



Sustaining Repositories

newsletter of the APSR Project

December 2007

www.apsr.edu.au

ISSN 1834-1098

In this issue:

- **ORCA providing a better 'view'**
- **Open Repositories 2008**
- **OA advocacy on show**
- **ORCA training**
- **Clever Collections**
- **Making an impression**
- **Hidden heroes – Tran Le**
- **Report from the field - Beijing**
- **Pulling it together**
- **Coffee anyone?**
- **Around the Traps**

■ **ORCA providing a better 'view'**

At last there is a program that opens the door to the data collections that exist all over the country. The ORCA Registry project (<http://www.apsr.edu.au/orca/index.htm>) was established as a way of sharing resources and information about collections between institutional repositories, archives and data centres – providing a better 'view' of collections.

Data collections exist at all universities, research organisations (like CSIRO) and in organisations with research interests, such as archives, museums and libraries. In addition, key national organisations hold research data collections, such as the Australian Social Science Data Archive (ASSDA). ORCA has created a discovery 'portal' for information about these collections. This is the 'human readable' output, but ORCA has also created a services registry that can facilitate machine-to-machine services. These will allow applications like alerting services, aggregation services, institutional and personal research portfolios, and other brokering services.



This is not the first collection registry in the world, with the most mature collections registry service being the Information Environment Service Registry funded by the Joint Systems Information Committee (JISC) in the UK. ORCA needed to ensure it was able to work with other collections registries, with a longer term aim of collaborating with the UK registry and with OCKHAM. Funded by the National Science Foundation, OCKHAM is part of the National Science Digital Library initiative.

Information about collections complements the item-level view of collections. It provides information about the collection-level resources themselves, contact details for the resource providers, technical details about how to access the resources, and software interfaces for this to happen automatically if required.

The registry is using a quality standard, ISO 2146, which defines a registry as a "collection of registry objects compiled to support the business of a given community. A registry includes the services on top of the data – not just the data."

Those involved in the setting up of the registry include APSR staff, most notably James Blanden, and also people from the National Library of Australia, ASSDA, the Australian Science Archives Project, ARROW and the ASL Partnership Laboratory.

■ **Open Repositories 2008**

One of the highlights on the northern hemisphere conference season in 2008 will be the third Open Repositories conference between 1-4 April. Hosted by the University of Southampton, which developed the EPrints repository platform, the conference aims to address the technical, managerial, practical and theoretical issues that arise from diverse applications of repositories in the increasingly pervasive information environment.

Open Repositories 2008 (<http://or08.ecs.soton.ac.uk/>) will provide focused workshops and tutorials, followed by general conference sessions and EPrints/DSpace and Fedora user group meetings.

OA advocacy on show

Don't let the silly season pass without booking your place on *Open Access Collections* in Brisbane next year. This will be a one day workshop on the challenges and opportunities of open access publishing for Australian universities.

This is a rare opportunity to hear speakers who work in the area of open access advocacy. This event will give repository managers and others working in the area of repositories or libraries some real-life examples to assist them in increasing the uptake and use of their repository.

The keynote speaker is Dr Peter Murray Rust, who works in molecular informatics at Cambridge University and has been promoting open source and open data. APSR is also pleased to confirm Kevin Stranack, Librarian with the Public Knowledge Project at the Simon Fraser University will attend. A highlight will be a panel discussion at the conclusion of the day. The panel, chaired by the always entertaining Belinda Weaver (University of Queensland), will include Paula Callan (Queensland University of Technology), Vicky Picasso (University of Newcastle), Simon McMillan (University of New England) and Margaret Henty (APSR).

Conveniently scheduled on Thursday, February 14, for the week after VALA2008, the workshop will be held at the historic Customs House, overlooking the Brisbane River.

More information about the event can be found at http://www.apsr.edu.au/open_access_collections/index.htm and a registration form is now available.



Pic credit:
<http://www.customshouse.com.au/>

ORCA training

An eResearch Forum was held at the University of Sydney on November 1 to raise awareness of eResearch issues and to enable networking and information sharing among those involved in eResearch. Approximately 90 people attended the forum organised by Ross Coleman and Rowan Brownlee from the University Library which featured an address by Dr Rhys Francis, Executive Director, Australian eResearch Infrastructure Council. There were also presentations by 23 University of Sydney academics and general staff.



