

Super Science – Research Data Storage Infrastructure Implementation Discussion Paper

Introduction

In May 2009 as part of its Super Science Initiative, the Australian Government announced funding of \$97 million to expand the electronic collaboration services and data fabric initiatives already underway in the Australian Research Collaboration Service (ARCS)¹.

Following the Budget announcements, and in response to consultations with the research sector, the Department of Innovation, Industry, Science and Research (the Department) will direct the \$97 million funding as follows:

- \$47 million to be allocated to:
 - providing core infrastructure and services that establish the national collaboration and data fabric managed by ARCS (\$35 million); and
 - furthering the aims originally set out in the National eResearch Architecture Taskforce (NeAT) process to create discipline specific infrastructure integrated with the national collaboration and data fabric (\$12 million); and
- \$50 million to be devoted to a data storage infrastructure to enhance regional data centre development and support retention and integration of nationally significant data assets into the national collaboration and data fabric.

The Department will enter into separate funding agreements for these two components. In part, this is to capitalise on likely institutional, and potential state government involvement and co-investment in the data storage initiative.

The Department is therefore seeking views from the sector to enable agreed implementation arrangements to be put in place for the data storage infrastructure.

Your Response

This discussion paper is focused on the \$50 million data storage component of the Super Science funding described above.

The Department requests written feedback from interested and relevant parties in response to the models proposed in this paper, with particular reference to the following:

- Three data storage infrastructure models have been proposed.
 - Which model would work best for the sector?
 - Are the proposed models feasible?
 - Are there other models that should be considered?
- Several questions have been posed for consideration in the development of criteria for assigning data storage capacity.
 - Are these questions focused on the appropriate points for determining access to data storage capacity?
 - Should additional, or different, points be considered?

¹ http://www.budget.gov.au/2009-10/content/bp2/html/bp2_expense-20.htm

- Are there any other sector developments not mentioned in the paper that may impact on the feasibility of the options proposed?

Responses are required no later than COB Friday 14 May 2010. Submissions may be made via the following methods:

Email

Please send written input via email to: e-research@innovation.gov.au.

Mail

Please send hardcopy input via mail to:

Clare McLaughlin

Manager, eResearch

Science and Infrastructure Division

Department of Innovation, Industry, Science and Research

GPO Box 9839

Canberra ACT 2601

Further Information

Further information on the consultation process is available online at www.innovation.gov.au/Section/science. Any questions or queries concerning the process should be directed to Ms Clare McLaughlin, Manager, eResearch, via email at e-research@innovation.gov.au or by phone on (02) 6213 6375.

Background

Vision

The \$50 million Super Science investment in data storage infrastructure is intended to protect and improve the availability of quality research data inputs for sharing and re-use and, as a result, to expand the scale and scope of problems that Australian researchers may seek to address.

The development of common data storage solutions across the country can also be expected to enhance system performance, improve research efficiency, reduce institutional data storage costs and enable more extensive collaboration.

The project provides the research sector with an opportunity to help frame a significant improvement in national research data retention and access.

Funding Context

The project will be implemented within a complex environment with competing discipline and institutional priorities and varying technological requirements.

Other projects funded as part of the Super Science Initiative include related investments in data management, high performance computing, research network infrastructure and a range of research infrastructures specific to discipline or capability requirements.

The data storage project will seek to realise economies of scale in a national data management and access system. While the data storage infrastructure may be composed of a combination of technologies from a number of suppliers, it is intended that all parts of it should be capable of being accessed through a common interface.

An effective and efficient use of the funding requires such an approach.

The resulting data environment should support specific identified data needs, but also be easily augmented by institutions, disciplines and research facilities, on a needs or interest basis, at an incremental cost.

Some of the drivers underpinning these objectives are as follows.

- At a very broad level, data intensive research and data re-use are important contributors to innovation excellence. In addition, improved data re-use also assists effective collaboration, providing a second contribution to innovation excellence and hence future prosperity.
- At a research sector level, the growing requirement to manage research data better and to more easily support collaboration is ubiquitous. However, for the first time, the acceleration in the generation of data is outstripping growth in data storage capacities. Therefore, over time, the research sector will need to devote more resources to data, taken as a whole, regardless of how funding is provided.
- Experience from the commercial sector shows that very significant economies of scale can be achieved, using highly sophisticated implementation strategies, while providing accessible data storage and data analysis capabilities. Large service providers, such as Google and Amazon, and the large providers of infrastructure such as Hitachi, Fujitsu and IBM, are developing new approaches to data infrastructure that the research sector might be able to adapt to its advantage.
- At the research discipline level, developments in data gathering and data management also trend towards the development and adoption of shared global standards and open access protocols, so that the re-use of data is becoming increasingly efficient and cost-effective over time.

