



# Conference Report: eResearch Australasia

University of Queensland, June 2007

## Contents

<b>Audience</b> .....	3
<b>Reference Documents</b> .....	3
<b>Version control</b> .....	3
<b>Review and Approval</b> .....	3
<b>Distribution</b> .....	3
<b>Introduction</b> .....	4
About eResearch Australasia .....	4
About this report .....	5
Acknowledgements .....	5
<b>Leadership, coordination and governance</b> .....	6
<b>Organisational change</b> .....	8
<b>Composition of the e-research community</b> .....	9
<b>Professional development</b> .....	11
<b>Usable sustainable technologies: meeting user needs; fostering uptake; software development and deployment processes; web services, Web 2.0 and social networks</b> .....	13
<b>Data storage and curation</b> .....	17
<b>Identity and access management</b> .....	21
<b>Workflows and semantic technologies (e.g. ontologies)</b> .....	22
<b>Appendix A: Sessions attended</b> .....	23
<b>Appendix B: Key technologies and standards noted</b> .....	24
<b>Appendix C: Initiatives noted</b> .....	25
<b>Appendix D: Other informal contacts made</b> .....	27





## Audience

The intended audience for this document are:

- REANNZ staff members
- REANNZ Board members
- members of the KAREN community.

## Reference Documents

No documents are referenced.

## Version control

VERSION	DATE	REASON FOR UPDATE	AUTHOR
1.1	14 Aug 2007		Sam Searle, Victoria University of Wellington

## Review and Approval

This document has been approved for release by the following:

NAME	ROLE	ORGANISATION	DATE

## Distribution

This document has been distributed to the following persons or parties:

NAME	ROLE	ORGANISATION

## Introduction

### About eResearch Australasia

eResearch Australasia took place at the University of Queensland in Brisbane on 26-29 June 2007. The conference was jointly organised by the Middleware Action Plan and Strategy (MAPS) and the Australian Partnership for Sustainable Repositories (APSR) projects, both of which are funded by the Australian Department of Education, Science and Technology (DEST).

The conference website is available at <http://www.eresearch.edu.au>.

More than 300 delegates from over 80 organisations (including 25 Australian and 5 NZ universities) attended. Around 9% of attendees were from government and 8% from industry. The programme was a busy one, with many concurrent sessions covering discipline-specific topics, e-research tools and technologies, and e-research governance and support structures, particularly those arising from DEST's National Collaboration Research Infrastructure Strategy (NCRIS) Platforms for Collaboration (PfC) programme.

Fellow attendees from New Zealand were:

- Mary Allan: University of Canterbury
- Nick Jones: BeSTGRID Project Manager, University of Auckland
- Gordon Mallinson: University of Auckland eResearch Taskforce
- Brian Opie: NZ Council for the Humanities; VUW E-Research Capability Subcommittee
- Jo-Anne Skinner: Otago University E-Research Advisory Group
- Alison Stevenson: New Zealand Electronic Text Centre; National Digital Forum
- David Thorns: University of Canterbury; Advanced Network Capability Building Advisory Panel.

The conference was a great opportunity for fostering the relationships amongst the NZ contingent, as well as providing networking opportunities with delegates from Australia and other countries, such as the UK.

Four of the NZ delegates presented at the conference, in two sessions on KAREN, BeSTGRID and trans-Tasman collaboration (Nick Jones and Sam Searle) and on Access Grid technologies and social science research (David Thorns and Mary Allan).





## About this report

This report takes a thematic approach to the insights gained during the conference, and covers:

- Leadership, coordination and governance
- Organisational change
- Composition of the e-research community
- Professional development
- Usable sustainable technologies: Meeting real user needs; fostering uptake; software development and deployment processes; the move towards web services, Web 2.0 and social networks
- Data storage and curation
- Identity and access management
- Workflows and semantic technologies (e.g. ontologies)

Suggestions for follow-up in New Zealand are made at the end of each section. Many of these suggested follow-up actions cover similar territory to the *Advanced Network Capability Building Roadmap 2007-2009* and could be mapped against the prioritised outcomes in that document.