Pic credit:
<http://www.usyd.edu.au>

The forum sought to provide an informal space for researchers to engage with policy makers. The morning session focused on the eResearch context at a national and institutional level and services for enabling and supporting eResearch at the University of Sydney. The afternoon sessions provided a fast-paced sampling of University of Sydney eResearch projects and associated products & tools. Professor Merlin Crossley, Acting Deputy Vice Chancellor of Research, concluded with an address discussing next steps for the University.

This forum was followed by workshops on digitisation, digital audio, GIS tools and high performance computing. The first one-hour workshop on November 7 was held by Tom Honeyman, a linguist from PARADISEC. Titled "Introduction to recording research audio quality in the field", it covered the basics of recording from microphone to recorder to computer. Participants discovered how to avoid simple mistakes, and how to create good quality and archival quality recordings.

A week later Andrew Wilson, GIS Coordinator from the Archaeological Computing Laboratory, held a one day workshop on geographic information systems. Targeted at beginners, the workshop provided an overview of options and approaches to mapping data.

There has been great interest in the forthcoming workshop on December 5, on digitisation for preservation and archiving. This workshop, originally developed for Belinda Weaver, University of Queensland will be presented by Andrew Stawowczyk Long, Manager Preservation Standards, National Library of Australia. The 3.5 hour introductory workshop will focus on images, and cover the essentials of digitisation for preservation.

More information about the workshops can be found at:
<http://www.library.usyd.edu.au/eresearch2007/workshops.htm>

Clever Collections

Clever Collections: a national showcase of technical innovations for digital collections was held on 28-29 November at the University of Melbourne with over 120 people in attendance. The conference explored the exciting and innovative approaches being taken around the country to play with and tend data. Professor Leon Sterling, Director of eResearch at the University of Melbourne opened the conference with a brief welcome, showing his commitment to the furthering of eResearch and his support for the aims of the conference.



Professors Leon Sterling, Noshir Contractor and Philippa Pattison

Four plenary speakers addressed issues of interest to all those present. The first of these was Professor Noshir Contractor, Jane S. & William J. White Professor of Behavioral Sciences at Northwestern University, who took the theme of clever collections leading to clever connections. Building technology is one thing, but we need to remember that in doing so we have to meet real human needs. By way of illustration of this, the main (legal) use of email is to set up face to face meetings. He gave examples of using technology to track social networks, a technique which can be used in all kinds of settings including networks of researchers.

Dr Elizabeth Webster, from the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne, put into context the dilemma of researchers and technologists who deal with intangible objects which do not seem to be as highly valued as tangible objects in an

economic environment. Jock Phillips from *Te Ara: the Encyclopedia of New Zealand* showed us the latest developments in this innovative online history to the delight of everyone. Dr Rhys Francis, Executive Director of the newly-created Australian eResearch Infrastructure Council, was keenly listened to as he outlined future plans for data management and stewardship.

The many clever presenters spoke of their work with considerable enthusiasm and demonstrated their expertise in handling data. Talks came from many different disciplines and types of organisations.

The presentations will be shortly placed on the website at:

<http://www.apsr.edu.au/clevercollections/>

Clever Collections was sponsored by APSR and ARROW who would like to acknowledge the considerable contribution of the University of Melbourne to the organisation of the event.

Making an impression in America

Matthew Walker recently travelled to North America on behalf of both the National Library of Australia (NLA) and the Australian Partnership for Sustainable Repositories (APSR). He reports here:

My journey presented a number of opportunities for promoting the work of APSR. During my various ports of call, I was able to talk about and discuss, in particular, the Automated Obsolescence Notification System (AONS) and METS profile work the NLA participated in over the past several months.

Firstly, I gave a presentation about AONS at the Digital Library Federation (DLF) Fall Forum (<http://www.diglib.org/forums/fall2007/>) in Philadelphia on 7 November. The presentation proved to be very popular, with approximately 70 forum delegates attending the session. It provoked lively and interested discussion not only in the software itself, but also in the risk questions that had been developed as an adjunct to the AONS project.

