World Wide Web; nevertheless, there is widespread agreement that the institutional rewards and training for digital scholarship are currently not of a sufficient level to hasten change (Kling and Spector 2003).

Secondly, there are significant technical challenges to creating and maintaining good metadata. These mostly relate to the widespread use of proprietary file formats and the downstream obsolescence and incompatibilities they introduce for repository managers. Content creators are not necessarily at fault; rather, they are locked into using the proprietary software applications and file formats that they typically use for their day-to-day work. The most commonly used software, such as Microsoft Office, is not geared to help users create good descriptive metadata, nor is it easily integrated with digital scholarly publishing platforms or repositories (Barnes 2007). In contrast, Adobe provides extensive support for descriptive and technical metadata in its software through its Extensible Metadata Platform (XMP) and enables users to create preservation-friendly file formats (see Adobe Systems 2005). But, to reiterate, as long as most scholarly authors and content creators remain untrained in, and unrewarded for, preserving their digital resources, then progress in this area will lag.

Even though the widespread use of proprietary file formats poses problems, repositories are capable of storing proprietary file formats. The problems arise when it comes to accessing and reusing these files at a later date. To do this successfully, users need the original proprietary software applications and operating systems (or special Web browser plug-ins); however, these are often unavailable or outdated. Similarly, repository managers face technical difficulties when converting proprietary file formats into preservation-friendly formats. A key to overcoming these technical barriers is for content creators to switch to using so-called open, non-proprietary, file formats where feasible. For example, instead of using Microsoft Word file formats, authors should adopt the Open Document Format (ODF) that has been established as standard format for both open source and proprietary word-processing software applications (see ISO 2006). Microsoft has publicly announced its support ODF in future versions of Microsoft Office software so this alternative should be available to scholarly authors (Boulton 2006).

The points discussed above by no means exhaust the issues; nevertheless, there is some room for optimism thanks to the steady adoption of open format standards and the growing capabilities of the open-source software development community to replace proprietary software and file formats. Overall, our capacity to innovate and overcome technical barriers to progress in digital preservation is constantly improving. In the following sections, we discuss how the repository development community is innovating in the metadata application domain.

Metadata for Preservation

A key function of digital repositories is to ensure that scholarly information and data is preserved for long-term access and use. However, when digital repositories were first designed, digital preservation was solely considered in terms of the needs and responsibilities of individual users and institutions. This still holds true but has been extended recently because digital preservation is increasingly being seen in government policy circles as a national priority covering the higher education and research sectors, if not the whole-of-government.

In Australia, for example, the recently published ‘Data for Science’ Working Group report to the Prime Minister’s Science, Engineering and Innovation Council (PMSEIC 2006) has highlighted the critical importance of research data collections on a national scale. This shift to a national perspective has emerged because research ‘collections’ are now seen as common resources that need to be shared between all researchers, research groups and universities to maximise their utility (Blackall 2007). By providing secure platforms for disseminating research information, digital repositories have become key technologies in our national information infrastructure (Burton 2007).

Paralleling the national concern over access to research collections has been the focus on sustainably curating and preserving such collections. Specialist national organizations have been established to deal these issues, such as the Digital Curation Centre (DCC) in the United Kingdom. We anticipate that the proposed Australian National Data Service will play a similar function nationally for the higher education and research sector (see ANDS Technical Working Group 2007).

At a more granular level, international leadership in digital preservation has come through the activities of PREMIS (PREservation Metadata: Implementation Strategies) Working Group. Although based in the USA, the PREMIS Working Group had wide international input, including representatives from the National Library of Australia. In May 2005, the Working Group published their final report, which comprised a “data dictionary for core preservation metadata needed to support the long-term preservation of digital materials" (PREMIS Working Group 2005). This document identifies the core information elements required for preserving files and gives repository developers a common language, and common data dictionary, for implementing preservation metadata.