Appendices cover sessions attended, technologies and initiatives of interest, and informal contacts made during the conference.

## Acknowledgements

My attendance at this conference was made possible by a travel grant from the KAREN Capability Build Fund, and the additional support of Victoria University of Wellington.

Sam Searle

E-Research Development Coordinator, Victoria University

14 August 2007



## Leadership, coordination and governance

- Government agencies at both federal and state level in Australia are fully engaged with e-research, and are investing heavily in this area, in collaboration with funding councils.
- Much of the conference focused on the National Collaborative Research Infrastructure Strategy (NCRIS), funded by DEST. This programme will clearly be the driver for e-research in Australia, just as various JISC programmes have driven e-science and e-research in the UK. It is significant not just for the amount of money involved (AUS\$500M+, which will be matched by research organisations, state governments and business) but also for the development of a new, consensus-driven investment decision-making process.
- E-research is seen by the government and by the research community as key for Australian research competitiveness (Mike Sargent).
- State government support is also strong for e-research, as evidenced by the number of state-based agencies such as VPAC (Victorian Partnership for Advanced Computing, VeRSI (Victorian eResearch Strategic Initiative) and QCIF (Queensland Cyber Infrastructure Foundation).
- Many initiatives require matching funding; NCRIS will rely greatly on matching contributions from universities and CSIRO.
- Joint ventures are a common way in Australia for organising collaborative initiatives that combine state and federal funding with contributions from research partner organisations like universities and CSIRO.
- AARNet is seen by the federal government's Department of Education, Science and Technology as foundational technology for e-research in Australia and as a "significant national investment" that will continue to be supported (Anne Marie Lansdown). The bulk of funding now is going towards developing capability other than the network.
- The business model for AARNet is moving to a subscription-based model rather than being based on traffic charges. A study at the University of Queensland had shown that the costs involved in monitoring and billing for usage made this less effective than a subscription-based model.
- JISC works closely with funding councils (including announcing joint funding calls), and are spending £10M on e-research infrastructure and £2M on virtual research environments.

### ***New Zealand follow-up***

- *Increase awareness and engagement levels within government agencies.*
- *Develop national leadership, coordination and governance structures.*
- *Build better relationships between policy agencies and research funding agencies to ensure that critical infrastructure and middleware is developed.*
- *Be prepared, within the KAREN membership, to match investment from government and to explore new business models such as joint ventures.*



## Organisational change

- “Cultural re-engineering” (the institutional factor) is required as well as capability (the human factor). There is a need to consider how to separate research structures from administrative structures; there are currently huge barriers to truly engaging in research across different units and across organisations, because processes are set up with schools and faculties in mind rather than collaborative activities (Mike Sargent).
- Lack of co-ordination within organisations is an issue: even though UQ is a leader in Australian e-research, their DVC Research regarded their activities to date as being “not particularly well coordinated” and saw this as an issue affecting other universities as well.

### ***New Zealand follow-up***

- *Address the e-research paradigm shift internally at KAREN member organisations, through strategic planning and removal of barriers to collaboration.*





## Composition of the e-research community

- The composition of the e-research community in Australia is quite different from in NZ.
- E-health (particularly at the interface between research and clinical practice) was much more integrated into wider discussions.
- It was evident from both the programme and the attendance list that cultural professionals such as librarians and archivists are very interested in e-research and in discovering their place within the e-research 'landscape'.
- The humanities and social sciences were well-represented in the programme, though less so amongst the general attendees.
- The greatest barrier to participation by the humanities is not cost, but is a split between the tech-savvy and the less tech-savvy humanities researchers. E-research is bringing about changes in expectations and work practices in the humanities that are more extreme and rapid than for the sciences. Technology is not as pervasive and the shift from lone researcher to collaborative research is only now underway. These changes are provoking anxieties and doubts.
- Humanities e-research relies on large-scale digitisation of primary resources; its development is therefore tied to the work programmes of cultural heritage organisations such as libraries, archives and museums.
- Potential collaborations involving the humanities and geospatial experts need to be explored.
- Individual humanities research centres are hotspots for humanities work and will be vital in bringing collaborators together.
- "In the humanities, the sensor networks are made up of people": this comment from the audience during a session by Paul Arthur resonated with many attendees.

