



Sustaining Repositories

newsletter of the APSR Project

February 2008

www.apsr.edu.au

ISSN 1834-1098

In this issue:

- APSR looks to the future
- Establishing ANDS
- Open Access Collections
- METS podcast and vodcast
- Surveying Data Management and eResearch
- APSR goes to Washington DC
- Scholarly Communications in Chicago
- DSpace Newsletter
- Lots of new project reports
- APSR software now available
- Useful aids for data curation
- Meet the team
- Around the Traps

■ APSR looks to the future

There are changes afoot in APSR where 2007 projects have been completed, the last federal funding has been received from the Strategic Infrastructure and the future of the partnership is considered. APSR funding was given originally for three years, to cover 2004 to 2006, and was then extended to cover 2007. This allowed APSR to fund its ongoing program of development and outreach activities.

The focus by APSR on outreach and training has been very successful, as can be seen by the very positive feedback received from participants, and to close down this program would leave something of a void in the Australian repository world. It has been decided, therefore, to continue these parts of the APSR agenda with some residual funding. Planning for this renewed focus is now in train and we would be more than happy to receive ideas for training and other activities. The partners will continue their involvement with APSR and provide input for future activities.

The APSR development portfolio projects for 2007 are now, for the most part, complete. Reports of these projects and the open source software developed are available. A full listing of new reports and software can be found on pages 6 and 7 of this newsletter. There is a new menu item on the APSR website for access to software – see <http://www.apsr.edu.au/software.htm>.

The APSR website will be maintained. It continues to be heavily accessed with substantial downloads of reports, presentations and podcasts. APSR_announcements will continue to be functional for announcements of interest. The APSR blog will, however, be discontinued.

■ Establishing ANDS

The period from January to June 2008 will see plans made for the establishment of the new Australian National Data Service (ANDS) to commence in July. APSR staff were actively involved in the planning which led to this establishment phase. *Towards the Australian Data Commons* was published in October 2007 and provides an outline and rationale for the new service (see <http://www.pfc.org.au/twiki/pub/Main/Data/TowardstheAustralianDataCommons.pdf>)

The Australian Government has contracted Monash University to undertake the ANDS Establishment Project with ANU and CSIRO. This six month project aims to open ANDS for business by July, 2008.

Two projects commenced in 2007 by APSR are continuing at the Australian National University with funding from the ANDS Establishment Project. These are the Collections & Services Registry and the Aggregated Statistics Service.

Adrian Burton and Vic Elliott of the Australian National University are serving on the Project Management Committee for the ANDS Establishment Project. Adrian Burton, Chris Blackall and Margaret Henty are all members of ANDS-ON, the working group charged with a range of tasks including development of a business plan.

Open Access Collections

Over eighty people assembled at Customs House in Brisbane on February 14 to investigate the issue of open access. Participants came from every state of Australia and from areas as diverse as federal and government departments, libraries, research departments and IT centres.



The day got off to a rousing start with a lively presentation by Professor Peter Murray Rust from Cambridge University. Peter has long been an exponent of open access, particularly for data. He discussed the difference between open access and open source saying that the former is a political term, while the latter is not. He talked about the importance of licenses and agreements in some detail, paving the way nicely for the two speakers from OAK Law later on the program, Professor Anne Fitzgerald and Dr Scott Kiel-Chisholm. And he concluded with the suggestion that one good place for universities to start is with the mandatory deposit of theses, avoiding any publisher copyright issues on the one hand, while introducing students to the importance of open access on the other.

Leanne Harvey from the newly created Department of Innovation, Industry, Science & Research reiterated the importance placed by the government on open access and reassured everyone that ASHER funding will be continued in 2008. Danny Kingsley, now completing a PhD at the Australian National University on barriers to open access, gave two take-home messages: mandates are a good start, and considering your audience is vital. This theme was taken up by Professor Ray Frost, a chemist from QUT who uses their open access repository extensively & has found that it has increased his citation count markedly. This is also the case for Professor Hubert Chanson, an engineer from the University of Queensland, who is an active user & beneficiary of UQ eSpace.

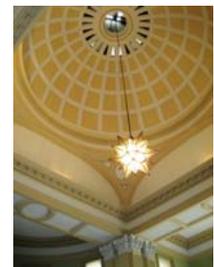
Kevin Stranack from the Public Knowledge Project in Canada was the final speaker for the day and talked about practical assistance for open access through the development of software to make it all easier. Both the Open Journal System (OJS) and the Open Conference System (OCS) are in use in a number of Australian universities (and many more around the world), and they are working on further innovations that will help to ease the path to open access.



