Australian Perspective: Opportunities and Responsibilities for Research Data

3rd International Digital Curation Conference
Curating our Digital Scientific Heritage: a Global Collaborative Challenge
Washington DC, 11-13 December 2007

Rhys Francis
Executive Director
Australian eResearch Infrastructure Council
Outline of Talk

• Government’s role
• Converging Agendas
• PMSEIC Data for Science Working Group
• Australian Government Framework Responses on Data
  – Research Quality Framework & ASHER
  – Accessibility Framework
  – Australian Code for the Responsible Conduct of Research
• National Collaborative Research Infrastructure Strategy
  – Platforms for Collaboration
  – Australian Access Federation
  – The Australian National Data Service
• Where to now?
Why Does ICT matter?

Our society was shaped by a truth: only people process information

- communication was between people
- everything made was hand made
- all decisions were a product of mind

Now, ICT gathers and processes information faster, more reliably and in inhuman quantities, and we have:

- economic globalisation
- remote monitoring and control

- integrated supply chains
- entirely new knowledge

And an emerging cyber world - understood by few

And as a by product we have eResearch:

- Using new technologies to create new knowledge in new ways
- Working on big-picture problems because for the first time we can
- Dealing with the loss of locality and the growth in participation
Government’s Role

• To engage stakeholders in building a trusted collaborative environment to collectively address issues of common interest
• To ensure that the policy and funding framework provides the right environment and incentives to encourage excellent research and foster greater collaboration and utilisation of data

Strategic Directions

• Research infrastructure and data should be accessible
  – to a range of users – research community, government, states and territories
  – under controlled, authenticated but simple access regimes
• Access should be independent of location – network infrastructure should therefore be robust and reliable
• Data should be discoverable, accessible, managed and long-lived
• The effect of the investments by governments and research communities in research and education should be maximised
## Converging Agendas

<table>
<thead>
<tr>
<th>Network</th>
<th>Data Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Australian Research and Education Network</td>
<td>• PMSEIC Data for Science Working Group</td>
</tr>
<tr>
<td>High Performance Computing Capability</td>
<td>• Accessibility Framework</td>
</tr>
<tr>
<td>• National Computational Infrastructure</td>
<td>• Australian Code for the Responsible Conduct of Research</td>
</tr>
<tr>
<td>Access &amp; Security</td>
<td>• Research Quality Framework</td>
</tr>
<tr>
<td>• Australian Access Federation</td>
<td>• Australian National Data Service</td>
</tr>
<tr>
<td>Coordination &amp; Collaboration</td>
<td>• The Australian Research Collaboration Service</td>
</tr>
<tr>
<td>• e-Research Coordinating Committee</td>
<td></td>
</tr>
<tr>
<td>• NCRIS Platforms for Collaboration</td>
<td></td>
</tr>
</tbody>
</table>
The PMSEIC** Data for Science Working Group

• Suggested a cooperative “Whole of Government” approach within a national strategic framework
• Key recommendations included:
  – National network of data repositories
  – Improved sharing and collaboration
  – Increase skilled workforce for best practice in data management
• The PMSEIC report is a key policy driver for future Australian investment in research infrastructure, particularly data management, federation and re-use


**PMSEIC: Prime Minister’s Science and Engineering Innovation Council
Research Quality Framework (RQF)

• Federal government funding has supported the development and deployment of digital repository technology

• The Australian Scheme for Higher Education Repositories (ASHER) is providing funding for
  – purchase of hardware and software to establish or update a repositories
  – support to meet the workload involved with populating institution’s repository

in order to make research outputs available for RQF assessment

• In opposition, the Australian Government expressed serious reservations about the proposed RQF, so “watch this space”

• However, the first round of payments to universities under ASHER has been made to support the deployment and population of institutional repositories
Australian Code for the Responsible Conduct of Research

- The objectives of the Accessibility Framework are also reflected in the new *Australian Code for the Responsible Conduct of Research*
- Published in 2007, the Code replaces the former Statement and Guidelines on Research Practice (1997) from the universities and research funding agencies
- It describes the responsibilities of institutions and researchers in the management of research data and primary materials
- Eg. Institutions are to retain research data, provide secure data storage, identify ownership and ensure security and confidentiality of research data

NCRIS: a National Collaborative Research Infrastructure Strategy

• NCRIS brings a strategic approach to Australia’s investment in research infrastructure
• Priority capability areas for support have been identified in the NCRIS Roadmap
• Investments will support core NCRIS goals: support research excellence, promote collaboration, provide national benefit
• NCRIS is a response to
  – Recognition of the need for a ‘small’ country to be strategic
  – Increasing importance of major research infrastructure to research and innovation
  – Increasing need for collaborative research
  – Increasing cost and complexity of research infrastructure
## NCRIS Investments

