



Australian Government

# NCRIS

National Research  
Infrastructure for Australia

## The Australian Research Data Infrastructure Strategy

**Research Data  
Infrastructure Committee**

The Data Revolution: Seizing the Opportunity



ISBN

978-1-74361-972-8 [PDF]



With the exception of the Commonwealth Coat of Arms, the department's logo, any material protected by a trade mark and where otherwise noted, all material presented in this document is provided under a [Creative Commons Attribution 3.0 Australia](http://creativecommons.org/licenses/by/3.0/au/) (http://creativecommons.org/licenses/by/3.0/au/) licence.

The details of the relevant licence conditions are available on the Creative Commons website (accessible using the links provided) as is the full legal code for the [CC BY 3.0 AU licence](http://creativecommons.org/licenses/by/3.0/au/legalcode) (http://creativecommons.org/licenses/by/3.0/au/legalcode).

The document must be attributed as *The Australian Research Data Infrastructure Strategy*.

# The Australian Research Data Infrastructure Strategy

---

## Preface

Globally, researchers are confronted with a veritable tsunami of data as a result of new means of gathering, collecting and generating data in increasingly complex forms. Data-intensive research is a critical aspect of cross-disciplinary research directed at major challenges. At a 2013 forum at the World Statistics Congress in Hong Kong on open data and reproducible research, three key requirements were identified as the backbone of collaborative research: the existence of appropriate infrastructure, a collaborative and open culture for research data, and researchers and data experts with the necessary skills.

Australian researchers have been well served by the significant national investment in infrastructure to support research. In Australia, operators and users of existing research infrastructure are seizing new opportunities presented by the data made available by these facilities, and the appreciation of the value of data underpins future research infrastructure development. For example, the Pawsey Centre supercomputer was built to deal with the masses of data arising from the largest astronomical project ever undertaken—the Square Kilometre Array—and cloud computing and virtual laboratories are responding to the data-intensive, multidisciplinary nature of environmental research.

In August 2012, recognising the importance of effective infrastructure to support research data, the then Department of Industry, Innovation, Science, Research and Tertiary Education established a Research Data Infrastructure Committee (RDIC) to review the current national research data landscape. The purpose of this committee was to provide advice to the Australian Government about optimising existing and future investments in research data infrastructure. Specifically, the department asked the committee to provide advice on high-level, strategic issues relating to data management; the roles of government, institutions and research facilities; and how to manage these roles effectively to encourage a collaborative, sustainable approach to research data and research data infrastructure in Australia.

In the course of its deliberations, the committee developed the Australian Research Data Infrastructure Strategy to advise the Australian Government on the current and future roles of research data infrastructure to support data-intensive research. The strategy considers the critical role of infrastructure in improving access to, and use of, data by researchers and others. It examines how research infrastructure supports a viable data-rich environment, what efforts need to be continued, sustained or initiated and makes specific recommendations about how this might be brought about.

Research data can be valuable for purposes other than research, including in operational or policy environments. For instance, some data collected in marine research is critically important to the operation of the Australian Bureau of Meteorology. Equally, data collected by government and other sectors, such as that collected by human services delivery agencies, can be crucial for research.

A key aspect of future infrastructure development will be how different pieces of machinery, equipment, tools, facilities and resources can be coordinated most effectively to improve outcomes for researchers, government and industry; increase productivity; and address societal challenges. This coordination depends on technical issues, including interoperability. More challengingly, it

depends on the capacity and ability of people to work across boundaries, disciplines, sectors and institutions to maximise the utility of new infrastructure. It also relies on sound approaches to investment planning across multiple sectors, and on sustained institutional commitment.

A connected network of research data infrastructure facilities supports cross-disciplinary research by addressing broader societal challenges. The benefits are immeasurable, and answering new research questions that arise from such challenges, in entirely new ways, could significantly increase Australia's national productivity.

Systematic, effective investment in research data infrastructure must be accompanied by an environment that enhances the use of data, such as open access policies and provisions. At the same time, existing principles for investment in research data infrastructure, such as those noted in the *2011 Roadmap for Australian Research Infrastructure*, will continue to apply. This coherent policy and infrastructure environment targets both an increased return on research investment through data usage, as well as enhanced research integrity through the reproducibility of research outcomes.

The committee emphasised the need for effective approaches to existing and future investments in data-intensive research infrastructure and has proposed a foundation for how governments, industry, and the research sector can work together productively to better develop research data infrastructure and enhance the benefits from data in years to come.

As Chair of the Research Data Infrastructure Committee, I would like to thank the members of the committee for their hard work over a number of months. In particular, I would like to thank the subcommittee members who worked on aspects of the strategy, and the members of the committee secretariat who worked indefatigably and with consummate professionalism throughout the committee's term. I would also like to thank the commentators listed in the acknowledgments for generously giving their time and their insights.