The PREMIS data dictionary provides a firm conceptual foundation for moving forwards, yet it doesn't give (and was never required to give) a specific list of metadata elements that are needed/desirable for use in digital repositories. Nor does it provide clear technical specifications, or a precise XML metadata schema, that software engineers could use to improve repositories to the point where they could be considered 'PREMIS compliant'. Such a metadata schema, however, is exactly what repository developers need.

We, the Australian Partnership for Sustainable Repositories through its partnership with the National Library of Australia, responded to this challenge when we funded the Preservation Requirements Statement (PRESTA) project (see APSR 2007). The stakeholders in the PRESTA project closely examined the PREMIS data dictionary to develop a requirements specification for preservation metadata that would inform subsequent repository development efforts (Lee, Clifton, and Langley 2006). Following the PRESTA project, APSR (again in partnership with the National Library of Australia) developed a draft metadata 'application profile' that implemented a subset of the PREMIS data dictionary (see next section for more details) (Lee 2006).

A metadata application profile is defined as "an assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema" (DCMI 2007). Its practical function is to describe the syntax and semantics of metadata elements that are required for software applications to process data for predictable and repeatable results. From the outset, we decided to create an application profile for PREMIS/PRESTA using the Metadata Encoding and Transmission Standard maintained by the Library of Congress (see METS). We decided to use METS because it is specifically designed to encode descriptive, administrative, and structural metadata of digital resources to be stored within digital repositories. Moreover, it is widely adopted within the repository community and the Library of Congress maintains a METS Implementation Registry for the maintenance of community developed application profiles.

To demonstrate the effectiveness of the PREMIS/PRESTA metadata application profile, a repository demonstrator application was jointly developed by the Australian National University and the University of Queensland in late 2006. This work included stakeholder involvement from the (ARROW) and (RUBRIC) repository projects, also funded through the Australian Government's Systemic Infrastructure Initiative.

We initially used the PREMIS/PRESTA application profile to examine the feasibility of transferring digital objects between a DSpace repository (at ANU, Canberra) to a Fedora repository (at the University of Queensland, Brisbane), and vice versa. The digital preservation issue examined was the feasibility of using METS to create a common data interchange format to automatically replicate and/or mirror repository collections. Indeed, we found that that the METS was ideal for this task. Using this approach, data (in the form of digital image collections and metadata) were automatically transferred between DSpace and Fedora repositories without problems—a world-first to our knowledge.

This experiment demonstrated that by applying the right metadata standards and common repository interoperability frameworks, repositories, data centres and web-based applications could automatically and reliably share information without technical barriers getting in the way.

This successful experiment inspired us to undertake further work on the PREMIS/PRESTA application profile through 2007, which is described in detail in the following section.

Metadata for Repository Interoperability

Having established that METS could be used for digital preservation, it also became clear to us that the PREMIS/PRESTA profile could be extended to greatly improve the interoperability of digital repositories. By interoperability, we mean the capacity of heterogeneous information systems to exchange information in highly automated ways, without losing or corrupting data in the process.

The point of improving the interoperability of repositories is to not only enable the scholarly community to share data, but also enable repositories to be better integrated with a wide range software applications and systems typically used for scholarly communications. We envisaged that it if we could lower the barriers faced by the scholarly community when they have to move newly created digital resources into repositories and other data storage facilities then this would encourage their wider use. These barriers exist because scholars typically have to upload new digital resources individually, or in small batches, to repositories. This process is time-consuming and error prone and is widely acknowledged as a major issue needing to be solved (McNamara and Buchhorn 2006).

Our approach to the solution was to see if we could provide better interoperability at the repository side of the upload process. The aims here were to establish automated services upload services between the repository and desktop applications, such as Microsoft Office, which would make uploading new resources a simple one-click operation. To do this, however, we realised that we needed a common data transfer file format that could contain the uploaded files, as well as descriptive, administrative and structural metadata. Within the repository community, these data transfer formats are referred to as Submission Information Packages (SIP). Based on our previous success, we decided to use METS as basis for the SIP.