$540M** over the five years: 2007-2011

<table>
<thead>
<tr>
<th>Evolving bio-molecular platforms and informatics</th>
<th>Networked biosecurity framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated biological systems</td>
<td>Population health and clinical data linkage</td>
</tr>
<tr>
<td>Characterisation</td>
<td>Terrestrial ecosystem research network</td>
</tr>
<tr>
<td>Fabrication</td>
<td></td>
</tr>
<tr>
<td>Biotechnology products</td>
<td></td>
</tr>
<tr>
<td>Optical and radio astronomy</td>
<td></td>
</tr>
<tr>
<td>Integrated marine capability</td>
<td></td>
</tr>
<tr>
<td>Structure and evolution of the Australian continent</td>
<td></td>
</tr>
</tbody>
</table>

**Platforms for Collaboration (allocated $75M)**

**Note: scaled to EU or US economies - analogous to 1B USD per annum**
NCRIS follows a decade of investment

1997: High Performance Computing Committee
   – Established the Australian Partnership for Advanced Computing to provide access to high
     performance computing capability

2000: Advanced Networks Programme
   – Established advanced demonstrator networks

2002: Higher Education Bandwidth Advisory Committee
   – Established the Australian Research and Education Network Advisory Committee, and the
     Australian Research and Education Network

2004: Research Infrastructure Taskforce Report
   – Established the National Collaborative Research Infrastructure Strategy Committee to
     implement a program of strategic investment in research infrastructure

2006: eResearch Coordinating Committee Report
   – Outlined an integrated program of skills development and of middleware and computer science
     research

2007: NCRIS Platforms for Collaboration
   – Commitment to an infrastructure program covering computing, data and inter-operation
     components, and supporting the development of the Australian Access Federation
eResearch Strategy

**Thematic Issues**

- Continuing Need for a Focus
  - through national coordination
- Human Capabilities
  - People, skills and understanding
- Linkage of eResearch Resources
  - seamless access to resources
- Access to Data
  - best practice data management and curation
- Structural and Cultural Change
  - evolution of organisational structures and cultures
- Awareness and Support
  - develop researchers’ ability to adopt eResearch

**Service Clusters**

- Data
  - outreach, curation, data management
  - meta-services, location, access, movement
  - practice, providers and users
- Computing
  - capability computing facilities
  - national computing environment
- Interoperation
  - discipline services (tools ((software))
  - user and operations support
  - collaboration services support
- Access
  - the Australian access federation
  - the Australian research and education network
Platforms for Collaboration

The Data Commons
*Data Federations*

Capability Computing
*Advanced models*

Interoperation

Compute

Data

Access

Research connectivity
*Seamless reach*

Collaboration services
*Research workflows*
Investments

Australia National Data Service  Data Federations
• Develop user and owner frameworks for a data commons
• Develop and operate national registries and data sharing services
• Help Institutions connect repositories to the data commons
• Help researchers share data through the data commons

National Computational Infrastructure  Advanced Models
• Develop and operate a shared national computational facility
• Develop domain oriented advanced modelling capabilities

Australian Research Collaboration Service  Research Workflows
• Develop and operate services linking systems and resources nation wide
• Develop and operate collaboration and workflow tools for researchers

AREN and AAF  Seamless Reach
• Connect researchers and research resources at required bandwidth
• Develop a shared authorisation framework for access to research resources
Seamless Reach: AREN

- Connect researchers and research resources at required bandwidth
- Develop a shared authorisation framework for access to research resources
**Seamless Reach: Australian Access Federation**

- Connect researchers and research resources at required bandwidth
- Develop a shared authorisation framework for access to research resources

**Trust Federation**
Advanced Models: National Computational Infrastructure

- Develop and operate a shared national computational facility
- Develop domain oriented advanced modelling capabilities