### ***New Zealand follow-up***

- *Strengthen KAREN community involvement with the health sector.*
- *Explore how best libraries and librarians can contribute to e-research, in partnership with the National Library and the managers of New Zealand's many institutional repository projects.*

- *Build on the work of BRCSS in ensuring that KAREN is seen as essential infrastructure across a wide range of disciplines, including the humanities.*
- *Partner with the cultural sector to ensure investment in digitisation of analogue materials that meet the needs of humanities scholars: e-research development and KAREN uptake in the humanities depends on this.*
- *Stimulate interdisciplinary collaboration e.g. between humanities scholars and GIS experts.*





## Professional development

- A recurring theme was the value of skilled people and the need for rewards and incentives to encourage their professional development. Concerns about lack of professional recognition were expressed by senior researchers (Alex Szalay), new researchers (Paul Arthur) and e-research business development specialists (Paul Davis). There were many mentions of 'carrots and sticks', though few concrete examples of how governments or individual institutions were providing incentives.
- Many speakers noted the impact of the Research Quality Framework (Australia's equivalent of New Zealand's PBRF) upon e-research. Many in the Australian e-research community feel that current definitions of RQF outputs do not reward e-research activities such as software development, database creation and maintenance and so on. Paul Arthur summarised this by saying: "The core task of building resources is not considered as a research activity."
- There is also a need to support a new breed of support professionals. Several presenters noted the lack of career paths and reward structures for non-academic staff within university environments.
- Recruitment and retention is an issue in Australia that is growing. There was anecdotal evidence of 'poaching' of staff amongst the NCRIS projects, and turnover amongst technical and support staff will have a big impact on projects (Rhys Francis). There is a lack of personnel with skills linking research disciplines and ICT (Paul Davis, Mike Sargent).
- The e-research paradigm shift causes very real professional and personal anxieties and fears as people's roles and responsibilities change: "self-transformation is required" (Paul Arthur).
- There are generational and attitudinal social factors that will have an impact on uptake. Support for new researchers was recognised as key, but there is also a need to ensure that established staff have the skills and knowledge that they need.
- Paul Davis from VeRSI said that we needed to "get past the hype" and recognise that raising awareness is still the most important thing we should be working on, as only a minority of researchers are aware of the paradigm shift.

### ***New Zealand follow-up***

- *Urgently develop strategies for workforce development: recruitment, retention and career development are likely to be issues based on the Australian experience.*
- *Discuss with TEC how researchers' contributions to the development of infrastructure, tools, outputs (e.g. databases and multimedia works) and e-research culture will be recognised and rewarded within the PBRF.*



## **Usable sustainable technologies: meeting user needs; fostering uptake; software development and deployment processes; web services, Web 2.0 and social networks**

*This is not about technology; this is about making technology usable.*

Paul Davis, Victorian eResearch Support Initiative (VeRSI)

*Grids are not just communities of computers; they are also communities of researchers.*

David Abramson, PRAGMA / Monash University

*If you don't have a deployment process it's not infrastructure.*


Lindsay Hood, APAC Grid

- Many of the sessions noted the importance of documenting the real needs of users and ensuring that these are met.
- There is a need to move beyond the model of the 'builder-researcher' to a scenario in which e-research tools are accessible and usable by a majority of researchers without the need for high-level technical skills (Rhys Francis).
- James Farnhill and Ann Borda from JISC discussed the importance of use cases as a tool, while others (e.g. Paul Arthur) noted the importance of good user interface design. Iterative design processes and use of demonstrators are vital as these enable continual user engagement: researchers are very time-poor and need to see immediate benefits (Andrew Treloar).
- Flexibility and understanding user needs are key; one project (ARCHER) focused on end-to-end development but discovered that ultimately the value in this was not demonstrated as most researchers only wanted some of the tools that were available.
- There was much talk about 'horizontal' vs 'verticals' and the need for building applications and services that met the needs of disciplinary communities and research groups. The focus on building up 'horizontal' layers of fairly generic technology (e.g. grid services) has been important, but in future needs to be shifted to ensure that researchers can do their



work (i.e. the technologies have to be 'vertically' integrated in ways that make sense to the researchers).