- The vast bulk of the research data that will need to be managed in 2015 has yet to be generated. Putting in place the foundations to better manage data in the future is an urgent requirement with potentially significant benefits.

Strategic Roadmap for Australian Research Infrastructure

The Strategic Roadmap for Australian Research Infrastructure (the Roadmap) proposed extending the existing Australian Research Collaboration Service (ARCS) to provide a national data capability to complement institutional data investments.

Data Storage: to provide data storage facilities (the hardware component) and data management plans (the principles for storage and long term preservation) to assist research.

Building on the ARCS, a national data grid is needed to provide for long-term preservation of data. This national data grid would provide seamless unstructured storage as well as collaborative storage spaces, particularly focusing on but not limited to active research projects and research collaborations. A dedicated high performance network would link the nodes in the grid, allowing researchers to move data rapidly from instrumentation to computing resources and to institutional storage.

This investment would extend to research organisations for the development of institutional nodes of the storage grid, on the condition that the storage is used for research data; the institutes co-invest in the infrastructure; each institute agrees a data management plan; and each institute ensures its researchers use and abide by the data management plan.

Super Science: Research Data Storage and Collaboration Infrastructure (2009)

In the 2009 Budget the Government announced the Super Science Initiative, which included funding for elements of the eResearch proposals put forward in the Roadmap. This function was distributed between:

- The high bandwidth research network investment (\$37 million)
- The acceleration of the Australian National Data Service (\$48 million)
- The provision of \$97 million for:
 - the development of ‘regional’ data capabilities (\$50 million); and
 - collaboration and discipline specific infrastructure (\$47 million)

Further information on the policy background for the project can be found at Attachment 1.

Current Environment

In the period since the Roadmap and the Budget, a trend towards aggregation of precinct and regional data storage facilities has been evident, including in Queensland, New South Wales, Victoria and Western Australia. This trend is supported by many of the sector’s IT directors. There are significant economies of scale that can be obtained through shared services and aggregations that will significantly assist the sector to minimise the rapidly rising cost and operational complexity of large data management.

The sourcing of suitable computing power to co-locate with any significant data holdings is likely to require further investment. Therefore, the development of common centres reduces risks and costs in the future provision of such data analysis capabilities. The lower environmental impact of those more modern data centres has the potential to reduce the reportable environmental impact of each institution.

For some disciplines, such as the climate modelling community, it will be highly desirable that a data centre can be easily accessed from high performance computing sites.

The funding made available under the Super Science Initiative, from the Education Investment Fund (EIF), is for the creation and development of infrastructure, and is unable to be applied to associated

operating costs. While the \$50 million data storage investment may result in the deployment of more than 100 PB of storage by 2012/13, it is insufficient to support all data needs, including the costs of operating the data stores and the broader infrastructure. It will therefore be necessary to prioritise the use of this storage.

Therefore, the manner in which content will be selected for support, the biases this may create for and against existing data investments, and the way in which necessary operating costs will be met are all questions to be explored.

For example, how will the funding relate to investments in disciplines previously funded for data retention, such as marine and astronomy? In addition, a significant question exists around the manner in which large data holdings should be associated with national scale computing capabilities. This option would need to be contrasted with a more diverse supply solution/s for data management that may have greater prospects of leveraging significant co-investment.

Objectives

The Department's intention is to fund a data environment which may be implemented as a national network of a limited number of data stores, where the content held in those stores is allocated capacity through a priority and merit process and where that content can be readily accessed, analysed and re-used in a coherently governed environment. It is expected that all stores would be accessed through a common interface and be connected by a high bandwidth network. It is expected that the data stores would be provided by agencies within the sector or commercial providers, or both and that all the models outlined in this paper are likely to provide opportunities for the participation of the Australian ICT industry.