Dr Ron Sandland  
Chair, Research Data Infrastructure Committee

## Contents

Contents.....	iii
List of breakout boxes.....	iv
Acknowledgments.....	v
Acronyms .....	vi
Executive summary .....	vii
1. Vision.....	1
2. Scope.....	2
3. Context.....	4
4. Principles.....	14
5. Needs .....	18
6. Challenges and opportunities .....	22
7. Recommendations .....	28
Appendix A: General references .....	33
Appendix B: Addendum on open data.....	36
Appendix C: Principles from the Strategic Framework for Research Infrastructure Investment .....	40
Appendix D: Recommendation 6—Proposed draft terms of reference .....	44
Appendix E: Glossary.....	46
Appendix F: Research Data Infrastructure Committee membership, role and terms of reference .....	49

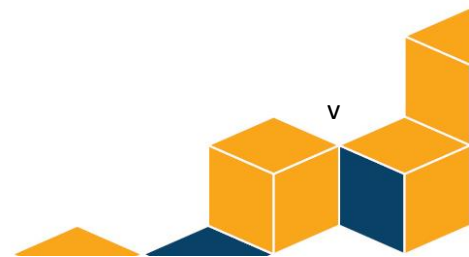
## List of breakout boxes

<i>Box 1: Research data infrastructure investments</i> .....	3
<i>Box 2: Data enables weather and climate modelling</i> .....	4
<i>Box 3: A virtual research vessel fleet</i> .....	5
<i>Box 4: Mapping Australia's soil diversity</i> .....	7
<i>Box 5: Linking Australia's stories with HuNI</i> .....	8
<i>Box 6: Virtual observatory standards enable advances in astronomical research</i> .....	9
<i>Box 7: Enhancing the usability of census data for urban research</i> .....	12
<i>Box 8: International collaboration and coordination in world climate research</i> .....	13
<i>Box 9: Tagged seals help solve 30-year mystery, serendipitously</i> .....	16
<i>Box 10: SURE: Demonstrating the importance of linked health data to population health research</i> .....	21
<i>Box 11: Challenges and opportunities for astronomy data collection on the Antarctic Plateau</i> .....	22
<i>Box 12: Terra Nova: Establishing a collaboration advantage for Australian researchers and policy makers</i> .....	30

## Acknowledgments

The Research Data Infrastructure Committee and the Department of Education thank the following people and organisations for their contribution, participation and feedback in the development of the Australian Research Data Infrastructure Strategy:

ARC Centre of Excellence for Climate System Science  
Australian Bureau of Statistics (ABS)  
Australian Government Chief Information Officer, Mr Glenn Archer  
Australian Government Information Management Office (AGIMO)  
Australian Institute of Marine Science (AIMS) Chief Executive Office, Mr John Gunn  
Australian Research Council (ARC)  
Australian Urban Research Infrastructure Network (Director, Professor Robert Stimson)  
Bureau of Meteorology (BoM)  
Chief Scientist for Australia, Professor Ian Chubb, and the Office of the Chief Scientist  
Climate Change Adaptation and Science Division, former Department of Industry, Innovation,  
Climate Change, Science, Research and Tertiary Education  
Commonwealth Scientific and Industrial Research Organisation (CSIRO)  
Department of Human Services (DHS)  
National Computational Infrastructure (Director, Professor Lindsay Botten)  
National Health and Medical Research Council (NHMRC)  
NSW Chief Scientist and Engineer, Professor Mary O'Kane  
NSW Deputy Vice-Chancellors (Research)



## Acronyms

2011 Roadmap	<i>2011 Strategic Roadmap for Australian Research Infrastructure</i>
AABW	Antarctic Bottom Water
ABS	Australian Bureau of Statistics
ACCESS	Australian Community Climate and Earth-System Simulator
ALA	Atlas of Living Australia
ANDS	Australian National Data Service
ARC	Australian Research Council
AREN	Australian Research and Education Network
AURIN	Australian Urban Research Infrastructure Network
BASE	Biome of Australian Soil Environments
BoM	Bureau of Meteorology
CMIP5	Coupled Model Intercomparison Project Phase 5
ESGF	Earth System Grid Federation
GBIF	Global Biodiversity Information Facility
HPC	High-performance computing
HuNI	Humanities Networked Infrastructure
ICT	Information and communications technology
IMOS	Integrated Marine Observing System
IVOA	International Virtual Observatory Alliance
IWGDD	Interagency Working Group on Digital Data (USA)
NBN	National Broadband Network
NCI	National Computational Infrastructure
NCRIS	National Collaborative Research Infrastructure Strategy
NeCTAR	National eResearch Collaboration Tools and Resources
PHRN	Population Health Research Network
QCIF	Queensland Cyber Infrastructure Foundation
RDIC	Research Data Infrastructure Committee
RDSI	Research Data Storage Infrastructure
SKA	Square Kilometre Array
SURE	Secure Unified Research Environment
TERN	Terrestrial Ecosystem Research Network



## Executive summary

Data is central to all research. Data, in its raw or processed form, from its original source (such as an ocean sensor) or via an analytical processor (such as the cores of a supercomputer), depends invariably on research infrastructure for its collection, generation, manipulation, characterisation, use and dissemination. Research data infrastructure refers to a range of facilities, equipment or tools that serve research through data generation, manipulation, curation, and access. It includes data itself.

The Australian Government has made significant investments in research data infrastructure, guided by principles set out in existing strategies. In light of newly developed sets of principles—in particular the Strategic Framework for Research Infrastructure Investment principles, which appear in the *2011 Strategic Roadmap for Australian Research Infrastructure*—the Government established the RDIC. The committee reviewed the national research data landscape to provide advice on how to optimise existing and future investments in research data infrastructure.

Developed as the Australian Research Data Infrastructure Strategy, this advice provides a basis for policy makers, investors, developers, operators and users to build and sustain an effective and holistic Australian research data infrastructure system. It is a system that *collects* data systematically and intentionally, *organises* data to make it more valuable, and *uses* data insightfully many times over.

The strategy proposes three key requirements for a successful national research data infrastructure framework:

- sustained infrastructure to support priority research data collections, data generation and management
- appropriate data governance and access arrangements
- delivery of enhanced research outcomes from effective data infrastructure arrangements.

Drawing on these requirements, a framework for an Australian research data infrastructure is proposed in Figure 1.





Figure 1: Australian research data infrastructure strategic framework

The strategy frames the way forward in 18 specific recommendations, which are aimed at all stakeholders with an interest in implementing them. They are the proposed means by which future investments in research data infrastructure can be made and used more productively.