In the process of working with the METS standard and consulting the METS Primer (METS Editorial Board 2007), we learnt that using METS for the SIP was not straightforward. The major challenge we encountered was that developing a usable SIP requires agreement within the repository community on how the METS standard is interpreted as an *application profile*. To reiterate, an application profile typically contains metadata elements drawn from other standards that are wrapped into the overall data encoding syntax, which in our case was provided by METS. How these metadata elements are selected and assembled requires careful thought and enormous attention to detail. For example, we decided to use the Metadata Object Description Schema (MODS) for the bibliographic sections of the SIP; however, there are no established guidelines for doing so within METS. This should not have come as a surprise to us because the METS Editorial Board makes it clear that developing community-wide METS application profiles requires community-wide agreements and policies.

Complicating matters further was the fact that DSpace and Fedora repository software development groups have implemented their own interpretations of the METS standard for processing SIPs. Indeed, we discovered that when we wanted to move data between DSpace and Fedora we had to create a different SIP for each repository, as well as some extra software code so that the SIPs could be processed consistently. Ideally, we wanted a single SIP format that could be processed by all repositories, or other software applications, without the need of modified software code, or on-off formatting tweaks.

The take home message for us here was that developing effective metadata application profiles requires national and international agreements, maintenance agencies and consistent implementation within repository software.

The Repository Interoperability Framework

With the experience of creating application profiles fresh in our mind, we began to develop of an application profile for a common repository SIP for the Australian higher education and research sector in early 2007. The focus was on improving the interoperability between repositories, particularly the DSpace and Fedora repositories used by APSR partner institutions. This was accomplished through a series of coordinated projects under the title of the Repository Interoperability Framework (RIFF).

The goals of the RIFF projects was to lower technical barriers involved in moving scholarly resources into repositories by making the submission process as streamlined and automated as possible for the academic community (Yeadon 2007). The RIFF project involved five institutional partners: the Australia National University, University of Queensland, University of Sydney, National Library of Australia and the Australian Partnership for Advanced Computing. Again, stakeholder involvement was gained from the ARROW and RUBRIC projects.

The RIFF projects began by defining a set of ‘genre workflows’ that were to be implemented independently of the internal submission process used by DSpace and Fedora. By genre workflows, we mean the combination of scholarly communications genres (scholarly research papers, conference papers, theses, image albums etc.), and automated software workflows consisting of software applications (word processors, databases etc.) that are connected to a series of software services, in particular, using Web Services standards and protocols. These services enable authors to literally pipe digital content from their desktop computers through to a repository or other data storage facility in a one-click operation from the desktop. This approach builds on the so-called service-oriented architectures now commonly employed in the business computing domain and, more recently, in the e-learning Web 2.0 domain (Dagger et al. 2007; Alexander 2006; Wilson, Blinco, and Rehak 2004).

This service-oriented approach aspect of the RIFF projects was accomplished through the interoperability layers of DSpace and Fedora that employ Web Services standards and protocols to communicate with third-party Web-based applications and services in the following stages. Firstly, the source digital files and metadata records are packaged together as a SIP using the agreed metadata application profile. The packaging of the SIP was done programmatically at the content provider's end of the process, after which it was transferred from the data provider to a repository. We developed an automated upload service for this task, which we named the RIFF Submission Service. Once stored in the repository, the SIP is then unpacked. By this we mean that the metadata records are stored in the repository database and the source files in the appropriate locations in the underlying data storage system. Once unpacked, repository users can easily locate and download the files by using the repository search engine, or browse functions.

To test this approach we took the digital content from known online sources and passed it through the RIFF Submission Service. For example, we took the PDF formatted articles and metadata from Open Journal System (OJS), a popular online e-journal editorial and production system, and automatically transferred them to DSpace and Fedora instances in a one-click operation (Christensen and Coleman 2007). The benefits of this approach were to give e-journal owners a solution for preserving their e-journals, as well as new options to make it more accessible to the public, at the same time providing a long-term preservation solution. Similar tests were successfully conducted with image collections, fieldwork data collections (Honeyman 2007) and word-processed scholarly documents (Barnes 2007).