The intention of the project is to increase the sharing and re-use of research related data, which might be sourced from any sector, and the collaborative activity based on the sharing of data, building off an improved framework for data retention as needed. A national data environment of the scale planned will also enable new questions to be asked on topics and at scales not previously possible.

Policy objectives that can be progressed by this investment include:

- assisting the sector develop a shared collaboration environment in which institutions and researchers can easily manage many separate collaborations, thereby materially reducing the cost of collaboration for all participants; and
- avoiding the costs associated with multiple divergent approaches to data management and data sharing arising across the sector.

In addition, it *may* be possible to develop sufficient economies of scale through this initiative that allow some of the storage capacity to be assigned to support universities and other research organisations in the following ways:

- assisting institutions sustain a quality of research in the digital age that includes the reproducibility of results; and
- assisting institutions comply with the research data provisions of Universities Australia's Australian Code for the Responsible Conduct of Research.

These objectives suggest the following general principles:

- National data stores should be managed within a national research data environment;
- National data stores should be able to hold data from multiple organisations, sourced from Australia and overseas;
- The primary purpose should be to hold data collections which have national applicability and which already have established national governance arrangements to control upload, access and re-use policies;

- The data held should be data intended for use by multiple researchers (both individuals and entities) and should be accessible for the purposes of research, subject to appropriate access rules;
- The quality of service should be able to span from temporary caching to archival retention. However, as the cost of storage capacity depends entirely on the proportion of each type of storage that will be available and the efficiency with which content is migrated between these storage types or tiers, further scoping work must be undertaken before implementation; and
- Access controls should support content providers in managing their own access rules.

More detailed requirements might include:

- The deployment of Australian Access Federation (AAF) based authorisation and tools to enable the governing bodies of each data set or collection to continue to manage access to their resources, researchers to identify their data wherever it is located, and research workflows to be implemented that inter-connect instruments, analysis tools and the data environment;
- The provisioning of an underpinning high speed network capacity that scales with growth in the data environment; and
- Providing an ability to attach access servers and analysis resources directly to the data stores. These resources may be owned and operated by different organisations than the operator of a particular data store.

Managing major data holdings is a significant undertaking requiring specialised expertise, and because significant economies of scale exist, it is possible that the early stages of the investment may need to focus on regional groups of institutions positioned to proceed in the short term.

The most likely early development sites might be where:

- a group of institutions is already separately managing significant collections and foresees a growing need to do so;
- such collections are meritorious within the context of the purposes set out above;
- the new services will drive a scaling up of the research supported by the collection/s in question; and
- the relevant institutions are prepared to co-operate into the future as a party to the initiative and enter into the necessary service arrangements.

Specific planning questions will need to be resolved as part of the development work, including agreements around the ongoing retention of data, for instance who can do what to data once it is lodged, and the deletion rights that should persist over that data.

Implementation Parameters

The factors above will be developed into appropriate requirements that can be used to contractually direct the implementation of the \$50 million investment.

The following broad requirements have been developed as a starting point.

Essential requirements

- A. The establishment of a sector based governance, management and implementation mechanism appropriate to the growing importance of research data retention across the sector that is capable of addressing longer term issues, beyond the life of this funding.
- B. The establishment of a process that has sector support, to identify data sets and collections that will be inputs to future research activities, and to focus and apportion resource allocation using the

National Research Priorities and national research infrastructure priorities (as identified in the Strategic Roadmap for Australian Research Infrastructure²).

Issues to consider in the development of criteria are:

- What data will be re-used by the research community?
 - What data sets make up the inputs to research?
 - Where are the relevant data sets sourced from and how?
 - For what period and with what access rights should data be retained?
 - What happens at the end of the retention period?
- C. The provision of data management facilities able to support the collaborative use of identified national research data sets and collections. It will be important to target these new systems at communities that can demonstrate a need to dramatically increase their data needs - by an order of magnitude
- D. The development and deployment of an access management and upload/download network that enables the effective use of, and collaboration around, the data sets and collections managed within the facilities. This would include support for a variety of access controls which can be exercised by contributing parties, for the data they contribute.
- E. The development of a scoping document describing the types of storage required and the management techniques to be employed.