A key recommendation is to establish a national research data infrastructure advisory committee (Recommendation 6). One of the committee's first tasks will be to prioritise the recommendations for implementation. The timescale for implementation will depend on availability of funding, capability, capacity and preparedness of all stakeholders, as well as perceived urgency and degree of complexity. The effectiveness of the recommendations and their successful implementation is dependent on the support and participation of key stakeholders across all levels of government; the research sector; industry; and other funders, operators and users of research data infrastructure.

The recommendations are:

**Recommendation 1:** Recognise the ongoing value of Australia's research data assets, and safeguard prior and future investment in them by developing a sustained approach to investment in research data infrastructure.

**Recommendation 2:** Adopt a rolling decadal planning and prioritisation approach to investment in research data and research data infrastructure. Planning and prioritisation should reference the data infrastructures described in the *2011 Strategic Roadmap for Australian Research Infrastructure*.

**Recommendation 3:** Require best practice in discoverability, accessibility and usability of Australian research data from all future investments in research data infrastructure, in line with leading data management policies and procedures. Set minimum data management requirements for nationally funded research facilities, to be incorporated into funding guidelines and agreements.

**Recommendation 4:** Design national research data infrastructure investments to enhance and complement capability at the institutional and international level, encourage institutional commitment, and foster sustained institutional capacity in data-intensive research.

**Recommendation 5:** Coordinate research data infrastructure investments to drive collaboration between, and commitment from, research institutions, government agencies, industry participants and community stakeholders.

**Recommendation 6:** Establish a national research data infrastructure advisory committee to review, coordinate and provide coherence to the implementation of research data infrastructure investments, including through assisting national research data infrastructure facilities to generate and build partnerships and collaborations. The committee will have particular regard to the recommendations in this strategy in delivering its terms of reference.<sup>1</sup>

**Recommendation 7:** With reference to Recommendations 1 to 5 above, require funded capability areas under the *2011 Strategic Roadmap for Australian Research Infrastructure* to include research data infrastructure plans consistent with this strategy. These plans should include national data strategies to coordinate collection, generation, manipulation and management of research data across institutions, programmes, sectors, and funding sources so as to maximise access, use and re-use.

**Recommendation 8:** Encourage and support custodians of government data collections to make sure that data is available, accessible and useable for research purposes, consistent with the principles of open government.

**Recommendation 9:** Ensure that research data infrastructure facilities are configured to enable appropriate access for data outputs for publicly funded research, consistent with the growing national and international policy momentum toward open access to research data.

**Recommendation 10:** Enable Australian researchers to access national and international research data, and data for research, through effective development and use of research infrastructure consistent with common and standard data approaches.

**Recommendation 11:** Encourage the adoption of appropriate national services that ensure open access to data for a broad spectrum of researchers, and require custodians of publicly funded research data to make that data available, accessible and useable.

**Recommendation 12:** Plan investments in research data tools strategically, with due consideration to existing investments, so that they are non-duplicative, and at a scale that is nationally and internationally relevant and can be supported.

**Recommendation 13:** Position research data infrastructure investments globally so as to provide Australian researchers with an environment in which they are seen as partners of choice in data-intensive research internationally.

---

<sup>1</sup> Draft terms of reference are in Appendix D.

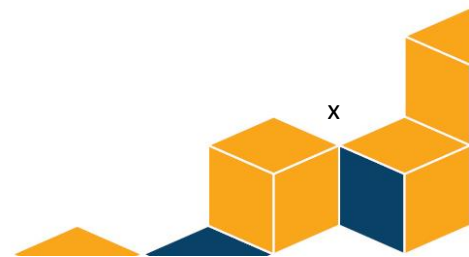
**Recommendation 14:** Ensure data services are flexible and adaptable, using standardised and open architectures, to help researchers explore research data assets effectively.

**Recommendation 15:** Urgently build the digital literacy of the research community through increased emphasis on use of research data infrastructure in research training, and links with higher education programmes to create a level of data competency in future generations of Australian scientists, researchers and academics that is globally competitive.

**Recommendation 16:** Provide incentives for researchers to contribute to and make use of effective research data infrastructure, by linking adoption to increased citations, emerging data impact metrics such as data citation, greater career recognition and further access to institutional support.

**Recommendation 17:** Build and maintain the human capacity— including that of support staff, technical staff and data specialists—to realise optimal benefit from investments in research data infrastructure. The competencies and skills required, the role of communities of practice, and career development paths should be emphasised. These efforts should occur in conjunction with and within the parameters of existing government activity to support research workforces.

**Recommendation 18:** Maintain and develop a coherent national infrastructure to aggregate and richly integrate information across domains, sectors and facilities, thus enabling researchers to navigate, explore and exploit Australia's evolving data assets.



## 1. Vision

Excellent Australian research data infrastructure fostering globally significant research

## 2. Scope

### Definition

Research data infrastructure refers to a range of facilities, equipment or tools that serve research through data generation, manipulation, and access. Research data infrastructure includes data itself, and relies on a skilled technical and research workforce for establishment, implementation, operation and use.

These facilities, equipment and tools include *eResearch infrastructure* as well as *data collecting and generating infrastructure* that encompasses large or systemic research infrastructure installations—such as high-performance computers, telescopes and marine observation systems, among others (see Box 1).

### Scope of the strategy

Research data infrastructure as defined above is the primary concern of this strategy. Within these parameters, the strategy has regard to:

- describing how research data infrastructure can be provided and used into the future in order to best support researchers and other stakeholders
- maximising the availability and delivery of, and the connections between, data generated and collected by research data infrastructure to researchers and others
- describing the technical, cultural and institutional environments most beneficial to the productive and effective use of such infrastructure
- suggesting the pathways by which data generated from outside the research sector, including within government, can also be made available to researchers more effectively.