This application profile—now appropriately named the Australian METS Profile—consists of a core set of metadata elements that mandate the use of PREMIS elements for administrative metadata, MODS elements for descriptive metadata and several metadata extensions for specific formats and content genres. A detailed technical report of the initiative is available online (see APSR 2007), and the Profile itself is registered with the Library of Congress, as the international maintenance agency for METS.

Overall, the RIFF projects have demonstrated the utility of using the Australian METS Profile to improve repository interoperability. With the help of more development work and publicity in the near future, we hope that the Australian METS Profile will become an internationally supported standard. Nevertheless, major work will be required to ensure that the leading repository software applications improve their support for registered METS profiles.

Metadata for Collection-Level Discovery

Our final example of developments in the metadata domain is the development of repository discovery services based on collection-level metadata.

This development has arisen largely because the Google search engine has transformed the public's expectations about what can be discovered on the Internet (Estabrook and Rainie 2007). Google, and a host of other innovative Web 2.0 applications, have demonstrated that search results can be enriched with a wide range of contextual information and hyperlinks. Inspired by these developments, libraries are pioneering new search services that seek to emulate the context-rich search results of Google. The WorldCat (see OCLC) and Libraries Australia (see NLA) search engines exemplify this approach.

Significantly, the library community has acknowledged the need to fundamentally rethink library cataloguing in light of the changing public expectations about what search engines can provide (Rosa, Dempsey, and Wilson 2004). The recent report from of the Library of Congress Working Group on the Future of Bibliographic Control (2008) highlights the scope of changes required. They recommend a shift from the current model of item level bibliographic control towards a relational model based on the Functional Requirements for Bibliographic Records (1998). However, before this recommendation can be realised, bibliographic records (and descriptive metadata generally) will need to be enriched with authoritative information about a wide range of information entities so that contextual relationships can be represented in search results. Hence, there is a growing realisation within the library community that registries of authoritative information will be required for many of the high-level entities in the FRBR information model; such as for people, places, organizations, collections and so on (Gorman 2003; Patton 2005).

Where do repository collections fit into this new information landscape? Currently, repository end-users are provided with search results based on descriptive metadata at the item level. This situation has arisen because most repositories support the Dublin Core Metadata Element Set (DCMI 2008), which comprises item level metadata elements only. By items we mean discrete digital files; for example, individual documents, images or data sets. This typically gives useful search results, but just as long the user has a fair degree of certainty about what they are seeking. This approach is less useful when the user needs more information to refine their search, or when seeking contextual information that might broaden its scope.

The repository community is aware of the limitations of the current Metadata Element Set; indeed, it has had some success in aggregating item level metadata for searching through the implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH 2004). Although widely used, OAI PMH powered search engines have also exposed the poor data quality of much of the metadata being harvested (due to missing, incomplete and/or duplicated information) and the lack of basic collection level information that could be used to help enrich search results.

This situation is about to change. The Dublin Core Metadata Initiative recently introduced the Dublin Core Collection Application Profile (DCCAP 2008), which covers collection-level metadata. The Profile document provides a broad definition of a collection when it states, "…the term 'collection' can be applied to any aggregation of physical or digital items." We anticipate a number of benefits that will flow from the support of the Collection Application Profile within the repository community. For example, in the higher education and research sector where most scholarly work is done in groups, or by broad discipline area, collections metadata will greatly improve our capacity to discover resources that are aggregated and sorted according to group criteria. Not only will the addition of collection-level metadata help refine user searches and improve browse and navigation structures throughout repositories, but it will also provide richer contextual linkages to information beyond the boundary of the repository community, such as Google and Wikipedia.