Additional requirements

A key question in setting out the requirements concerns the degree to which resources should be directed to any of a number of additional requirements, including the following:

- The provision of basic analysis facilities that can support common data processing and analysis needs of researchers using data sets and collections managed within the facilities.
- The provision of robust archive and retention capability for selected data sets, in addition to the collaboration, re-use, and access management capabilities listed under 'essential requirements'. This requirement may extend to long-term active preservation and the appropriate archiving of technologies associated with the retrieval of data from collections of particular national importance.

Implementation Models

Broadly speaking three factors are likely to influence the implementation of this project.

Firstly, the nature and growth of operating costs for large data centres needs to be better understood, as is the means by which this funding support can be provided. This issue is particularly important in the context of EIF funds and as cost factors such as the price of power may undergo significant change in the near term.

Secondly, the nature of required analysis facilities and the relationship between data volume and associated analysis capacity will need to be developed over time. Hence the manner in which those facilities should be associated with the national data environment will likely evolve, so that an implementation model needs to provide that flexibility.

Thirdly, the presumption of ownership of data present in the sector is likely to influence the rate and manner in which data services will be taken up. The model needs to be able to respond to this issue.

²Department of Innovation, Industry, Science and Research: Strategic Roadmap for Australian Research Infrastructure, http://ncris.innovation.gov.au/Documents/2008_Roadmap.pdf

Model 1 – New Peak National Capability

In this model a ‘new’ national facility is created to be the Australian peak research data service.

- The envisaged service could be based (in a minimal configuration) around two physical sites supporting a single fully replicated data service open to all researchers.
- Because data is held remotely, some form of operating cost contribution would be required from data contributors, subscribers or sector participants.
- The cost and funding model would need sector agreement and some risk would be present for the operators. However, the cost for the volume of data envisaged would be significantly less than any possible in-house solution because of the substantive EIF funds and because of the large economy of scale factors.
- The location of the service and the operators could be decided using an Expression of Interest (EOI) process.
- A significant national developmental opportunity might arise if the new facilities included ‘cloud’ data analysis capabilities and the investment also supported data caches at relevant major computing and supercomputing centres. The result could form into a national compute/data network with four or five major facilities and a national data/analysis backbone in which nodes could specialise around different disciplines over time.

It is important that an inclusive, sector based data-centric governance body is established, and a competitive process by definition would isolate parties into separate bids. Therefore, the use of an EOI process would raise some issues around governance, as it is highly likely that a final governance arrangement, set up to mirror sector interests, would need to include a broader range of stakeholders than respondents to the EOI.

Model 2 – Regional Strength

In this model, regionally focussed services would be developed on the basis of existing regional associations. A particular advantage could be gained from building on associations in which state governments have an interest, as this may assist state government agreement to co-locate copies of state generated research related data.

A key benefit of this model would be the leveraging of local co-investment. Economies of scale would still be delivered given that a small set of shared data stores would result. This option would provide better opportunities for state and territory government data ingest.

- A common national data-centric governance, and associated standardisation and service inter-operation platform would be required. Those structural arrangements could not be easily established through an open EOI process, so that significant government involvement in establishing the nodes is likely to be needed.
- A regionally distributed data service maximises the potential for ingest (including physical data transfer options), improves access and data upload/download capacity, and has the potential to enhance the perception of relevance to each set of local Universities. Participants are likely to view locality as important when considering independently adding capabilities to such centres, such as co-locating data analysis systems.
- Regional location may also allow operating support to be provided on an in-kind basis or through locally negotiated mutual resourcing and an EOI process could be applied region by region to identify operators and co-funders. However, the ability for data to migrate over time to build discipline centric collections could represent a challenge. This approach may lead to significant complexity as a “slice” of each discipline specific data collection may need to be delivered across the federation of regional centres.
- The funding support for the operating components of the necessary layer of inter-operation services and overall co-ordination activities would need to be developed and agreed, but some fund flow from the regions to the centre for overall management would be necessary.

- The provision of capacity for data movement/replication/caching to bring data from any regional location to differently located analysis capabilities would be a major design question to be resolved.

The specification at the outset could include either a single ‘national’ replicate site (copy of all data) with ‘cloud’ data analysis capability to ensure the purposes are fully achieved or the use of mirrored pairs of sites to achieve a similar outcome. The logic might be that the Commonwealth is supporting regional strength, ensuring robust retention and providing national data analysis capability, noting that participants could provide further regional specialised data analysis capacity, as they desire. Such an extension may be difficult as the operational funding for the ‘national’ framework has no obvious contributor.