The scope of this strategy extends to:

- describing how data is exchanged and used in an environment where the researcher relies not only on data created and collected in their sector, discipline or institution, but also on data produced elsewhere, by other disciplines or in government and private sectors
- underpinning initiatives to make publicly funded data more available to a wider audience, such as government users, non-government users and the private sector
- guiding, rather than determining, future investments in research data infrastructure in view of the relevance of other government strategies and policies that may influence outcomes.

The audience for this strategy is:

- policy makers in the Australian Government
- funders of research infrastructure
- funders of research data
- funders of other data (for example, government)
- custodians and owners of data, including government agencies
- research institutions including universities
- implementers of research data infrastructure
- users of research data infrastructure
- other sectors (for example, industry).

## Box 1: Research data infrastructure investments

### *eResearch infrastructure*

National-scale investments in eResearch infrastructure, under the National Collaborative Research Infrastructure Strategy (NCRIS) and the Super Science Initiative, have developed a robust national capability to make data available and useable across the research sector.

The work of the Australian National Data Service (ANDS), including its development of an Australian Research Data Commons, is transforming research data from being unmanaged, disconnected, invisible and single-use; to managed, connected, findable and reusable.

Further investment under Super Science is developing Australia's first national research *data storage* infrastructure, which will identify, store and support research data holdings of lasting value and importance. The Research Data Storage Infrastructure (RDSI) project will result in cost-effective, scaled-up, shared storage services, aimed at improving research collaboration.

Continued investment under Super Science in the Australian Research and Education Network (AREN) is improving research *data transfer* and collaboration across the nation and internationally, including for the large datasets produced from disciplines such as radio astronomy and environmental science.

Investments under NCRIS and Super Science in two national high-performance computing (HPC) centres—the National Computational Infrastructure (NCI) at the Australian National University and the Pawsey HPC Centre led by iVEC in Perth—will provide petascale capacity for *data analysis and modelling* for the next five years.

Finally, Super Science investments in research tools, research cloud capacity, virtual research laboratories and a national research server provide researchers with the means to *manipulate and handle research data* from a distance or in multipart collaborations with other researchers. These investments under the National eResearch Collaboration Tools and Resources (NeCTAR) project complement previous investments in research tools, particularly under the NCRIS-funded National eResearch Architecture Taskforce (NeAT).

### *Data generating research infrastructure*

Complementary investments have occurred in specific domains or disciplines where the creation of a data-centric research infrastructure is strategically important. Under NCRIS and Super Science, a range of infrastructure capabilities emphasise data as the output, and include:

- direct *data collecting infrastructure* such as astronomy investments in telescopes
- *data gathering infrastructure*, which brings together, collates or makes coherent data collected through other means, such as investments in the Australian Urban Research Infrastructure Network (AURIN) and the Population Health Research Network (PHRN)
- *infrastructure that performs a combination of both functions* such as the Integrated Marine Observing System (IMOS)
- *infrastructure*, like the two HPC facilities, which as well as providing powerful tools for data manipulation themselves, *generates data* in large quantities and of significant value to researchers.

### 3. Context

The Australian Research Data Infrastructure Strategy proposes a framework to sustain, coordinate, and build Australia's research data system from a solid foundation of investment and capability.

Research data infrastructure, which includes data itself, is a valuable national asset that supports the pursuit of research in all fields. This strategy considers data holistically in the context of the creation, collection, manipulation, integration and re-use of data, together with the knowledge frameworks and infrastructure capabilities that are needed to translate our data assets into research outcomes.

#### Box 2: Data enables weather and climate modelling

Data underpins the models that allow us to predict weather, climate and the risks of extreme weather now and in the future. Weather and climate extremes affect society, our economic competitiveness and our capacity to adapt to change on timescales that span days, seasons, decades and centuries. Observed data ensures that models used to predict weather and climate correctly represent key processes. Data assimilation, for example, has led to dramatic advances in weather forecasting in recent decades. Data also underpin model evaluation, including diagnoses of model strengths and weaknesses, thereby developing enhanced capacity to forecast future threats and opportunities.

Australia's weather and climate models depend on data from Australia and the surrounding region: in particular, satellite observations (critical given the data-sparse Southern Hemisphere); meteorological observations (air temperature and rainfall); ocean temperature, salinity and fluxes (for example, the Integrated Marine Observing System (IMOS), Box 3); in situ land–air fluxes (OzFlux) and atmospheric composition (green-house and reactive gases, aerosols). Data infrastructure is critical so that data is securely archived, discoverable and quality assured, and can be shared and used in a sustained and enduring way.

Current investments are building better national observational data infrastructure. They have also provided Australia with access to unprecedented weather and climate model data from model simulations produced by the Australian Community Climate and Earth-System Simulator (ACCESS) and similar modelling systems around the world (see Box 8). The vast quantity and breadth of these model datasets demand a sophisticated array of data infrastructures to support their storage, access and use, as well as to facilitate linkages between research and operational communities.

A well-integrated data infrastructure that strongly connects observed data systems to the major supercomputing and petascale data systems will ensure researchers and policy makers capitalise on work to date. Such a system will allow data to be mined and exploited for model development and evaluation, and enable ACCESS model simulations to be interwoven with other international model simulations. Physical data infrastructure and data are both key to advances in weather and climate research. However, they need to be combined with tools, software and supporting protocols to enforce version and release control, digital object identifiers, data publishing and documentation. An integrated data system, building on the newly established petascale data and computing environment, will enable the optimum and effective use of the explosion in observed and simulated data, and so improve decision making and policy responses based on this information.