Furthermore, the adoption of collection-level metadata would allow repository managers to initiate automated administrative and curatorial processes; for example, to control access to collections or migrate/replicate them to other repositories. Indeed, we see great many opportunities in the Web 2.0 space to expose collection-level information to third-party Web applications to create new services for our end-users (Burton 2007).

The Online Research Collections Australia Registry

In this section we describe another APSR project named the Online Research Collections Australia (ORCA) Registry that demonstrates the possibilities for improved collection level discovery described above.

The aims of ORCA-Registry project were to provide a national registry for collection-level metadata 'harvested' from repositories and other data sources. The ORCA-Registry design anticipated the need for a federated network of repositories and data storage facilities in Australia as set out in the National Collaborative Research Information Strategy (NCRIS 2008). The Registry will not only provide a 'human-readable' interface to the harvested information, but will also form the basis for a wide range of machine-to-machine 'services' that intersect with Web 2.0 developments. Thus, we refer to the ORCA-Registry as a collections and services registry. The Registry will be operational in 2008 as part of the Australian National Data Service (ANDS) as part of the NCRIS investment.

We faced numerous design challenges when specifying the ORCA-Registry. Firstly, it had to maintain detailed information about institutional owners/managers of collections, and the various rights, authorisation and access conditions covering them. Secondly, it had to provide Web Services interfaces that would enable communications with third-party applications, such as repositories. Finally, it needed to be user-friendly for registry administrators and the providers of collection-level metadata records.

To meet these design challenges we originally envisaged using the Dublin Core Collection Application Profile (DCCAP); however, we found that it did not include sufficient scope to include the type of descriptions for high-level business entities that we required and so we searched for a better solution. We plan to retain DCCAP as a metadata exchange protocol instead. In the end we adopted the ISO2146 (Registry Services for Libraries and Related Organisations) standard (see NLA 2008). We decided to adopt the ISO2146 standard because of its comprehensiveness and fitness as a generic registry model (Pearce and Gatenby 2005).

After closely analysing the ISO2146 standards document we developed a functional data model of the ORCA-Registry, which was then mapped to a relational database structure for testing. This was followed by the development of an XML schema to assist the automated submission of third-party metadata into the Registry. These technologies were successfully demonstrated as a functional Web-based ORCA-Registry application in September 2007. The collection-level metadata for the Registry was provided by the ORCA-Network, an institutional network of Australian higher education and research data providers that are also the target users of the ORCA-Registry (see APSR 2007).

Another key design goal of the ORCA-Registry was to support inclusion in, and expansion of, a global network of interoperable collections and services registries (Burton 2007). To this end, representatives of collecting institutions met in Washington in December 2007 to discuss the opportunities for such a network. It was agreed that discussion and planning toward implementation would take place through 2008. We think that the ORCA-Registry project will provide an inspiration for the development of global collection-level discovery services. In other words, we hope that it will be to research collections what Google is to generic web searching.

Naturally, as the first functional registry developed using the ISO2146 standard, we encountered some issues during implementation. These are being raised with the relevant ISO standards committee and are being addressed. The ORCA Registry software was released to the public in late 2007, and this will undoubtedly aid other registry developers and further refine the ISO standard (see APSR 2008).

Conclusion

In this paper we have discussed developments in digital preservation, repository interoperability, and collection-level discovery, and examined a range of innovative repository projects guided by the Australian Partnership for Sustainable Repositories that are designed to improve metadata creation, management and sharing within the Australian higher education and research sector.

The developments described in this paper suggest that the prevailing assumptions that digital repositories as discrete, monolithic, entities must change. Indeed, the public success of Google's web-based applications almost assures this change. This success indicates that repositories must become more agile, service-oriented and interoperable if they are to be relevant to end-users in the higher education and research sector. The world we are undoubtedly moving toward is one of Web-based 'mash-ups', that is, networked applications that can combine data in real-time from multiple service providers. Clearly, digital repositories must become active providers and consumers of these services. A good place to start this effort would be to investigate, if not implement, some of the metadata concepts and applications outlined in this paper.