Following the DLF Forum, I participated as a guest in the METS board meeting and described the Australian METS Profile work. The work on the Australian METS Profile seemed to have been well-received by the group and they were very interested in the approach that the Australian collaborators had taken. At the meeting, there was also detailed discussion about METS capabilities for interoperability and machine action-ability. The METS board wishes to investigate the possibilities of this further, and is setting up a group for discussion of the issues. They would very much like an Australian representative to be part of the working group.

I also met with Stephen Abrams and Gary McGath at the Harvard University Library to discuss AONS and the JHOVE format validation tool. They were very interested in the work we had done, and we also discussed the overlap that may exist between the new JHOVE 2 project and the AONS outcome. I raised the possibility of future collaboration between those involved in the JHOVE 2 project and those working on the future development of



Jock Phillips and Stewart Wallace

AONS. We were all in agreement that there was some overlap between the tools, and a definite potential for collaboration and future interoperability.

One of the other key reasons for making the trip was to participate in the Global Digital Format Registry (GDFR) Governance workshop (<http://www.archives.gov/era/research/global-digital-format-registry.html>), hosted by NARA in Washington DC on 14 and 15 November. The purpose of the workshop was to outline the issues to be addressed in the formation of governance for the operation of the format registry in the global community. As such, the workshop was structured as a set of presentations relating to the need for a registry of format information in a variety of disciplines ranging from scientific to cultural. There were also a couple of presentations that were specifically describing research and issues relating to collaboration outside the context of the GDFR. The workshop was certainly successful in giving all participants a good understanding of the extent of the application of the future GDFR network, as well as producing a list of issues to be addressed in the setting up of governance.

After the GDFR workshop, I had made arrangements to meet with Library of Congress staff who work with their file formats website, as well as other technical staff. Before I went there, I discussed the AONS project and software with someone from Tassella, who has previously been involved in working on PRONOM for The National Archives (UK) and is now involved with the PLANETS project in Europe. He was very interested in the work done and also in collaborative opportunities for the future.

Library of Congress staff were also very impressed with AONS. The key reason for meeting with them was that they are changing the structure of the information they publish, and which AONS uses. They are willing to share the content in XML format for the purposes of use within the AONS software, once they have completed the data migration work. They also agreed in principle to provide resources to build an AONS adaptor for the new XML data structure. This informal agreement was dependent on some resource being left at the end of their data migration project. However, I think it shows the kind of impression that AONS, and APSR, has made in the international community.



Matthew in front of Library and Archives Canada's Gatineau Preservation Center

Lastly, I'd like to thank the APSR and the NLA for supporting my trip. The response that both AONS and the METS profile have received in the international community is a testimony to the efforts of those who have been involved throughout the projects.

■ Hidden heroes – Tran Le



Our Hidden Hero for December is Tran Le, who joined Australian Partnership for Sustainable Repositories (APSR) late in 2007. Tran studied Computing Science at the University of Sydney, but his interests lie elsewhere, he explains:

Pianoforte and I have been friends for a long time :) I've also played the guitar for many years as well. My pride and joys are my Variax, Maton and custom Gibson Les Paul guitars, and of course my piano. I particularly interested in new flavours of music (usually fusing many different styles but still recognisable as music!) that push the envelope.

My other passion is for flying. I've been a part of the aviation industry for a good number of years, until other commitments whisked me away to Canberra. I might return to general aviation in the near future...perhaps something based here in the Capital Territory. In the meantime, I make do with riding my motorcycle while making whooshing aircraft sound effects!

My experience working for APSR and the Federal Government has been a rewarding experience. There's a real synergy and enthusiasm here that is definitely infectious!. My role as the technical business analyst and webmaster keeps me on my problem-solving toes and my team is here to steady me whenever I stumble.

Report from the field – Beijing

David Pearson from the National Library of Australia was able to attend iPRES2007 in Beijing with the help of APSR. His report is below:

I was given the opportunity to attend iPRES2007 (International Conference on Preservation of Digital Objects) (<http://ipres.las.ac.cn/>) which took place at the National Science Library, Chinese Academy of Science, in Beijing on 11-12 October 2007. Whilst in China I aimed to accomplish a number of tasks for APSR. The first was to acquire for Margaret Henty a suitable photograph for this travel report and present a paper on the AONS II (Automatic Obsolescence Notification System – version II) software developed by the National Library of Australia (NLA) in collaboration with APSR.