Despite good progress, Australia still lacks an overarching data infrastructure to enable the integration and uptake of data from the weather and climate research community so it can be fed into the development of ACCESS. An overarching data infrastructure will provide a platform for managing data storage, maintenance and access; the integration of hardware, software and people; and links between research, operational and management communities.



The data accumulated from centuries of observation, and the pace of technological change over the last half century, have transformed the processes and nature of knowledge discovery. With the advent of computational modelling and simulation in the mid-twentieth century (referred to as the third paradigm of scientific discovery<sup>2</sup>) data outputs from simulations of complex phenomena have increased enormously. The development of computers capable of building detailed simulations and solving huge numbers of equations very rapidly has enabled researchers to discover and investigate fields of study previously impervious to experiment and direct observation—such as ecosystem, climate, and deep-earth modelling; the mechanics of planetary formation; or the evolution of the cosmos (for example, see Boxes 2, 6 and 8).

In turn, such a data-rich environment has driven increasingly data-intensive research. This period of rapid acceleration in the amount and complexity of data available and vastly expanded possibilities for data creation and manipulation has been referred to as the fourth paradigm of scientific discovery.<sup>3</sup> Researchers in this new environment require an integrated infrastructure system which can seamlessly translate data assets into research outcomes. To support this data-intensive research and optimise the outcomes for researchers in all fields and for the nation, funders and infrastructure designers and operators need to provide better ways to generate, organise, manipulate, share, use and re-use data.

The solution will involve a connected national research data infrastructure system that allows integration throughout the data lifecycle: from processing, to collection, to curation and storage, to re-use. It will also encourage discoverability, and promote open and flexible access arrangements (see Boxes 2 and 3), while allowing funders, operators and users of research data infrastructure to capitalise on future transformative technologies. Policy-makers, as well as funders, designers, operators and users of research data infrastructure, will need new approaches and solutions which take account of changing technologies and environments, including current and future national and international drivers.

### Box 3: A virtual research vessel fleet

Although the earth's oceans cover 70 per cent of the surface of our blue planet, they are massively under-observed. Until fairly recently, ship-based observations provided virtually all of the empirical information we had about the oceans' fundamental role in making our planet habitable.

Satellite technologies have been addressing this gap over the last two decades for the surface oceans, and robotic technologies such as autonomous profiling floats and piloted ocean gliders have more recently been revealing the secrets of the water column to a depth of 2 kilometres.

Ship-based observations remain vital, however, to ground truth satellites, to collect highly integrated or spatially explicit data that cannot be remotely sensed, and to measure the deep ocean (down to 6 kilometres), which comprises the majority of the global water mass.

All data collected by research vessels is therefore highly valuable. Historically, however, technical, logistical, cultural and institutional constraints have prevented researchers from fully exploiting the collective value of research vessel observations taken within the Australian region.

<sup>2</sup> Hey, T 2010, 'The big idea: the next scientific revolution', *Harvard Business Review*, November; Hey, T, Tansley, S & Tolle, K (eds) 2009, *The fourth paradigm: data-intensive scientific discovery*, Microsoft Corporation, pp. xx-xxi, 177, 180; Strawn, GO 2012, 'Scientific research: how many paradigms?', *EducauseReview* (May/June), pp. 26–34.

<sup>3</sup> Hey, 2010; Hey et al., 2009; Strawn, 2012, p. 34.

With funding from the National Collaborative Research Infrastructure Strategy (NCRIS) and the Super Science Initiative, the Integrated Marine Observing System (IMOS) has been gradually instrumenting all research vessels regularly operating in the Australian region with common equipment. In collaboration with the Australian National Data Service (ANDS), IMOS has developed a common delivery system for 'underway' data.

As a result of this work, research vessels are all streaming data into the IMOS Ocean Portal. These vessels include the blue-water marine national facility operated by CSIRO, RV *Southern Surveyor*; the polar research and supply vessel operated by the Australian Antarctic Division, RSV *Aurora Australis*; the shelf-scale vessels operated by the Australian Institute of Marine Science, RVs *Cape Ferguson* and *Solander*; the New Zealand research vessel, RV *Tangaroa*; and the French polar research vessel, *L'Astrolabe*.

By focusing on the research data as infrastructure rather than on the vessels themselves, IMOS, ANDS and their collaborators have created a 'virtual fleet' that is now servicing a national, regional, and global research community concentrated in the Australian region. There is significant potential to expand this concept with partnerships across the Southern, Pacific and Indian ocean basins.

## National drivers

Developments in information and communications technologies (ICT), such as those described above, are revolutionising science, knowledge and ultimately society at large. The research sector is benefiting from the transformative potential of high-speed networks, computational power, ubiquitous sensor networks, and smart tools. Developments in the research sector parallel and build on initiatives such as the rollout of the National Broadband Network (NBN). The NBN is providing key assistance in certain targeted segments of the sector's high-speed AREN, and the increased promotion and uptake of the cloud computing technologies through the National Cloud Computing Strategy.

The significant increase in the rate of research data being created and captured; the range of disciplines and capability areas depending on data; and the substantial potential benefits offered by integrated data generation, analysis, manipulation and re-use require coordination across research infrastructure initiatives and between stakeholders. This includes coordination within the research sector, and also with governments, non-government organisations, the private sector and the community.

#### Box 4: Mapping Australia's soil diversity

Working closely with CSIRO, various government departments, universities and research and development corporations, Bioplatforms Australia (established through NCRIS and supported by the Super Science Initiative) has launched an important project to map soil biodiversity in Australia.