Model 3 – Industry Partnerships

Models 1 (a new peak capability) and 2 (building regional strength) imply that the data stores are likely to be located with stakeholder institutions as part of their data infrastructure and potentially managed by them while being made accessible both regionally and nationally.

An alternative approach would be for the sector to work with commercial suppliers to build all or part of the required infrastructure, whilst retaining the provision of an appropriate interface layer within the sector. This approach could contribute to either the new peak national capability model, or the regional store model, described as Model 2.

Some advantages of this approach include:

- A purely commercial arrangement would assist the development of a robust service level agreement for the infrastructure in data stores
- Cost effective 24 hour a day/7 days a week support could be obtained by leveraging the economies of scale of a supplier already providing significant data management capacity
- Cost effective pricing for additional capacity for individual institutions in the sector, for any purpose, could be achieved by leveraging a large initial procurement
- Further economies of scale might be generated if state or territory governments wished to make use of regionally situated and commercially supported national data stores.

It is difficult, at this stage, to predict the pricing of this option but it could be considered a useful reference against which to measure other expressions of interest. A significant amount of work would be required to address the viability and trust issues associated with this model that could impact the implementation of the service.

Other Models

A variety of more highly distributed campus or research centre based models could be envisaged.

However, any cost model of data stores and their management systems shows that very significant overheads arise as the number of data stores grows, especially in the case of new or additional data stores - given that each store attracts a fixed cost in base equipment, staffing levels and high throughput network access services.

Conversely, the economies of scale that can be achieved with very large stores grow with the sophistication of the implementation techniques that only become practical at scale, so that there will be a multiplying gain in the volume of data that will be able to be managed for the same cost, as the number of centres is reduced.

Just as importantly, research projects, research centres and indeed individual institutions are not the only source of national data sets and data collections or the sole users of national data sets and national data collections, even if they do currently host their retention. Fundamentally, the model of national data sets is broader and requires broader governance and delivery solutions that are more related to research techniques, research fields or disciplines.

For these reasons, more highly distributed models do not satisfy the requirement for an effective, efficient use of public funds and have not been elaborated further.

Future Planning Activity and Implementation Timeline

An initial examination of the proposed models suggests that a regional but integrated solution offers the greatest opportunity for local enhancement and support, for contribution of copies of state data assets, and for the co-location with vital data intensive resources, such as supercomputers, synchrotrons and already established major data holdings. These benefits may be more important than the clear national impact of the first model (a new peak national capability).

The attractiveness of the last model (industry partnerships) should also be explored with the sector.

To assist the Department to understand these trade-offs and thus to refine an implementation strategy, further inputs will be gathered from the sector.

Consultation Process

Research communities will be engaged through appropriate bodies such as National Collaborative Research Infrastructure Strategy (NCRIS) capabilities, Australian Research Council (ARC) Research Networks and the Learned National Academies, to identify:

- the (current and future) data sets and collections that meet the purposes of the funding and which are significant to disciplines
- any community based preference around possible providers of related data services

Research institutions, including universities and large and small research institutions will be approached to provide feedback on:

- the three models outlined above and their various expected impacts that might arise, both positive and negative, on institutions, disciplines and the sector overall
- a preferred model that should be further developed into an implementation plan and the essential characteristics that such an implementation must meet
- the nature of the governance mechanism that should be put in place, and
- criteria to determine which data sets are to be supported.

The results of this consultation and the proposed steps forward will be reviewed by the Department and canvassed through a stakeholder meeting arranged specifically for that purpose.

Implementation Plan development

Following consultation, the Department will work with the sector to develop an Implementation Plan that executes the preferred model and supports the prioritised data collections noting that not everything can be done at once. There are varying levels of funding, experience, agreement and skills across the sector and between states and regions. Consequently, an *initial* focus on those communities that are likely to benefit from early engagement, will be appropriate and necessary.

The implementation planning process will include the following steps:

- The development of a governance model and a description of the infrastructure.
- Agreement on appropriate service agreements for both the national data environment and the underlying data stores. Those agreements will need to reflect both community expectations and the community's ability to fund them, which may be different for different stakeholders.
- Determination of a funding model acceptable to the sector that would allow data to migrate and be aggregated on a discipline basis over time.
- Development of a suitable process to identify institutions and regional partnerships willing and able to engage as a provider in the proposed service, and/or commercial suppliers both capable and willing to engage in a partnership approach.