A panel session in the afternoon covered practically every issue known to the repository manager. Questions covered meeting user concerns, the HEW level of staff who process deposits, handling datasets, the use of creative commons licenses, impediments to the furthering of open access and more.

Feedback from the day was positive and included comments such as

- I hope you continue to always ask researchers to come along and talk about their OA or repository experiences - these in both Forums I've attended have been very valuable
- Being new to the area of digital repositories, I found this workshop of immense value, not only listening to the presentations but talking with those experienced in the area.



Photographs on this page. Top: Paula Callan (QUT), Ray Frost (QUT), Peter Murray Rust (Cambridge). Middle: The panel - Margaret Henty (APSR), Paula Callan (QUT), Belinda Weaver (UQ), Simon McMillan (UNE), Vicki Picasso (U Newcastle). Bottom – John Byron (Australian Academy of Humanities); Kevin Stranack (Simon Fraser University); lunch; chatting; the beautiful domed ceiling of Customs House.

Many of those at the forum were interested in finding out what has happened to the RQF. Leanne Harvey from the Department of Innovation, Industry, Science and Research was sorry to have to disappoint the audience in not having up to date information about developments but these have since appeared. A press release from



Peta Hopkins (Bond University) and Danny Kingsley (Australian National University)

Senator the Hon Kim Carr has now announced plans for a new research quality and evaluation system to be known as Excellence in Research for Australia (ERA). The press release can be found at

<http://minister.industry.gov.au/SenatortheHonKimCarr/Pages/NEWERAFORRESEARCHQUALITY.aspx>.

Open Access Collections also hit the blogosphere in a big way. See <http://wwwm.ch.cam.ac.uk/blogs/murrayrust/> and <http://librariesinteract.info/>. Peter Suber's blog contains both comment and the presentations list and Charles W Bailey picked the presentations up for his online bibliography. For those who missed it, the event was recorded on video so that podcasts and vodcasts will be become available in due course. All the overheads, plus papers by Peter Murray Rust and Hubert Chanson are already available at the APSR website at http://www.apsr.edu.au/open_access_collections/presentations.html.

[ations.html](http://www.apsr.edu.au/open_access_collections/presentations.html).

■ METS podcast and vodcast

Judith Pearce and David Pearson of the National Library gave a two hour presentation at The Australian National University in early December, 2007 about the Australian METS Profile. Their journey through a forest of acronyms, including METS, MODS, PREMIS, MIX, DIPs and SIPs, held the audience fascinated for nearly two hours. Those who are part of the Australian data commons need to know about these standards and how they work together to ensure interoperability in the Australian repository environment.

This presentation is now available, not just by podcast, but also by vodcast or streaming. Overheads can be downloaded too from the APSR website at <http://www.apsr.edu.au/nla-mets/mets-presentation.htm>.

■ Surveying Data Management and eResearch

In late 2007, the University of Queensland undertook a survey of data management practices among the university's researchers. This was done in response to the increasing realisation that repositories need to include research data, in addition to the research outputs in print form already included, and to provide information which would enhance the support provided for those engaged in eResearch. The survey was carried out using the Apollo software developed at the Australian National University and adapted by APSR. Two other universities, the University of Melbourne and the Queensland University of Technology, have now replicated the survey among their own communities, while adding some questions of local interest.

The survey covered questions such as the types of digital data being created (spreadsheets, documents, experimental data, images, fieldwork data, etc), the size of the data collection, software used for data analysis, data storage and backup, application of a data management plan, roles and responsibilities around data management, copyright frameworks, usage of high capacity computing, and much more.

The three surveys had high response rates, and the results have been summarised for internal use within each university. The results show a remarkable consistency between the three universities. The picture emerging is one of enormous variability within each of the universities, with some researchers showing considerable expertise and confidence in data management and others showing a need for, and in some cases actively requesting, further support. The results will be aggregated and published later in the year.

If there is anyone who would like to replicate the survey in their own institution, and contribute their results to the national picture of data management practice, a copy of the University of Queensland survey can be found on the APSR website at http://www.apsr.edu.au/currentprojects/data_management_at_uq_blank_survey.pdf. The survey can be customised for local application and the data collected and analysed using Apollo. For further information, contact Margaret Henty (Margaret.henty@apsr.edu.au).