Soil hosts diverse microbial communities that play a critical role in the many ecological processes that underpin agricultural enterprises and influence our natural landscapes. Despite this fundamental role, soil communities are not well characterised in Australia or the rest of the world. This new project is bringing together leading researchers in a novel investigation of the diversity and ecological function of Australian soils. The Biome of Australian Soil Environments (BASE) project offers unique opportunities to catalogue and describe the communities of microscopic organisms that exist in soil, and define their intrinsic relationship with plants, soil health and agricultural productivity. Under BASE, soil samples from different regions and land uses are collected and analysed to create a reference map of Australian soil and enable detailed research on the microbial communities extracted from each site.

Comprehensive mapping of Australian soils has not been undertaken before and offers many discovery opportunities. Researchers will be able to investigate the role of soil microbial communities in ecological processes such as carbon cycling, degradation of contaminants and defence against soil-borne diseases. BASE will provide the datasets needed to define and model different microbial communities and relate their structure and function to contrasting environments, vegetation and land use. Such data is critical to achieving ecological stability and sustainable agricultural production and has a range of other vital applications. For example, soil datasets are critical for investigating ways to manage soils for carbon sinks and they can be used to investigate the management of crop vulnerability.

Bioplatforms Australia will create large genomics datasets for BASE in collaboration with soil experts. The datasets will be linked with contextual data such as soil chemistry, GPS information and environmental observations. This will give an expanded view of soil communities and their symbiotic and co-evolutionary relationship with plants. Ultimately, it will also allow researchers to quantify and compare different soil communities across Australia. Soil samples are collected from national reserves and agricultural monitoring sites. Access to these and other sites, together with land-use history, will ensure a continent-wide inventory of biodiversity and enable relevant research into soil resilience and agricultural productivity.

BASE data will be publicly available for the benefit of broader research applications. Soil datasets can be linked with existing overland surveys, meteorological data, geological data and other knowledge of the Australian continent and its land use. BASE will also align and partner with the Earth Microbiome Project, an international effort which aims to characterise more than 200,000 microbial samples from around the world.

Source: Bioplatforms Australia,  
[www.bioplatforms.com.au/special-initiatives/environment/soil-biodiversity](http://www.bioplatforms.com.au/special-initiatives/environment/soil-biodiversity)

Specific advances in ICT for research have positioned Australia well in the global context. A broader approach to the creation, management, storage and re-use of research data is required as Australian and international research data continue to grow to monumental size and complexity. Part of that complexity arises from the fact that some significant data can only be obtained or may only be readily available from sources outside the research sector, such as industry or government (see Boxes 4 and 7).

The Australian Government has invested significantly in research infrastructure to support the emphasis on data in research. These investments include ICT-based research infrastructure (such as supercomputers and high-speed networks) known as eResearch infrastructure, and domain-based research infrastructure (such as telescopes and marine sensor networks) that focus on data as a resource to be collected, analysed and used to enable research outcomes. We now have a suite of state-of-the-art national facilities to boost Australia's ability to conduct outstanding research, to collaborate internationally, and to attract the best researchers from around the world (see Boxes 1, 3, 4, 5, 6, 7 and 9).

### Box 5: Linking Australia's stories with HuNI

The Humanities Networked Infrastructure (HuNI) is a national virtual laboratory project being developed as part of the National eResearch Collaboration Tools and Resources (NeCTAR) project, in a partnership between 13 public institutions, led by Deakin University. HuNI is using a linked data framework to combine information from 28 of Australia's most significant cultural datasets. These datasets comprise more than 2 million authoritative records relating to the people, objects and events that make up the country's rich heritage, covering fields as varied as literature, art and design, theatre, film and visual media, history, biography, music and archaeology. These datasets have been developed and used by subject and technical experts over many years.

HuNI is also deploying an integrated suite of software tools to enable researchers to work with this large-scale aggregation of linked data. Drawing on an extensive collection of user stories and a detailed analysis of user requirements, these tools cover key tasks for working with large and complex datasets in the humanities and creative arts, and include such functions as visualisation, annotation, browsing, sharing and mapping.

HuNI will enhance researchers' ability to work collaboratively or independently with the data. Cutting-edge analytical tools will yield new scholarly outcomes and deepen our understanding of Australian culture across time. Through HuNI, cultural data will be available for linking with data from the sciences and the social sciences. Designed for future expansion, HuNI will transform research methods in the humanities and creative arts.

Source: HuNI Project Management Plan, [www.huni.net.au/](http://www.huni.net.au/)

## International trends and drivers

Funding agencies worldwide are re-positioning their priorities in response to the new data environment. Recent examples include reports and strategies such as *Riding the wave: How Europe can gain from the rising tide of scientific data* (Appendix A) and the National Science Foundation's *Cyberinfrastructure vision for 21st century discovery*. Further, there is growing acknowledgement internationally of the benefits of opening access to publicly funded research. Such acknowledgments include the G8 Science Ministers Statement (London UK), released on 13 June 2013, which included support for collaboration and agreement on open scientific research data and increasing access to the peer-reviewed, published results of scientific data (Appendix B).

Vast benefits will accrue to countries which embrace the opportunities presented by data-intensive research. Similarly, there are strong messages about the long-term costs of not embracing these new opportunities, doing this poorly, or doing it too late. A number of the breakout boxes describe the value of international data collaboration and access arrangements for research in Australia, and

the value of Australian research data for addressing matters of critical importance internationally. For example, Box 6 discusses Australia's involvement in the International Virtual Observatory Alliance (IVOA) which comprises 17 countries developing and promoting astronomy data interoperability standards.