- JISC has been working on a study of user priorities across three main areas: software, policy, and support. Recurring issues are: distributed file management and data curation; virtual organisations; tools and training; authentication, licensing, and other intellectual property issues; and tools. JISC are fostering community engagement via establishment of e-Science centres (now including centres for the social sciences and for the arts and humanities). They are focusing on building capability through the development of testbeds, support organisations and processes, policy groups, and governance structures.
- In working with users, JISC are looking to: identify clear community needs, support early adopters, encourage perceptions of the grid as a tool rather than as a technology, deploy more diverse 'plug and play' applications, provide user support and training, and develop strategies for marketing, awareness raising and outreach. JISC are utilising Geoffrey Moore's Technology Adoption Model to guide their approach.
- VeRSI is a five-year programme in Victoria aimed at "providing a coordinated and accelerated approach to uptake". Fostering uptake of e-research is all about change management: building trust, awareness, exemplars and enabling technologies (security and access management; storage infrastructure).
- Projects must move from development to deployment. Sustainability and scalability of services is seen to be crucial by the Australia government in funding programmes (Ann Marie Lansdown). Similarly, in the UK, all funded projects must take deployment and sustainability into account. JISC pilots are now routinely evaluated and projects are assessed on a "development to service" basis that takes into account business cases, needs and gap analyses, and cost-effectiveness (Ann Borda).
- There is clearly a shift away from grid services as they have been conceived in the past to web services (David de Roure, Lindsay Hood).
- There is scope for participation in international standards development and software development projects: small countries should be looking to leverage existing technology, not to develop from scratch. Lindsay Hood noted that standards work is not well connected to the work of on-the-ground practitioners, and this is a problem.

- 
- Faults are normal (hardware, software and network); 80% of NIMROD code is for dealing with faults and failures (David Abramson): we need to manage the expectations of users in new and experimental environments.
  - Technology cannot solve all the problems of distance: when you have users across 12 timezones (e.g. PRAGMA), things are still not easy.
  - Collaboration is not just about communication; collaboration takes place over *artefacts* (David de Roure): this has to be built in.
  - “Mashups are workflows”; real world users making data do what *they* want it to do (David de Roure). Drawing on examples from Web 2.0 and social software in the e-research space (e.g. MyExperiment as a “workflow bazaar” for sharing and re-use of scientific workflows), David suggested that the most benefits would be reaped when developers empowered researchers with collaborative working spaces and easy-to-use tools. He argued that we need to “create an ecosystem of participation”; Science Commons could be a model for this kind of research environment.
  - Phil Bourne spoke about the emerging generation of researchers: the leaders of the future are Web 2.0 savvy and are immersed in working with digital media. This emerging generation are likely to drive changes in scholarly communication. Current work at the Public Library of Science open access journal and its sister database Protein Databank is trying to better integrate data and the final publication (e.g. by enabling linking from a scholarly paper to original datasets and modelling tools); foster professional networking akin to social networking; and drive publication models towards new media paradigms (e.g. by enabling authors to contribute podcast presentations that serve as an audiovisual ‘abstract’ for their work). This work falls under the banner of the BioLit: Tools for New Modes of Scientific Dissemination project.
  - Alex Szalay talked about Web 2.0 technologies and social networking in the context of astronomy, arguing for a “Virtual Observatory Flickr”. He noted the emergence of bottom-up initiatives such as Life Under Your Feet (<http://lifeunderyourfeet.org>) and Galaxy Zoo (<http://galaxyzoo.org/>), which make data available freely via web services and involve participation from the general public in scientific endeavours.
  - Although there are positives to greater involvement from the general public in e-research endeavours, there may also be qualms from researchers concerned about the erosion of their status as experts (Paul Arthur).