Ecosystem focus

Biotech focus

NCI Peak Facility
- managed by ANU-SF
- national merit access

Geoscience focus
**Data Federations: Australian National Data Service**

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons

<table>
<thead>
<tr>
<th>Organisation/Project</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>APAC National Facility</td>
<td>APACF</td>
</tr>
<tr>
<td>ATLAS Experiment</td>
<td>ATLAS</td>
</tr>
<tr>
<td>Australian Bureau of Statistics</td>
<td>ABS</td>
</tr>
<tr>
<td>Australian Science and Technology Heritage Centre, University of Melbourne</td>
<td>Austehc</td>
</tr>
<tr>
<td>BlueNet</td>
<td>BlueNet</td>
</tr>
<tr>
<td>Bureau of Meteorology</td>
<td>BOM</td>
</tr>
<tr>
<td>Centre for Cultural Research at the University of Western Sydney</td>
<td>CCR</td>
</tr>
<tr>
<td>Centre for Public Culture and Ideas, Griffith University</td>
<td>CPCI</td>
</tr>
<tr>
<td>Council of Australian University Librarians</td>
<td>CAUL</td>
</tr>
<tr>
<td>CSIRO</td>
<td>CSIRO</td>
</tr>
<tr>
<td>Geoscience Australia/Corporate Information Management and Access</td>
<td>GA</td>
</tr>
<tr>
<td>Institute of Molecular Bioscience – The University of Queensland</td>
<td>IMB</td>
</tr>
<tr>
<td>Melbourne Health – Molecular Medicine Informatics Model</td>
<td>MMIM</td>
</tr>
<tr>
<td>Molecular and Materials Structure Network</td>
<td>MMSN</td>
</tr>
<tr>
<td>Monash University</td>
<td>Monash</td>
</tr>
<tr>
<td>National Library of Australia</td>
<td>NLA</td>
</tr>
<tr>
<td>Radio and Optical Astronomy Research Capability</td>
<td>ROAR</td>
</tr>
<tr>
<td>Reserve Bank of Australia</td>
<td>RBA</td>
</tr>
<tr>
<td>Securities Industry Research Centre of Asia-Pacific</td>
<td>SIRCA</td>
</tr>
<tr>
<td>Sesame Scientific Data Management System, Bio21 Molecular Science and Biotechnology Institute</td>
<td>Sesame</td>
</tr>
<tr>
<td>Tasmanian Partnership for Advanced Computing</td>
<td>TPAC</td>
</tr>
<tr>
<td>The Australian Social Science Data Archive</td>
<td>ASSDA</td>
</tr>
<tr>
<td>University of Melbourne</td>
<td>MU</td>
</tr>
<tr>
<td>Victorian Partnership for Advanced Computing</td>
<td>VPAC</td>
</tr>
</tbody>
</table>
Institutional model

Requirements of funded programs, e.g. ARC
Requirements of National Privacy Principles (NPP)
Rights and responsibilities of researchers and University

ARROW Access Management Working Group, e.g. for vocabulary for encoding access rights

Analytical data
Metadata

Treated data
Metadata

Published data
Metadata

Preserved data
Metadata

End Use
Other researchers, public, Government, etc

Academic research projects, units, collaborations

Collaborative agreements, e.g. with CSIRO

Institutional DM policies & Guidelines as part of the Monash University Information Management Strategy
Costing model for central storage space

Policies for retention and disposal
Data Federations: Australian National Data Service

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons
Data Federations: Australian National Data Service

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons
Data Federations: Australian National Data Service

- Develop user and owner frameworks for a data commons
- Develop and operate national registries and data sharing services
- Help Institutions connect repositories to the data commons
- Help researchers publish and share data through the data commons

The Frameworks Program
- is expected to be implemented through a small number of centres of expertise and will need to engage strongly with federal government agencies and national institutions

The Utilities Program
- is likely to define services to be outsourced to suitable suppliers, which may be suppliers of similar national services such as for the Australian Access Federation.

The Repositories Program
- will need to provide staff able to consult with and deliver on-site support to higher education institutions and significant research organizations, though the function may be hosted at a small number of institutions with existing expertise

The Researcher Practice Program
- is expected to strongly engage with research communities in activities mixing expertise between data management and research skills, and will require project implementation capabilities and consulting support in all regions, such as would be available through the Australian Research Collaboration Service
Where to now?

• Elections of 24 November resulted in a change of Government in Australia
• The Department of Education, Science and Training (DEST) is now the Department of Education, Employment and Workplace Relations
• NCRIS is joining the new Department of Innovation, Industry, Science and Research (DIISR) along with the Office of the Chief Scientist, CSIRO, ANSTO, AIMS (equivalent to national labs)
• The change presents opportunities:
  – A stronger connection with the broader research community outside of universities
  – The possibility of working more closely with CSIRO and other publicly-funded research agencies
  – An ongoing connection with PMSEIC
Where to now?

- The $6 billion Higher Education Endowment Fund (HEEF) remains in the new education related department
- The dividend from the investment in HEEF will support capital expenditure and research facilities in Australian universities
- HEEF does not replace NCRIS; it has a different scope and priorities
- Any future NCRIS funding would be expected to come from a review of the Innovation System
- Almost all of the elements required to respond to such a review are now consolidated in innovation related department
A closing thought - It’s all about data

• In consultations for NCRIS and particularly for Platforms for Collaboration, the overwhelming message from the research community was that data is just going to be more and more important

• Perhaps enduring data and advanced models are new forms of infrastructure that need better support
  – Does labelling data as infrastructure help?
  – What mechanisms apply if we treat it that way?
  – What is the economic model for ‘data’ infrastructure?
Thank you

For further information, contact:

Rhys Francis
Executive Director, Australian eResearch Infrastructure Council
Email: rhys@pfc.org.au
Website: www.pfc.org.au