In a number of other areas, Australian research data infrastructure is highly regarded internationally, as evidenced by Australian leadership in global research and infrastructure collaborations such as the Research Data Alliance, the Global Biodiversity Information Facility (GBIF) and the Global Ocean Observing System.

#### **Box 6: Virtual observatory standards enable advances in astronomical research**

The International Virtual Observatory Alliance (IVOA) comprises 17 countries developing and promoting astronomy data interoperability standards. These standards provide the framework for a worldwide virtual observatory that enables astronomers to share, discover, use, and reuse data. The IVOA was a founding member of the Research Data Alliance (supported by the Australian Government through the Australian National Data Service) and represents one of the most advanced domain-specific alliances in this new cross-discipline effort. The success and rapid evolution of IVOA is largely due to the widespread adoption of IVOA standards by data and service providers, and uptake of virtual observatory-enabled tools by astronomers.

Australia was a founding member of the IVOA and has made key contributions to the development of data standards and services. Next-generation radio telescopes like the Square Kilometre Array (SKA) and the Australian SKA Pathfinder (ASKAP) will produce unprecedented floods of data, and Australia is in an exciting position to drive the development of new and improved standards to handle massive 3D radio data cubes.

Not only do IVOA standards and services facilitate data use, they are powering a new era in astronomical research and discovery, by helping astronomers access and combine enormous datasets—spanning radio, optical, ultraviolet and X-ray wavelengths, along with theoretical data—in order to construct a complete picture of cosmic evolution.

Australia is currently involved in several projects that link a broad range of datasets from telescopes around the world and in space, and store them in a centralised repository (for example, the Galaxy and Mass Assembly Survey and the Australia Telescope Large Area Survey). However, the growth of data volumes requires moving to a model of distributed storage and seamless query and access. This approach is currently being implemented within the All-Sky Virtual Observatory project, funded by the Australian Government, through the National eResearch Collaboration Tools and Resources (NeCTAR) project. The first phase of this project will make simulation data (housed at Swinburne University on an Education Investment Fund supercomputer) and optical survey data (housed at the National Computational Infrastructure facility, Canberra) accessible and analysable via IVOA-compatible services. The next phase of this project will aim to include other datasets of national significance as cornerstones of a growing Federation of National Astronomy Datasets.

Investment in data infrastructure aims to make Australian researchers 'collaborators of choice' in a global research environment where data is the new currency.<sup>4</sup> In many cases, Australian researchers must have access to robust data infrastructure if they want to be collaborators on global research

<sup>4</sup> World Economic Forum Annual Meeting, Davos-Klosters, Switzerland, 22–25 January 2014, see [www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development](http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development).

projects funded through international programmes that require data management and dissemination as part of grant conditions (for example, Boxes 6 and 7).<sup>5</sup>

International agencies increasingly recognise that data, being a pervasive and potentially long-lived information asset for all of society, needs planning and coordination. For example, in the United States, the National Science and Technology Council chartered the Interagency Working Group on Digital Data (IWGDD) to 'develop and promote the implementation of a strategic plan for the [US] Federal government to cultivate an open interoperable framework to ensure reliable preservation and effective access to digital data for research, development, and education in science, technology, and engineering'.<sup>6</sup> The IWGDD recognised the need for a whole of government approach to research data infrastructure, policy, and investment. In addition, on 22 February 2013, the US Government, through the Office of Science and Technology Policy (OSTP) released an open access policy memorandum to promote easy access to the results of publicly funded scientific research. Federal agencies with more than \$100 million in research and development expenditure have been directed to develop plans to make the published results of federally funded research freely available to the public within one year of publication and require researchers to account for and manage the digital data resulting from federally funded scientific research.<sup>7</sup>

## Investment to date

Australia has made substantial research data infrastructure investments that have delivered significant advantage to Australia's research sector, including in eResearch infrastructure and data-generating research infrastructure. The breakout boxes describe recent Australian examples.

As a result of some of these investments, in many quarters Australia is considered a global partner of choice for data-intensive research. These partnerships reflect our global involvement; for example the IMOS is a partner in a major investment in marine data initiated by the European Union's Seventh Framework Programme (EU FP7)—the Ocean Data Interoperability Platform; the Terrestrial Ecosystem Research Network (TERN) has been invited to participate in the National Science Foundation-sponsored National Ecological Observatory Network; the Atlas of Living Australia (ALA) is GBIF's Australian node; and ANDS facilitates Australia's participation in the internationally focused Research Data Alliance, of which Australia, the United States and the European Union are foundation members.

This investment in research data infrastructure covers acquisition of new data through programmes such as flux towers or shared microscopes; the assembly of data as implemented by facilities like the ALA; or the enhancement of data availability and usefulness of data through improved storage, tools, computation and access, and broader application (for example, this is enabled by NCRIS and

---

<sup>5</sup> See Appendix B. Some examples: the Intergovernmental Group on Earth Observations' *Global Earth Observation system of Systems (GEOSS) Data sharing principles*; the UK Open Data White Paper: *Unleashing the Potential* (2012); the UK Royal Society's report, *Science as an Open Enterprise* (2012); the 2011 Interim Report to the International Council for Science (ICSU) Committee on Scientific Planning and Review; the 2010 Wellcome Trust *Policy on data management and sharing*; and the 2007 *OECD Principles and Guidelines for Access to Research Data from Public Funding*.