### ***New Zealand follow-up***

- *Put greater emphasis on user needs analysis, use cases, and cost-benefit analyses to ensure that activities meet real user needs.*
- *Put greater emphasis on clear plans for deployment and support to ensure the sustainability of projects and services.*
- *Develop products and services in a web services framework.*
- *Investigate and deploy, where appropriate, Web 2.0 technologies that enable participation from the R&E community and from non-experts.*



## Data storage and curation

*It's a data world: data is the future.*

Andrew Treloar, ARCHER/Monash University

- The Australian government accepts that urgent action is required with regards to data storage and curation: a whole-of-government approach is emerging. The Australian government is establishing an Australian National Data Service (ANDS) as part of NCRIS at an estimated cost of AUS\$45M (\$21M to be provided by central government). An Accessibility Framework is planned to ensure that the public, government, business, and other researchers can access publicly-funded research (Ann Marie Lansdown).
- There are benefits to good data management, e.g. new research opportunities, less wastage of resources, social good (e.g. cultural heritage), and serious risks associated with poor data management (loss of irretrievable data, inability to verify research results). Data management requires changes in organisational cultures and increased capability amongst staff. (Marcus Buchhorn).
- Copyright, privacy and other intellectual property rights issues are a huge barrier and concern for many in the e-research community in Australia: the OAKLaw Project at QUT (Open Access to Knowledge) is attempting to address some of these issues, and there is interest in Creative Commons and its science-specific equivalent Science Commons.
- Science is evolving: from empirical to theoretical to computational, and now to data exploration: science now is about synthesising theories, experimenting and applying new algorithms to existing data as well as collecting new data (Alex Szalay).
- The size of datasets is increasing; this is driven by the emergence of generations of inexpensive sensors and compute power. The growth in data is outstripping the growth in compute power. Data in astronomy doubles every year; already there are several hundred TB. New simulation tools are adding to this: astronomy simulations can be 30TB at a time (Alex Szalay).
- Despite the emergence of very large datasets in some disciplines, most research data is still contained in small datasets that are manually maintained (Alex Szalay; Markus Buchhorn).

- The humanities are building up data too, in the form of online reference works, digital collections and humanities gateways: “The database is the aesthetic form of our era” (Paul Curtis).
- New data is easier to cope with than old data: we may be better off building policies and processes for the data yet to be created, than expending resources on ‘retrofitting’ existing data (Markus Buchhorn).
- The conference focused on data curation as a set of methodologies, skills and technologies required to manage research data. Data curation is related to, but distinct from, other types of information management (Andrew Treloar). The role of information professionals (e.g. librarians, archivists, data managers) is unclear. Some see data curation as an obvious extension of work already happening in institutional repositories, with libraries providing continuity of service, networks of useful relationships, and expertise in managing intellectual property. However, most libraries are not resourced or staffed to cope with the new demands of data archiving functions on top of their current activities.
- Depending on the discipline, there may be a need to combine information management skills with disciplinary knowledge in new types of ‘data scientist’ roles: e.g. BlueNet, a marine science data archive, has a dedicated data management professional with subject expertise.
- Managers of institutional repositories noted that aligning research output archiving with the administrative requirements of RQF reporting was vital. This minimised effort and ensured the sustainability of repositories by tying them to business processes perceived by senior management as critical.
- Several papers noted the importance of provenance: “The details of the origins of data are just as important as the values” (David de Roure). Provenance metadata supports the validation of results through repeating the research, as well as the sharing and re-use of the data.
- The need for clarity around responsibilities for retention and disposal was mentioned several times, from both the researcher perspective and the information manager perspective. An anecdote during the conference demonstrated the current state of play: when asked how he decided when to dispose of research data, a researcher replied that when he couldn’t read the media the data was on, he just threw it in the bin!
- In some disciplines (e.g. astronomy) there is a need to archive multiple ‘released’ versions of datasets, so that researchers have stable access to a





single dataset over time (e.g. for a PhD or research project, 3-5 years was a typical timeframe).