Naturally, both tasks required some effort to complete.

I arrived in China a number of days before the conference commenced. During this time I visited a number of tourist sites with the goal of acclimatising and finding an appropriate location for photographs for APSR and the introduction to my conference paper. The most exciting and memorable event of the pre-conference activities was a five hour walk along the Great Wall between *JinShanLing* and *Simatai*. This 5km (as the crow flies) stretch of wall is located around 120km north of Beijing. During this time, I took a number of photographs, using one in my conference presentation.

The second, and most important part of the trip, was presenting a paper at iPRES2007. The 200 conference participants consisted of 40 international delegates from European, African and North American libraries and archives. I was the only person from the Oceania region. The remaining delegates consisted of around 160 Chinese representatives of Libraries and Archives all over China. The sessions and papers were diverse, fitting into the following categories: Infrastructures for Digital Curation; Institutional Repositories; Preservation Strategies and Policies; Trusted Repositories & Certification; Ingest, Migration and Emulation; and Education and Training. My presentation was in the session called "Preservation of Research Data – Recommendations and Challenges for Digital Curation".



The paper I gave was entitled "AONS II: continuing the trend towards software 'Nirvana' ". This paper reported on the AONS II project, undertaken by the NLA in conjunction with APSR in 2007. The project aimed to develop a software tool which allowed users to automatically monitor the status of file formats in their repositories, make risk assessments based on a core set of obsolescence risk questions, and receive notifications when file format risks change or other related events occur. I stressed that the prototype software was in Beta release and needed further funding to become a mature, suitable and sustainable tool for Repository Managers. I showed screen shots of the APSR federated and local deployment modes and spoke about the

file format obsolescence risk questions which developed as part of the project. Both the software and risk questions were well received. I was grateful that I had listened to my colleagues who said, "don't do live software demonstrations, work with animals or children" as my laptop died the night before the conference.

Pulling it all together

It is pleasing to see a positive result for a series of APSR projects drawing to a close in 2007. As the name suggests, the Repository Interoperability Framework (RIFF) initiatives aim to design systems that integrate current work practices into repository use, regardless of the repository platform being used (such as DSpace or Fez+Fedora). The goal is to prevent 'double handling' of digital content that can be automatically incorporated into repositories, such as online journals.

One of the projects in the series harnessed the Open Journal System (OJS), which is an open source web-based e-journal system developed by the Public Knowledge Project in Canada. The team, managed by the University of Sydney, have now developed a workflow allowing journal editions (once complete) to be migrated into DSpace or Fez+Fedora. The nature of repositories means the journals editions then have the benefits of long-term preservation, persistent identifiers for bibliographic control, and common search/discovery mechanisms.



Other workflow projects include Open Conference System (OCS) version 2 materials, image collections and document submissions.

The conference project addresses a long-standing problem with the dissemination of conference papers. While conferences are the primary or a major method of scholarly communication in a number of disciplines such as computing, engineering and architecture, using traditional print publication of proceedings can mean a long lag time. The papers can also be very difficult to find. The dissemination packages published papers for archiving. This project used a similar system to the OJS project to automatically ingest the conference outputs into DSpace and Fez+Fedora repositories.

Another project looks at the proper management of digital images used in teaching and research. Images have specific needs in storage, so having clear metadata on copyright, access control and preservation needs is important. This project turned to the National Library of Australia's schemas to develop a proper workflow to specify what is required to store digital images correctly. The result is a system which allows image collections to be moved between repository systems such as iSpheres, DSpace and Fedora to enhance sustainability and preservation.



Regular readers of the APSR newsletter will be familiar with the FieldHelper tool, which allows researchers to capture and maintain raw data collected in the field. This part of the RIFF project has developed a system to organize, index, preserve and present portfolios of field work. Underpinning the RIFF projects was the highly successful METS program, discussed in the October issue of the newsletter (<http://www.apsr.edu.au/news/newsletter2007oct/newsletter.pdf>).