■ APSR goes to Washington DC

December was a busy month for Adrian Burton, Chris Blackall and Margaret Henty as they set off to Washington DC for meetings and conferences. High on the agenda was a meeting to discuss the establishment of a global network of interoperable registries. "Why do we need a global network of registries?" you might ask.

A major project for APSR during 2007 and into 2008 has been the creation of the Online Research Collections Australia (ORCA) registry. This is designed to improve the ability of institutional repositories, archives and data centres to create and share information and resources. Its aims are to develop a discovery portal for information about research collections and a services registry that can improve collection management.

Research does not respect national boundaries, so it would make sense to create a network to link registries in other countries. Some are already being created, such as the OCKHAM Initiative in the USA and the Information Environment Service Registry (IESR) in the UK.

As well as your three intrepid APSR representatives, the meeting was attended by Jeremy Frumkin (Oregon State University), Todd Carpenter (NISO), Kris Carpenter Negulescu (Internet Archive), Dan Rehak (eFramework), Dave Carpenter (National Science Foundation), Lucy Nowell (National Science Foundation), Ann Apps (IESR), Catherine Kott (DLF Aquifer), Andrew Treloar (ARCHER) and Andreas Stanescu (OCLC). An apology was received from Herbert van de Sompel (Los Alamos).

The meeting addressed two key themes: the desirability to stakeholders of a global network of interoperable registries; and the mechanisms for getting it into place. Current registry solutions are constrained by the fact that they have limited (local or national) scope and are typically implemented or configured as standalone applications, with little or no interoperability with other information systems. The goal should be to transform these stand-alone registries into a network of federated registries with global reach.

There are many challenges to be addressed in the process of creating a global registry. For example, not all data is currently put into a repository, intellectual property and rights management issues would need to be addressed, more countries than those represented at the meeting would need to be included, collection description issues abound, how to deal with the relationship between data collections and data processing, and so on.

Metadata issues were singled out for lengthy discussion. Although metadata standards for registry developers are improving, there are still issues associated with interoperability of registries with third party software applications through the use of Web Services protocols and standards. Governance, policy and stakeholder engagement strategies were discussed with as many questions as answers

In order to move the initiative forward, an interim coordination group was established: with Jeremy Frumkin to coordinate the US stakeholder involvement, Ann Apps to coordinate the UK stakeholders and Adrian Burton to coordinate Australian stakeholders. This group will seek opportunities to publicize the initiative at international conferences and events, investigate the possibility of having workshops during 2008 with funding from NSF and ANDS and establish working groups to address specific tasks and issues where relevant.



Global registry meeting, Renaissance Hotel, Washington DC. L-R: Lucy Nowell, Anne Apps, Andrew Treloar, Adrian Burton, Chris Blackall, Andreas Stanescu, Jeremy Frumkin (standing), Kris Carpenter Negulescu, Todd Carpenter, Dan Rehak.

The Digital Curation Centre Conference and the Coalition of Networked Information (CNI) Fall Meeting were both held in the same week as the Global Registries meeting. APSR was represented on the program of the DCC Conference by David Pearson of the National Library of Australia (photo right) who spoke on the topic of AONS, the Automated Obsolescence Notification System. His paper, "*Defining File Format Obsolescence: a risky journey*" was written in association with Colin Webb.

Also on the program of the DCC Conference was Dr Rhys Francis, Executive Director of the newly formed Australian eResearch Infrastructure Council (AeRIC). Dr Francis was one of four speakers, from Australia, the United States, United Kingdom and European Community who spoke on national perspectives for data management and curation.



■ Scholarly Communications in Chicago

In December 2007, the Association of Research Libraries (ARL) and the Association of College and Research Libraries (ACRL) held the first regional Institute for Scholarly Communication, hosted by the Consortium of Academic and Research Libraries in Illinois (CARLI) in snow-covered Chicago. This immersive learning experience was designed to prepare participants to become local experts within their libraries.

The thrust of the workshop was the need to address the ever-increasing costs of academic journals by offering an alternate means of access through open access publishing in institutional repositories. Two of the attendees were Margaret Henty of APSR and Lorena Kanellopoulos, Manager of the ANU ePress. Their aim in attending the Institute was to develop a plan to publicise scholarly communications at ANU, and to learn more about the issues from a North American perspective.