- There is a shift in thinking away from centralised archives towards distributed data storage and management; the focus is on finding data, enabling its re-use and then storing new outputs. A range of organisations can retain data, whilst curation can be done centrally as part of disciplinary or wider collaborative structures. Alex Szalay said that in astronomy, “Data will NEVER be centralised. It has to reside with the projects, where the analysis is taking place.”
- CSIRO is engaged in e-health projects that focus on distributed data. The goals of these are to improve researchers’ access to data and also to provide better care to patients and to improve medical training. The health system in Australia is very fragmented, with federal and state authorities responsible for different components of research and healthcare service provision. Data privacy is a huge issue, with non-aligned federal and state legislations, as well as varying organisational practices. The Health Data Integration project aims to bring various data sources together: it provides remote query processing (with privacy/security measures and audit trails built in) so that the data custodian maintains control of the data, but researchers and clinicians are still able to find and access databases.
- Another e-health data linking project is the Cancer Stage Interpretation System, which will link clinical and genomic data.
- An Australian Partnerships for Sustainable Repositories (APSR) project has interviewed researchers about their data archiving needs.
- E-research is producing new forms of intellectual output, including compound digital objects that may have components of varying genres, media types, file formats, network locations and relationships with each other. Jane Hunter is working on OAI-ORE standards for defining compound objects across multiple repositories using ‘resource maps’ constructed using standards such as RDF, ATOM, YADS (a simplified RDF for describing collections) and TriX (a way of expressing RDF ‘triples’ logic).

### ***New Zealand follow-up***

- *Urgently start investigating data storage, sharing and curation in the New Zealand context. Our Australian colleagues are in no doubt that this is THE key issue moving forwards, and will require technical solutions as well as skilled staff and significant cultural change. There is scope to learn from*

*the Australian experience and to keep an eye on the development of the Australian National Data Service as part of NCRIS PfC.*

- *Raise awareness, and address where possible the intellectual property rights and privacy concerns that pose a significant challenge to data sharing and re-use. There are a raft of issues – legal, technical and professional development – and all will require attention.*
- *Support and promote Creative Commons NZ.*



## Identity and access management

- The Australian Access Federation (which has grown out of the Meta Access Management System project at Macquarie University) has received around AUS\$5M to work on policies and governance, as well as technology deployment (Shibboleth/PKI). There will be a phased approach to enable early member participation.
- The intention is to be self-sustaining via member subscriptions in time.
- They have produced an animation about identity and access management that makes these issues very easy to understand.
- There is opportunity for NZ participation in three working groups (AU.eduperson; grid; levels of assurance). AAF are also seeking participation through feedback processes, conferences and training workshops, and mailing lists.
- AAF are working with David Bannon (APAC) on a migration plan from APAC grid certification to AAF grid certification.
- AAF are in discussion with browser vendors (e.g. Microsoft, Mozilla) with regards to AAF certificate integration with popular browsers.

### ***New Zealand follow-up***

- *Ramp up participation in the Australian Access Federation, e.g. through participating in working groups, attending AAF training events and running local capability building activities in the area of identity and access management.*

## Workflows and semantic technologies (e.g. ontologies)

*Workflows are the new rock 'n' roll.*

David de Roure, University of Southampton

- Workflows (e.g. TAVERNA, Kepler) reduce the amount of manual effort researchers need to put into experimentation, making the repetitive and mundane easier.
- There has been a high uptake (relative to other e-research tools) amongst researchers of workflow tools, but they need to be available on the desktop and easy to use.
- Key projects in this area included SmartTea and Coakting.
- As workflows are shared (e.g. through MyExperiment) they “transcend their application” (David de Roure) as other users add value.
- RDF/OWL and other ontologies featured in many presentations.

### ***New Zealand follow-up***

- *Build NZ expertise and applications in the area of workflows, e.g. Kepler, and other semantic technologies and standards, e.g. ontologies.*

## Appendix A: Sessions attended

Presentation slides for most sessions are available via the conference website at <http://www.eresearch.edu.au/program>.

