

Research Data Management Consultation – Scoping Statement

Background

Any investment plan for Platforms for Collaboration for Research Data Management will be based on a long-term National vision for research data management. This vision would address policies and technologies around data access and discovery, storage and management¹ including:

- Policy/Standards/protocols development
- Technology/tools development underpinning services/processes/workflow
- Trusted repositories

The following national level activities have been undertaken recently or are under way in this area.

Activity	Contact	Partners
AERES project - Australian Partnership for Sustainable Repositories (APSR)	Adrian Burton and Markus Buchhorn	APAC, ANU, University of Melbourne, University of Queensland, National Library of Australia, University of Sydney, University of Technology Sydney.
Dataset Acquisition, Accessibility and Annotation e-Research Technology Project (DART)	Andrew Treloar	Monash University, University of Queensland, James Cook University.
Regional Universities Building Infrastructure Collaboratively (RUBRIC)	Alan Smith	University of Southern Queensland, University of Newcastle, University of New England, University of the Sunshine Coast, Massey University (NZ), Flinders University, Macquarie University, Murdoch University.
Australian Research Repositories Online to the World (ARROW)	Andrew Treloar	Monash University, National Library of Australia, UNSW, Swinburne University of Technology.
Middleware Action Plan and Strategy (MAPS)	Nick Tate and Patty McMillan	University of Queensland, Macquarie University, Monash University, ANU, CAUDIT, CAUL, AARNET, GrangeNet.
OAK Law Project	Professor Brian Fitzgerald	QUT, NICTA (Qld)

The purpose of information gathering on research data management within Platforms for Collaboration is to bring together the leading practitioners from these activities and other relevant informants to provide guidance on:

1. The state of national research data management policies and underpinning standards.
2. The state of frameworks and platforms that could enable national research data management.

Questions that need to be answered during the consultation process:

- What might constitute a long term National vision for research data management:
 - What are the problems/issues around research data management?
 - What are the core functions around data accessibility, short term sustainability and long term preservation?
 - What are the component services needed to support these functions?
 - Who might reasonably be seen as responsible for providing the component services within the overall functionality?
 - How might those services be provisioned?
 - How might they be accessed?

¹ Authentication and authorisation will be addressed by the AAA consultation under this capability.

- What will a 10 year plan around reviews of policy, functions, services and technologies look like?
- What are the possible or probable technologies and systems that could be deployed over the next 5 to 10 years and the level of functionality that could therefore be expected?
- What might be reasonable estimates of the time for the development of these policies and technologies?
- What activities that must be undertaken to ensure those timelines?

The consultation process will ensure that a comprehensive record of views of practitioners is available to the committee so that the level of agreement over issues in research data management can be ascertained.

Consultation Activity

Pre-workshop Consultation – September 11 to 26

The activity will consult widely to scope the data management problem via a mapping process.

Key questions will be framed around the types of systems/frameworks that need to be put in place to manage different types of data. Please refer to page 5 of this document for suggested questions for the consultation.

More specifically the consultation will attempt to identify the factors driving complexity in:

- Classes of research data
- Stages in the lifecycle of research data
- Research data rights ownership
- Possible systems/frameworks for managing research data
- Roles and responsibilities around these different systems/frameworks

Finally, a primary goal is to determine if an 80:20 rule might apply when we come to provision supporting infrastructure, i.e. which issues most act to increase the variety of systems needed for managing data?

Consultation will include:

- National Research Organisations, ARC, NHMRC
- Government/Private Research bodies, CSIRO, RIRDC (Rural Industries Research & Development Corp), DPI (State government departments of Primary Industry), Australian Bureau of Statistics (ABS), and Reserve Bank (RBA), others as identified during the activity.
- National Archiving Organisations/Groups, National Library (+PANDORA), National Archives, National Museum Archives, National Film and Sound Archive, Collections Council of Australia, others?
- National Universities committees, e.g. CAUL, CAUDIT, others?
- The Academies: Humanities, Social Science, Science, Technological Sciences and Engineering.