Lorena Kanellopoulos, Julia Blixrud and Margaret Henty

Nearly one hundred participants attended the Institute, coming from all over the USA. While many came from libraries, others were faculty members and administrators on the front line. The issues covered in the workshop included copyright, the researcher perspective, publishing models, the legislative and policy environment and program evaluation and reporting.

Among the presenters were Julia Blixrud, Assistant Executive Director (External Relations) of ARL and Assistant Director for Public Programs for SPARC. The Scholarly Publishing and Academic Resources Coalition, Maria Bonn of the Scholarly Publishing Office of the University of Michigan and Joy Kirchner from the University of British Columbia Library.

■ DSpace Newsletter

Users of DSpace will be pleased to know that there is now a regular DSpace newsletter. The first issue appeared in January 2008. In introducing NewSpace, Michelle Kimpton, the Executive Director of DSpace, wrote

Happy New Year and welcome to the first edition of the DSpace Foundation Newsletter! Through this monthly newsletter, we will keep our community informed about current news and events in the DSpace digital archive world. In each issue we will highlight DSpace installations and contributors. We will also provide updates on DSpace software development and projects sponsored by the Foundation and others in the community. The newsletter will link to relevant conference and event information.

The newsletter, and subscription information, can be found at

http://www.dspace.org/index.php?option=com_content&task=view&id=512&Itemid=171.

■ Lots of new project reports

November and December were busy months for APSR staff and partners as many projects came to completion and many reports were written and published.

You are sure to find something of interest among this list. They are all available in pdf format on the APSR Publications page at <http://www.apsr.edu.au/publications/projects.htm>.

Report of the METS Profile Development Project, published November 2007, by Judith Pearce, David Pearson and Megan Williams (National Library of Australia) and Scott Yeadon (Australian Partnership for Sustainable Repositories)

RIFF Submission Service, published November 2007, by Peter Raftos and Scott Yeadon (Australian National University)

The Digital Scholar's Workbench Project: Final Report, published December 2007, by Ian Barnes (Australian National University)

ANU DSpace Dissemination Service, published December 2007, by Peter Raftos and Leo Monus (Australian National University)

Field Helper 2007 Report on Activities, published December 2007, by Steven Hayes (University of Sydney)

APSR Image Workflow Final Report, published December 2007, by Tristan Gutsche (University of Sydney)

Project Report W2 Conferences Workflow, published December 2007, by Gary Browne (University of Sydney)

Project Report W1 Journals Workflow, published December 2007, by Gary Browne (University of Sydney)

Fez/Fedora Support for Submission and Dissemination Services as part of the APSR RIFF Framework, published December 2007, by Andrew Bennett, Christiaan Kortekaas, Lachlan Kuhn & Matthew Smith (University of Queensland)

COSI-Framework v1.0 Documentation, published November 2007, by James Blanden (Australian Partnership for Sustainable Repositories)

ORCA Collections Registry Reports, published November 2007, by James Blanden (Australian Partnership for Sustainable Repositories)

This is a series of four documents which together describe the technical aspects of the ORCA Collections Registry project. These comprise: ORCA Registry v1.0 Documentation, ORCA Registry v1.0 Implementation Notes, Database Schema v1.0 and ORCA xsd

Project Report of the Benchmark Statistics Service (BEST) published December 2007, by Laura Vallee (Australia Partnership for Sustainable Repositories)

■ APSR software now available

All of the software developed by APSR is now available through a new page on the APSR website at <http://www.apsr.edu.au/software.htm>. All software is open source and free to all. It is however, licensed under the Apache License, Version 2.0 so you may not use this software except in compliance with the license. This can be found at <http://www.apache.org/licenses/LICENSE-2.0>.

From this page you can find the software for the following, as well as links to further information about the projects on which they were based.

ANU DSpace: Dissemination Service Source Code

Automated Obsolescence Notification System (AONS) II - this notifies repository managers about formats within digital resources in their repositories and alerts them to potential problems relevant to obsolescence and long term usage.

AONS II COSI Module - requires an instance of the AONS application with which to interact.

Collection Services Infrastructure (COSI) Framework - this is a web applications framework built to provide authentication (via either LDAP or the built-in service) and roles-based access to modules of grouped functionality (or sub applications).