References

Adobe Systems. 2005. Adobe XMP for Creative Professionals. San Jose, CA: Adobe Systems Incorporated [cited 10 Jan. 2008]. Available from http://www.adobe.com/products/xmp/pdfs/xmp_creativepros.pdf.

Alexander, Bryan. 2006. Web 2.0: A New Wave of Innovation for Teaching and Learning? *Educause Review* 41 (2):32–44.

ANDS. 2008. *Australian National Data Service*. Australian Government [cited 10 Jan. 2008]. Available from http://www.ncris.dest.gov.au/capabilities/collaborative_investment_plan_platforms.html.

ANDS Technical Working Group. 2007. Towards the Australian Data Commons: A proposal for an Australian National Data Service Canberra: Department Education, Science and Training [cited 10 Jan. 2008]. Available from http://www.pfc.org.au/twiki/pub/Main/Data/TowardstheAustralianDataCommons.pdf.

APSR. 2007. PREMIS Requirement Statement (PRESTA). Canberra: Australian Partnership for Sustainable Repositories and the National Library of Australia [cited 10 Jan. 2008]. Available from http://apsr.anu.edu.au/presta/index.htm.

———. 2007. Report of the METS Profile Development Project. Canberra: Australian Partnership for Sustainable Repositories [cited 10 Jan.]. Available from http://www.apsr.edu.au/nla-mets/mets_profile_report.pdf.

———. 2008. *ORCA-Registry Software*. Australian Partnership for Sustainable Repositories 2008 [cited 10 Jan 2008]. Available from http://www.apsr.edu.au/software.htm.

ARROW. 2008. *Australian Research Repositories Online to the World* [cited 10 Jan. 2008]. Available from http://arrow.edu.au/.

Barnes, Ian. 2007. The Digital Scholar's Workbench. Canberra: Australian Partnership for Sustainable Repositories [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/presentations/barnes_elpub07_paper.pdf.

Barton, Mary R., and Margaret M. Waters. 2004. Creating an Institutional Repository: LEADIRS Workbook. Cambridge, MA: LEarning About Digital Institutional Repositories (LEADIRS), MIT Libraries [cited 10 Jan. 2008]. Available from http://dspace.org/implement/leadirs.pdf.

Blackall, Chris. 2007. Digital Repositories and the Australian Higher Education Sector: Where to Next? Melbourne: EduCause Australasia 2007 Conference [cited 10 Jan. 2008]. Available from http://www.caudit.edu.au/educauseaustralasia07/authors_papers/Blackall-125.pdf.

Boulton, Clint. 2006. Microsoft Backs Open Document Format: InternetNews.com [cited 10 Jan. 2008]. Available from http://www.internetnews.com/dev-news/article.php/3618176.

Burton, Adrian. 2007. E-research National Perspectives. Melbourne: EduCause Australasia 2007 Conference [cited 10 Jan. 2008]. Available from http://www.caudit.edu.au/educauseaustralasia07/authors_papers/Burton-230.pdf.

———. 2007. In Support of Digital Collections: The APSR Development Agenda. Washington, DC: CNI Spring Taskforce Meeting [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/presentations/burton-cni-spring-07.pdf.

———. 2007. Services and Infrastructure for Digital Collections. University of North Carolina at Chapel Hill: DigCCurr2007: an International Symposium in Digital Curation [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/presentations/burton_digccur_06.pdf.

Christensen, Sten, and Ross Coleman. 2007. Open access publishing and the repository - a strategy for sustainability. Vancouver, British Columbia: PKP Scholarly Publishing Conference [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/presentations/christensen_pkp2007.pdf.

Crow, Raym. 2002. The case for institutional repositories: a SPARC position paper. Washington, DC: Scholarly Publishing and Academic Resources Coalition (SPARC).

Dagger, Declan, Alexander O'Connor, Séamus Lawless, Eddie Walsh, and Vincent P.Wade. 2007. Service-Oriented E-Learning Platforms: From Monolithic Systems to Flexible Services. *IEEE Internet Computing*.