- Viviani Paz (AAF): Pre-conference Australian Access Federation Forum
- David Siddle (UQ DVC Research): Welcome
- David de Roure (Southampton): eScience is about scientists too
- Phil Bourne (California): Thoughts on the future of scientific dissemination
- Lindsay Hood (APAC): National grid and applications
- Ian Atkinson (JCU): Sensor grids and ARCHER
- Jane Hunter (UQ): The Open Archives Initiative Object Reuse and Exchange (OAI-ORE)
- Anthony Maeder (CSIRO ICT Centre): Data linking and integration for health applications
- David Abramson (Monash): Grid enabling 'real' science and engineering
- Alex Szalay (John Hopkins): Science in an exponential world
- Paul Arthur (Curtin): Going Digital: humanities and the eresearch revolution
- Anne-Marie Lansdown (DEST): All aboard, destination seamless
- Rudolf Dimper (European Synchrotron): High performance computing for synchrotron radiation research
- Ann Borda (JISC): Expanded uptake and sustainable communities of use – JISC's role in shaping the UK eInfrastructure for Research
- Andrew Treloar (ARCHER/Monash): Supporting the eResearch lifecycle from acquisition through to annotation: the DART/ARCHER experience
- Paul Davis (VeRSI): Victorian eResearch Support Initiative
- Markus Buchhorn (ANU): The preservation & sustainability of research data
- Rhys Francis (NCRIS): Platforms for Collaboration
- Mike Sargent (MA Sargent & Associates): eResearch strategic positioning for the future
- Workshop: The researcher/librarian nexus: the challenges of research data management in institutional repositories

## Appendix B: Key technologies and standards noted

Technology	Comments
dSpace, Fedora and other repository tools	Currently used for institutional repositories but increasingly seen as a potential tool for wider data archiving initiatives.  e.g. Fedora will be used as the data management tool for the Distributed and Integrated Multi-Sensor Instrument Middleware (DIMSUM) toolkit.
Storage Resource Broker (SRB)	Mentioned along with SRM and dCache by APAC Grid staff.  Will be used by VerSI as part of an SRB federation – each member is supplying up to 100TB of storage each.
Gridsphere portals	Described by APAC Grid staff as “harder than expected” due to need for Enterprise Java knowledge and skills.
Globus	Used by APAC Grid but not exclusively; they also have 4 other stacks of grid technology from Europe.
Common Instrument Middleware Architecture	Part of NSF Middleware Initiative; aimed at grid enabling instruments as real-time data sources  Adopted in Australia for use with Great Barrier Reef Observation System.
Shibboleth/PKI	Now well-established in Australia, and the foundation technologies for the Australian Access Federation.
IAMSuite AND VOMS	Developed out of MAMS project for managing virtual organisations.  Will be utilised by VerSI (amongst others).
Sakai	Generic “virtual research environment” (VRE) that also provides e-learning functionality.  Being used by APAC Grid as partial solution (along with IAMSuite) for management of small and dynamic virtual organisations.
Open Geospatial Consortium (OGC)	Consortium of companies, government agencies and universities that develops specifications for spatial information.  Staff from the Australian DART/ARCHER project are involved in the OGC Remote Instrumentation Group.
OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	Standards for defining compound objects across multiple repositories using ‘resource maps’ constructed using RDF, ATOM, YADS (a simplified RDF for describing collections) and TriX (a way of expressing RDF ‘triples’ logic).