Internationally a range of reports and studies exist, specific contact will also be made with a few representative informants (using an amended question format):

- Brewster Kahle, Internet Archive Project, US
- Adil Hasan, CCLRC, UK
- Regan Moore
- Robert Jones, Enabling Grids for eScience (EGEE) Technical Director
- Local participants in international data management activities in physics, astronomy, geosciences and earth systems science
- Others as identified during the activity

Pre-workshop Consultation Output:

1. Data gathered from informants will be collated into a workshop discussion paper and distributed to workshop participants on **October 10.**
2. As it will be impractical for all to present at the workshop, participants will be asked to work in clusters based on background and interest and address key features identified in the Discussion Paper:
 - a. What are the classes of data and ownership rights around data and how does this impact on systems/frameworks and management problems?
 - b. What lifecycles for data are in common use and what if any changes to lifecycles for data are thought to be desirable?
 - c. What access regimes are in common use and what if any changes to access regimes are thought to be desirable?
 - d. Who has current ownership of resolving the problems?
 - e. What should happen in the future?
3. A presentation from each cluster on each issue will be invited at the workshop and some appropriate representation nominated to go forward into workshops 2 & 3.

Workshop Day One – Monday October 23

This will be a large workshop to ensure representation of the diverse research data management issues for researchers.

- To scope the research data management problem and to relate it to a vision of future “platforms for collaboration”
- To establish the elements of the problem over which there may be consensus.
- To establish what work could be performed over the next few months to further improve our understanding and planning, particularly covering non-capability areas (e.g., Humanities, Social Sciences)
- To gather and review expert views, to provide a foundation for advice to NCRIS

Workshop Structure:Part 1

- Report of consultation process and review of Discussion Paper content
- Reports from clusters on their activity and summary views
- Evaluation of the level of agreement on issues identified in the Discussion Paper

Part 2

- Some representative groups identified as likely to hold responsibilities in research data management in any proposed framework will be invited to present their view on requirements and responsibilities in data management. These groups will be identified from the pre-work and suitable speakers will be chosen to represent the groups.

Part 3

- Evaluation of the level of agreement on roles and responsibilities for data management
- Review of the day and forward planning

Key issues which will inform the above include consideration of:

- What do we have already – what have the AERES and PMSEIC found?

- What infrastructure is needed:
 - Storage Facing Issues: Underlying formats; Data signatures; managing technology migration; Persistent identifiers. Preservation policies: Data ‘ownership’; Cross-generation preservation planning; Policy regulation. Storage timelines – should these be set or should the infrastructure have capability to meet all parameters? Digitisation protocols. Researcher contribution of existing/new storage; international collaboration and interoperability, middleware.
 - User Facing Issues: metadata; standards; supporting AAA; Access mechanisms (portal, command line, API); middleware, and digitisation.
- Federated Access versus Federated Storage? Can researcher data held both inside and outside nationally managed storage be accessible via the same data management infrastructure (Federated Access)? Can researchers contribute existing storage and collections, and what policies are necessary for this to occur?
- Ownership and rights – how does this impact on policy frameworks, systems, roles and responsibilities?

Workshop Day One Participants

Proposed participants:

- Mike Sargent, NCRIS, Chair eResearch Co-ordinating Committee
- Rhys Francis, Facilitator, Platforms for Collaboration
- Linda O’Brien, Platforms for Collaboration Committee
- John O’Callaghan and Stephen McMahon, APAC
- NICTA
- CSIRO ICT
- Google Australia.
- Nick Tate, MAPS, University of Queensland
- Kim Finney, Working group on data for Science, PMSEIC
- Adrian Burton, APSR
- Markus Buchhorn (multi projects)
- Roze Frost, CSIRO
- Department of Primary Industry (DPI)
- CAUL executive representative
- CAUDIT executive representative
- AARNet representative– nominee from Director
- Andrew Treloar (Monash ARROW, DART)
- Warwick Cathro, NLA
- Professor Brian Fitzgerald, OAK Law Project, QUT
- Paul Davis, VerSi
- Deb Mitchell or Sophie Hollaway, Australian Social Science Data Archive, (ASSDA)
- Michael Briers, Securities Industry Research Centre of Asia-Pacific (SIRCA)
- The Australian Academy of the Humanities
- Academy of Social Science in Australia
- Vladimir Likic, Bioinformatics, Bio21 Molecular Science and Biotechnology Institute
- Geoff Taylor, HEP, University of Melbourne
- Gavan McCarthy, AUSTEHC
- Margaret Birtley, CEO, Collections Council of Australia Ltd
- Representatives from data gathering disciplines such as geosciences, astronomy and earth system sciences
- Facilitators from other NCRIS capabilities
- Other members of the PfC reference group and DEST representatives who wish to participate

Workshop Day Two – Monday November 27

This will be a smaller working group taken from Day One participants (including cluster representatives) to assist integrate the information gathered. Information from Day Two will be sent out to original informants for comment.

- To review summaries of the implications of issues identified in the process, particularly around classes of research data, classes of access rights and classes of ownership.
- To confirm the expected broad development and deployment timelines (if possible).
- To identify any investment actions that might be needed either immediately or at an appropriate time over the next 5 years.
- To review the appropriateness of data management infrastructure and data management activities within NCRIS investments in light of the broad community consultation.
- To review an outline of PfC recommendations on research data management

Workshop Day Two Participants

- Rhys Francis, Facilitator
- Other members of the PfC reference group and DEST representatives who wish to participate
- Selected participants from Day One Workshop

Workshop Day Three – short day Thursday February 15

Day Two participants will be re-assembled to review a final version of the PfC recommendations on research data management and determine if further comments or recommendations should be sent forward to other appropriate bodies.

Arrangements

The timing will be October 23rd, November 27th and February 15th and Melbourne has offered to host the meetings. Workshops 1 and 2 will be held at Melbourne Airport Hilton and the February workshop will be held at the University of Melbourne.

The PfC budget will cover PfC Reference Group travel, any facility hire and catering charges.

Participants from research institutions or DEST funded activities will be expected to cover the cost of their participation from their own project or institutional resources.

QUESTIONS FOR PRE-WORKSHOP CONSULTATION

1. If only a small set of data management systems can be economically implemented, what are the different classes or types of research data created/acquired/used in your research/discipline/organisation that must be supported?
2. If only a small set of different access restrictions can be economically supported, what different classes of access must be supported for your data and how would you describe, in the simplest way possible from your perspective, which can do what to your data?
3. Who are the owner(s) and who are the custodian(s) of the research data (indicate relationships if many)?
 - **Funder**, e.g. Corporation, Government, general public.
 - **Research Institution**, e.g. University, CSIRO.
 - **Research Collaboration**, e.g. ATLAS, AusVO, research project group.
 - **Researcher** – individual.
 - **Research subject/participant (human)** , e.g. patient/individual, indigenous community.
 - **Instrument owner**, e.g. telescope, satellite.
 - **Other?** Please describe the relationship.
4. How does ownership and access to data change across different stages of the data lifecycle? Who should have responsibility for data retention? For data curation?
5. What are the roles and responsibilities around the management of research data for your users/research communities? How does this vary at different stages of the research data lifecycle?
6. If you generate or store research data, what current volumes do you have and what growth do you anticipate over the next 3-5 years?
7. What other key problems exist around research data management for your users/research communities/users?
8. What are you doing or planning to do to address these problems?
 - 8.1 Policy problems
 - 8.2 Service Delivery problems
 - 8.3 Infrastructure problems
9. What solutions need to be local? National? International?
 - 9.1 Policy domain
 - 9.2 Service delivery domain
 - 9.3 Infrastructure domain
10. Any other comments