DCC. 2008. *Digital Curation Centre* [cited 10 Jan. 2008]. Available from http://www.dcc.ac.uk/.

DCCAP. 2008. Dublin Core Collections Application Profile: Dublin Core Metadata Initiative [cited 10 Jan. 2008]. Available from http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/2004-08-20/.

DCMI. 2008. *DCMI Tools Glossary*. Dublin Core Metadata Inititiative 2007 [cited 10 Jan. 2008]. Available from http://dublincore.org/groups/tools/glossary.shtml.

———. 2008. Dublin Core Metadata Element Set (Version 1.1): Dublin Core Metadata Initiative [cited. Available from http://dublincore.org/documents/dces/.

Estabrook, Leigh, and Lee Rainie. 2007. Information searches that solve problems: How people use the internet, libraries, and government agencies when they need help Washington, DC: Pew Internet and American Life Project [cited 10 Jan. 2008]. Available from http://www.pewinternet.org/pdfs/Pew_UI_LibrariesReport.pdf.

Gorman, Michael. 2003. Authority control in the context of bibliographic control in the electronic environment. Paper read at Authority Control: Reflections and Experiences, at Florence, Italy, 10-12 February.

Henty, Margaret. 2007. Ten Major Issues in Providing a Repository Service in Australian Universities. Review of Reviewed Item. *D-Lib Magazine* (5/6), http://www.dlib.org/dlib/may07/henty/05henty.html.

Honeyman, Tom. 2007. FieldHelper: a tool to assist the collation of field data and its ingestion into repositories. Brisbane: eResearch Australasia 2007 Conference [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/presentations/tom_honeyman.pdf.

IFLA Study Group on the Functional Requirements of Bibliographic Records. 1998. *Functional Requirements of Bibliographic Records: final report.* München: K. G. Saur.

ISO. 2006. Open Document Format for Office Applications (OpenDocument) v1.0. : International Organization for Standardization, ISO/IEC 26300:2006 Information technology [cited 10 Jan. 2008].

Kling, Rob, and Lisa B. Spector. 2003. Rewards for scholarly communication. In *Digital scholarship in the tenure, promotion and review process*, edited by D. L. Andersen. Armonk, NY: M.E. Sharpe.

Lee, Bronwyn. 2006. Preservation Metadata: Adapting or Adopting PREMIS for APSR. Ithaca, NY: 3rd International Conference on the Preservation of Digital Objects (iPRES 2006) [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/presentations/ipres_lee.pdf.

Lee, Bronwyn, Gerard Clifton, and Somaya Langley. 2006. The PREMIS Requirement Statement Project Report. Canberra: Australian Partnership for Sustainable Repositories and the National Library of Australia [cited 10 Jan. 2008]. Available from http://apsr.anu.edu.au/publications/presta.pdf.

Library of Congress. 2008. *Australian METS Profile*. Library of Congress 2007 [cited 10 Jan. 2008]. Available from http://www.loc.gov/standards/mets/profiles/00000018.html.

Library of Congress Working Group on the Future of Bibliographic Control. 2008. On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control. Washington, DC: Library of Congress [cited 10 Jan. 2007]. Available from http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

LOC. 2008. *METS Implementation Registry*. Library of Congress [cited 10 Jan. 2008]. Available from http://www.loc.gov/standards/mets/mets-registry.html.

McNamara, Paul, and Markus Buchhorn. 2006. Sustainability Issues for Australian Research Data: The report of the Australian e-Research Sustainability Survey Project. Canberra: Australian Partnership for Sustainable Repositories [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/documents/APSR_Sustainability_Issues_Paper.pdf.

METS. 2007. *Metadata Encoding and Transmission Standard*. Library of Congress [cited 10 Jan. 2007]. Available from http://www.loc.gov/standards/mets/.