## Appendix C: Initiatives noted

Initiative	Comments
Australian Access Federation (AAF) <a href="http://www.aaf.edu.au">http://www.aaf.edu.au</a>	Will provide identity and access management services to Australian (and other) organisations.
CombeChem <a href="http://www.combechem.org/">http://www.combechem.org/</a> TAVERNA <a href="http://taverna.sourceforge.net">http://taverna.sourceforge.net</a> SmartTea <a href="http://www.smarttea.org/">http://www.smarttea.org/</a> Coakting <a href="http://www.aktors.org/coakting/">http://www.aktors.org/coakting/</a>	Various workflow projects, part of or related to the JISC eScience programme.
MyExperiment <a href="http://www.myexperiment.org">http://www.myexperiment.org</a>	Web service enabling the upload, annotation and sharing of workflows.
ScienceCommons <a href="http://sciencecommons.org">http://sciencecommons.org</a>	Exploratory project to apply the philosophies and activities of Creative Commons to scientific publishing, licensing, and data.
BioLit: Tools for New Modes of Scientific Dissemination Worldwide Protein Databank - <a href="http://www.wwpdb.org">http://www.wwpdb.org</a> Public Library of Science - <a href="http://www.wwpdb.org">http://www.wwpdb.org</a>	Project investigating new modes of publishing that will take advantage of digital media and try to integrate research publications with original datasets (e.g. through links from an article PDF to datasets and analytic/visualisation tools).
AuScope <a href="http://www.auscope.org.au/">http://www.auscope.org.au/</a>	Australian NCRIS-funded geosciences capability building programme; strong sensor network component.
NCRIS-IMOS <a href="http://www.imos.org.au/">http://www.imos.org.au/</a>	Australian NCRIS-funded marine science capability building programme; strong sensor network component, as well as emphasis on creating information architecture to support data sharing and preservation.
DART (Dataset Acquisition, Accessibility and Annotation e-Research Technology) – <a href="http://www.dart.edu.au">http://www.dart.edu.au</a> ARCHER (Australian Research Enabling Environment) – <a href="http://www.archer.edu.au">http://www.archer.edu.au</a>	DART investigated and developed requirements and software to help researchers collect, capture and retain datasets and streams; also focused on issues of intellectual property and publishing (including open access). ARCHER will productionise tools developed under DART to assist researchers in NCRIS priority areas.
SUPER: Study of Users' Priorities for e-Infrastructure for Research <a href="http://wiki.nesc.ac.uk/read/super?HomePage">http://wiki.nesc.ac.uk/read/super?HomePage</a>	JISC study into user requirements for e-research; primary goal is to radically improve usability and uptake of grid tools.

<p>VeRSI (Victoria eResearch Strategic Initiative)  <a href="http://versi.edu.au">http://versi.edu.au</a></p>	<p>Joint venture between Multimedia Victoria and Melbourne, La Trobe and Monash Universities to promote the uptake of eResearch.</p>
<p>BlueNet  <a href="http://www.bluenet.org.au">http://www.bluenet.org.au</a></p>	<p>National data management project in ocean sciences, providing federated access and data curation.</p> <p>They have a dedicated data management professional with subject expertise and 'data facilitators' at each organisation in the consortium.</p>
<p>Australian Partnership for Sustainable Repositories  <a href="http://www.apsr.edu.au">http://www.apsr.edu.au</a></p>	<p>Developing systems required for managing research data and information; initial focus seems to have been on institutional repositories but this is now broadening to encompass a wider range of activities relating to research assets, including datasets.</p> <p>One of their projects involved interviewing researchers about their data archiving needs.</p>
<p>HDI (Health Data Integration)  <a href="http://e-hrc.net/hdi/">http://e-hrc.net/hdi/</a></p>	<p>Application developed by CSIRO's eHealth Research Centre (~20 person-years of effort).</p> <p>Provides novel methods for linking multiple databases; allows data custodians to retain control of data, while queries and reporting are done remotely at data locations; manages privacy and security restrictions and has audit trails.</p>



## Appendix D: Other informal contacts made

Person	Topics discussed
Paul Arthur, Curtin University	Humanities and e-research; KAREN Forum and future NZ visits
David Bannon, VPAC	APAC/VPAC postgraduate internship programme; recruitment and retention; summer schools; NCRIS
Ann Borda, JISC	Uptake issues; JISC Middleware Assisted Take-Up programme (MATU); social sciences and humanities e-research
John Byron and Sarah Howard, Australian Academy of the Humanities	Trans-Tasman humanities coordination; survey of current usage and future needs
Paula Callan, E-Research Access Coordinator	Institutional repositories and research datasets
James Farnhill, JISC	Validity of MyProxy and OpenID as part of wider 'hybrid' IAM landscape; importance of use cases
Chris Hancock, AARNET	AARNET/KAREN; international link between Australia and NZ – coordinated approach required to improvement; AARNET participation in KAREN Forum
Ian Johnson, University of Sydney	Humanities and e-research; GIS and humanities; social software for humanities scholars
Dirk van der Knijff, Research Computing University of Melbourne	Research computing support in Australian universities