- Establishment of the data storage capacity allocation process
 - Given the complexity in this investment, and that an acceptable valuation process for data has yet to emerge, co-investment, or provision of appropriate facilities may need to be included in the mechanism for determining priority.
 - A large variety of data management layers and interfaces needs to be supported as disciplines will naturally continue to develop specific data management solutions and data sharing protocols and analysis tools, to meet their particular needs. The readiness of different disciplines to engage a national data environment is likely to be a further factor in the capacity allocation process.
 - The final architecture will also need to account for substantial data sources, such as existing large data holdings, High Performance Computing (HPC) systems, the synchrotron, telescopes and emerging bioscience facilities. However, the peculiarities of the interfaces into these systems may not be economic to provide via the national data storage infrastructure. Consequently, the ability of large data intensive investments to adapt their own infrastructure may also be a factor in the capacity allocation process.

The intention is to progress agreement through the above steps to the extent that a suitable funding agreement can be put in place to implement the further detailed design process, installation and operation of the national data environment described in this paper.

Attachment 1 – Policy Context

eResearch context

The decision by Government to support the establishment of national research data stores follows a continuous period of investigation and reporting, both in Australia and overseas, in support of such action.

The development of the National Collaborative Research Infrastructure Strategy (NCRIS) roadmap and the consultation processes leading to the establishment of the Australian National Data Service (ANDS) confirmed those requirements and the importance of data to contemporary research.

A specific and substantive summary report was prepared for the Prime Minister's Science, Engineering and Innovation Council (PMSEIC) by an expert panel in 2006. The report, entitled *From Data To Wisdom: Pathways to Successful Data Management for Australian Science* is available at www.pfc.org.au/pub/Main/Documents/PMSEIC_Data_to_Wisdom.pdf.

Much of that report relates to policy and data description and access requirements for research use of data and specific recommendations were made related to a national network of federated digital repositories. These were:

Recommendation 2:

That a high-level expert committee be established to provide the leadership role in progressing the formation of the long-term strategic framework for scientific data management.

Recommendation 3:

That the necessary policy and programmes be implemented with a view to establishing a sustainable publicly funded national network of federated digital repositories.

Recommendation 4:

That the expert committee consider the development of a strategic roadmap for the implementation and evolution of the national network of federated digital repositories.

The PMSEIC report lists a variety of international sources supporting its recommendations.

Recent international examples continue to reinforce this perspective. For example, the US National Science Foundation (NSF)'s Cyberinfrastructure Council report *Cyberinfrastructure Vision For 21st Century Discovery* defines a vision in which

...science and engineering digital data are routinely deposited in well-documented form, are regularly and easily consulted and analysed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved, NSF's five year goal is twofold:

- To catalyse the development of a system of science and engineering data collections that is open, extensible and evolvable; and
- To support development of a new generation of tools and services facilitating data mining

The resulting national digital data framework will be an integral component in the national cyberinfrastructure framework.

As the NSF wording highlights, the overall goal for data in research is ambitious and there is much to be done in finding a path to that future.

- The most recent review of the National Innovation System, *Venturous Australia*, also recommended a new infrastructure:
- Recommendation 7.10: A specific strategy for ensuring the scientific knowledge produced in Australia is placed in machine searchable repositories be developed and implemented using public funding agencies and universities as drivers.
- Recommendation 7.11: Action should be taken to establish an agreed framework for the designation, funding models, and access frameworks for key collections in recognition of the national and international significance of many State and Territory collections (similar to the frameworks and accords developed around Australia's Major Performing Arts Companies).
- Recommendation 7.14: To the maximum extent practicable, information, research and content funded by Australian governments – including national collections – should be made freely available over the internet as part of the global public commons. This should be done whilst the Australian Government encourages other countries to reciprocate by making their own contributions to the global digital public commons.

Australian Industry Participation

The Australian Industry Participation (AIP) framework is aimed at increasing opportunities for capable and competitive Australian and New Zealand small and medium enterprises (SMEs) to be suppliers for government funded projects. Compliance with the framework is required for major direct Commonwealth Government procurements and encouraged for projects funded, but not directly procured, by the Commonwealth. More information is available at www.aip.gov.au.