METS Editorial Board. 2007. Metadata Encoding and Transmission Standard: Primer and Reference Manual. Canberra: Library of Congress [cited 10 Jan. 2008]. Available from http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf.

MODS. 2008. *Metadata Object Description Schema*. Library of Congress [cited 10 Jan. 2008]. Available from http://www.loc.gov/standards/mods/.

NCRIS. 2008. National Collaborative Research Infrastructure Strategy. Canberra: Australian Government [cited 10 Jan. 2008]. Available from http://www.ncris.dest.gov.au/.

NLA. 2008. *Libraries Australia*. National Library of Australia [cited 10 Jan. 2008]. Available from http://librariesaustralia.nla.gov.au/apps/kss.

———. 2008. *ISO 2146 Project*. National Library of Australia 2008 [cited 10 Jan 2008]. Available from http://www.nla.gov.au/wgroups/ISO2146/.

OAI PMH. 2004. Open Archives Initiative Protocol for Metadata Harvesting (Version 2): Open Archives Initiative [cited 10 Jan. 2008]. Available from http://www.openarchives.org/pmh/.

OCLC. 2008. *WorldCat* [cited 10 Jan. 2008]. Available from http://www.worldcat.org.

OJS. 2008. *Open Journal Systems*. Public Knowledge Project [cited 10 Jan. 2008]. Available from http://pkp.sfu.ca/?q=ojs.

OSI. 2004. A Guide to Institutional Repository Software (3rd ed.): Open Society Institute [cited 10 Jan. 2008]. Available from http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf.

Patton, Glenn E. 2005. FRAR: Extending FRBR Concepts to Authority Data. Dublin, Ohio, USA: OCLC [cited 10 Jan. 2008]. Available from http://www.ifla.org/IV/ifla71/papers/014e-Patton.pdf.

Pearce, Judith, and Janifer Gatenby. 2005. New Frameworks for Resource Discovery and Delivery. Wellington, New Zealand: Standards Australia IT-19 Seminar, Technical Standards for Libraries and Education: Solutions and Emerging Frameworks. National Library of New Zealand [cited 10 Jan. 2008]. Available from http://www.nla.gov.au/nla/staffpaper/2005/pearce1.html.

PMSEIC. 2006. Data for Science. Canberra: Prime Minister's Science, Engineering and Innovation Council (PMSEIC), Australian Government [cited 10 Jan. 2008]. Available from http://www.dest.gov.au/sectors/science_innovation/publications_resources/profiles/documents/Data_for_Science_pdf.

PREMIS Working Group. 2005. Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group. Dublin, OH: PREservation Metadata: Implementation Strategies Working Group, OCLC [cited 10 Jan. 2008]. Available from http://www.oclc.org/research/projects/pmwg/premis-final.pdf.

RIFF. 2008. *Repository Interoperability Framework*. Australian Partnership for Sustainable Repositories [cited 10 Jan. 2008]. Available from http://www.apsr.edu.au/currentprojects/index.htm.

Rosa, Cathy De, Lorcan Dempsey, and Alane Wilson. 2004. *2003 OCLC Environmental Scan: Pattern Recognition*: Online Computer Library Center.

RUBRIC. 2008. *Regional Universities Building Research Infrastructure Collaboratively* [cited 10 Jan. 2008]. Available from http://www.rubric.edu.au/.

Wilson, Scott, Kerry Blinco, and Daniel Rehak. 2004. Service-oriented frameworks: modelling the infrastructure for the next generation of e-learning systems: DEST (Australia), JISC-CETIS (UK), and Industry Canada [cited 10 Jan. 2008]. Available from http://www.jisc.ac.uk/uploaded_documents/AltilabServiceOrientedFrameworks.pdf

Yeadon, Scott. 2007. Content Interchange and the Invisible Repository. San Antonio, Texas: Open Repositories 2007 Conference [cited 10. Jan 2008]. Available from http://www.apsr.edu.au/presentations/yeadon_open_repositories_07.pdf.