Field Helper – this is a desktop application that enables you to quickly view and categorise groups of related digital files and then submit the resulting package to a repository for long term preservation and access. It is designed for researchers working in the field.

Fez - an open source project to produce and maintain a highly flexible web interface to FEDORA for any library or institution to configure and publish or archive documents of any type sustainably.

Online Research Collections Australia (ORCA) Registry - provides for a registry of collection level (and associated service, party and activity) metadata that is based on the ISO 2146 draft standard. The ORCA Registry software consists of a PostgreSQL database that is managed/utilised by a PHP module housed within an instance of the COSI framework.

Repository Interoperability Framework (RIFF) COSI Module - this requires an instance of the RIFF application with which to interact.

Repository Interoperability Framework (RIFF) Submission Service Source Code - provides a service-oriented framework to support the packaging and routing of content and metadata from a source application to a target repository.

■ Useful aids for data curation

For those who are engaged in the business of data stewardship, there are two new documents which might be of help.

Michael Witt and Jake Carlson of Purdue University Libraries have produced a list of 10 questions that “a librarian can use as a starting point” for a data interview. They see this as a tool to assist with the identification of data sets to add to the repository and to help with the creation of the necessary infrastructure to do so. See http://docs.lib.purdue.edu/lib_research/81/

From EPrints comes a poster designed to persuade researchers to ensure that their data is properly managed and curated. It's a great poster with a strong message. Take a look at <http://www.eprints.org/resources/rgirr2.pdf>

■ Meet the team

In late November, the APSR staff met to celebrate the achievements of 2007 and, in some cases, to say goodbye. David Berriman has moved on to the Centre for Mental Health Research where he is using his expertise in designing a new software interface for a product under development. Laura Vallee has taken up a place at the University of Melbourne to study nursing. Danny Kingsley is completing her PhD in the Centre for the Public Awareness of Science at the Australian National University. The rest of the group, Adrian Burton, Chris Blackall, Margaret Henty, Scott Yeadon, Leo Monus, James Blanden and Fiona Nelson Campell are now engaged in continuing some of the work of APSR and contributing to the establishment of the new Australian National Data Service (ANDS).



*Front row (l-r) David Berriman, Adrian Burton, Leo Monus.
Centre: Danny Kingsley, Tran Lee, Fiona Nelson Campbell & Margaret Henty
Back: Laura Valle, Chris Blackall, Scott Yeadon & James Blanden*

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

Conferences and events

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>

■ **Publications (reading)**

The European Repository Landscape: Inventory study into the present type and level of OAI compliant Digital Repository activities in the EU by Maurits van der Graaf and Kwame van Eindhoven, DRIVER Studies, 2008. ISBN 9789053564103, <http://dare.uva.nl/aup/nl/record/260225>

The study presents a complete inventory of the state of digital repositories in the 27 countries of the European Union and provides a basis to contemplate the next steps in driving forward an interoperable infrastructure at a European level.

A DRIVER's Guide to European Repositories; Five studies of important Digital Repository related issues and good practices edited by Kasja Weenink, Leo Waaijers and Karen van Godtsenhoven, DRIVER Studies, 2008. ISBN 9789053564110, <http://dare.uva.nl/aup/nl/record/260224>

This study contains comprehensive and current information on digital repository-related issues that are particularly relevant to repository managers, decision makers and funding agencies.

The Investigative Study of Standards for Digital Repositories and Related Services by Muriel Foulonneau and Francis André, DRIVER Studies, 2008. ISBN 9789053564127, <http://dare.uva.nl/aup/nl/record/260226>

This study by CNRS, the French partner in the DRIVER project, reviews current standards, protocols & applications in the domain of digital repositories. Special attention is paid to the interoperability of repositories to enhance data exchange.

Journal of Electronic Publishing, Winter 2008. <http://journalofelectronicpublishing.org/>

Special issue on communications, scholarly communications and the advanced research infrastructure

Judith Pearce, David Pearson, Megan Williams and Scott Yeadon, "The Australian METS Profile – A Journey about Metadata", *DLib Magazine*, March/April 2008 (forthcoming) <http://www.dlib.org/>

David Pearson & Colin Webb, "Defining File Format Obsolescence: a Risky Journey", *International Journal of Digital Curation* (IJDC), Vol 3, No 1, June 2008 (forthcoming)

Contributors to this newsletter were Margaret Henty, Chris Blackall and James Blanden.

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/>