Much of the RIFF work is invisible to the regular user of a repository. In order to extract information out of repositories once it has been ingested, Leo Monus and the development team have developed a theme editor for Manakin, an XML-based repository user interface developed by Texas A&M. The theme editor is a software application that can be used with Manakin-compliant software packages, that is, those packages that can produce XML in the way Manakin expects. For example, with the OJS project, the data is taken from the repository and presented in attractive, user-friendly ways that can also be specifically 'branded' to reflect institutional and other affiliations. Similar themes are being developed for image collections and conferences.



These packages have been tested on DSpace and Fez+Fedora, and have been designed to not preclude other platforms. The software packages will be available as open source progressively until the end of the year and will be placed on the APSR website. Updates will be made over the APSR Announcements email service.

Information on the RIFF projects can be found at: <http://www.apsr.edu.au/currentprojects/index.htm#riff>

Coffee anyone?

Imagine Matthew Walker's surprise, when on the way to a meeting with people from MIT, came across this café. COSI is famous world-wide!

The COSI-Framework is a PHP/PostgreSQL web application framework which has been designed and built to provide shared and consistent functionality across the projects of the Australian Partnership for Sustainable Repositories (APSR) Collection Services and Infrastructures (COSI) initiatives of 2007. The APSR COSI Framework Documentation is now available, details below.



■ **Around the Traps....** Send your suggestions to contact@apsr.edu.au

Conferences and events

2008 30-31 January, DRF International Conference 2008, Open Access and Institutional Repository in Asia Pacific, Osaka, Japan

https://www.tulips.tsukuba.ac.jp/DRFIC2008/index_en.php

2008 5-7 February, VALA2008: Libraries/Changing Spaces, Virtual Places, Melbourne, Australia

<http://www.vala.org.au/conf2008.htm>

2008 1-4 April, Open Repositories 2008, Southampton, United Kingdom

<http://or08.ecs.soton.ac.uk/>

2008 21-23 April, Fourth Nordic Conference on Scholarly Communication, NCSC 2008, Lund, Sweden

<http://www.lub.lu.se/en/ncsc.html>

2008 25-27 June, ELPUB, 12th International Conference on Electronic Publishing, Toronto, Canada

<http://t08.cgpublisher.com/>

Software

■ Visit the new APSR software site at <http://www.apsr.edu.au/software.htm> and download software from recent APSR projects. All of the software for download below is licensed under the Apache License, Version 2.0 (the "License"); you may not use this software except in compliance with the License.

Publications (reading)

■ James Blanden, APSR, Nov 2007, *COSI-Framework v1.0 Documentation*,

Downloadable from: <http://www.apsr.edu.au/publications/projects.htm>

■ James Blanden, APSR, Nov 2007, *ORCA Collections Registry Reports*,

Downloadable from: <http://www.apsr.edu.au/publications/projects.htm>

■ James Blanden, APSR, Nov 2007, *Report of the METS Profile Development Project*,

Downloadable from: <http://www.apsr.edu.au/publications/projects.htm>

■ The ANDS Technical Working Group, DEST, Oct 2007, *Towards the Australia Data Commons – a proposal for an Australian National Data Service*.

Earlier in the year DEST's National Collaborative Research Infrastructure Strategy (NCRIS) published an investment plan for "Platforms for Collaboration" which included significant Federal Government investments in high performance computing, an interoperation and collaboration environment, and a national data infrastructure. The investment plan for the national data infrastructure recommended the establishment of an Australian National Data Service (ANDS). This new document more clearly defines the rationale and activity areas for ANDS. ANDS activities are expected to start in 2008.

Report downloadable as pdf: <http://www.pfc.org.au/twiki/pub/Main/Data/TowardstheAustralianDataCommons.pdf>

Next issue?

Be the first to know ... join the APSR Announcements email service at:

http://mailman.anu.edu.au/mailman/listinfo/apsr_announcements

Contributors to this newsletter: Danny Kingsley, Margaret Henty, Scott Yeadon, Tran Le, Matthew Walker & David Pearson.

The APSR partners



Supported by



An Innovative Action Plan for the Future.



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License:

<http://creativecommons.org/licenses/by-nc-nd/2.5/au/>