<sup>6</sup> Statement of the Director, National Coordination Office for Networking and Information Technology Research and Development to the Committee on Science and Technology, US House of Representatives, 1 April 2009 [www.whitehouse.gov/files/documents/cyber/Congress%20-%20Christopher%20Greer%20-%20NITRD%20Greer-reauth-testimony3.0.pdf](http://www.whitehouse.gov/files/documents/cyber/Congress%20-%20Christopher%20Greer%20-%20NITRD%20Greer-reauth-testimony3.0.pdf)

<sup>7</sup> OSTP, 2013, Memorandum Increasing Access to the Results of Federally Funded Scientific Research [www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)



Super Science supported eResearch capabilities including the RDSI, NeCTAR, NCI, Pawsey Centre and ANDS projects, mentioned in Box 1).

The Australian Government, together with co-investors from state and territory government agencies, the research sector and industry, has provided substantial investment in research infrastructure, including data and data infrastructure. This includes Australian Government investments through the \$542 million NCRIS and the \$1.1 billion Education Investment Fund supported Super Science Initiative. In its May 2013 Budget, the Government announced additional NCRIS funding of \$185.9 million over two years (2013–14 and 2014–15) to support the operation and maintenance of the most critical projects established by NCRIS and Super Science. This additional NCRIS funding is intended to ensure the continued operation of established facilities for two years. In its May 2014 Budget, the Government announced a further \$150 million over one year (2015–16) for the operation and maintenance of critical research infrastructure.

These investments have been underpinned by successive consultative roadmapping of strategic priorities for research infrastructure investment grouped as capabilities, through the 2006 NCRIS Roadmap (in Appendix A of the *2008 Strategic Roadmap for Australian Research Infrastructure*), and the *2011 Strategic Roadmap for Australian Research Infrastructure*. These roadmaps placed a fundamental focus on access, collaboration and the ability to fund operating costs and thus support system-wide access. The latest 2011 Roadmap articulates the priority research infrastructure areas on a national scale (capability areas) to develop Australia's research capacity and enhance research outcomes over the subsequent five to 10 years.

Outside the NCRIS and Super Science model, data-holding institutions are similarly investing significantly. BoM, the ABS, and Geoscience Australia are among Australian Government agencies that invest heavily in the generation and management of data that is crucial to research (for example, Boxes 2 and 7). In addition, research institutions, including publicly funded research agencies and many universities, are investing in infrastructure to manage their research data and are keen to maximise the opportunities afforded by access to internationally significant data holdings (for example, Box 8). The Australian Research Council's annual Linkage Infrastructure, Equipment and Facilities scheme provides funding for research infrastructure, equipment and facilities to eligible organisations. The scheme enables higher education researchers to participate in cooperative initiatives so that expensive infrastructure, equipment and facilities can be shared between higher education organisations, and also with industry. The scheme also fosters collaboration through its support of the cooperative use of international or national research facilities, consistent with the principles of NCRIS and Super Science.

In respect to data generated by the Government, the *APS200 Project: The place of science in policy development in the public service* (2012) systematically reviewed the ways in which scientific input is used to inform policy development in the Australian Public Service.<sup>8</sup> This 2012 report noted 'a need to facilitate access to and use of scientific data and research services to support policy', and that 'government can maximise its investments in research and data by encouraging data access, sharing and integration to support further research and policy development'.

Government agencies and organisations have also collected valuable data. Their combined data holdings now form a substantial investment. These datasets collectively and individually are a significant resource that should be available to researchers if possible. It is clear from existing public sector/research sector partnerships that data generated and collected by the public sector is an

---

<sup>8</sup> Australian Government, 2012, *APS200 Project: The place of science in policy development in the public service* [www.industry.gov.au/science/Pages/APS200ProjectScienceinPolicy.aspx](http://www.industry.gov.au/science/Pages/APS200ProjectScienceinPolicy.aspx)

important asset for research, and cannot be dealt with in isolation from research developments. If these datasets' utility for research is to be realised, their management, custodianship and protection will need to be undertaken in recognition of their potential use for purposes other than that for which they were collected. Box 7 provides an example of AURIN's use and re-use of data generated by the ABS and others in urban research with positive outcomes for researchers, policy makers and analysts.

### **Box 7: Enhancing the usability of census data for urban research**

The Australian Urban Research Infrastructure Network (AURIN), which was established in the second half of 2010, is a \$20 million project funded under the Super Science Initiative. AURIN is building an e-Infrastructure capability that will integrate data from multiple sources and use open source eResearch tools to visualise data and conduct statistical and spatial analysis and modelling of data. Aimed at urban and built environment researchers, the project will facilitate online access to diverse data at various levels of spatial scale held by public agencies, a number of private sector organisations, and generated by researchers.

The AURIN e-Infrastructure will facilitate secure access to individual data, and enable unit record data to be integrated with spatial objective data for interrogation online. To ensure individual identity is protected, the researcher will be provided with the results but not direct access to the unit record data.

Among 30 or so projects so far in progress, AURIN is collaborating with the Australian Bureau of Statistics (ABS) for a federated data hub to provide users with online access to 2011 census data and other ABS data products. The project is innovative because it provides capability to conduct online manipulation of data—such as that generated by the 2011 census—through the application of analytic tools developed in open source by AURIN. These tools automate the conversion of count data into the sort of derived variables that researchers typically use, and support the online analysis of the data and its visualisation through GIS-enabled mapping routines.

The project is demonstrating how such transformation in machine-to-machine interaction and cloud computing can enhance the way ABS and other Australian Statistical Geography Standard (ASGS)-embedded datasets may be used in research and policy analysis. It employs an eResearch approach that overcomes the necessity for users of ABS data to download the data from the ABS website and re-load it into the users' own (usually proprietary) data analysis and GIS visualisation packages.

The National Collaborative Research Infrastructure Strategy (NCRIS)-funded Australian National Data Service is a collaborating partner in the AURIN/ABS project, along with groups in a number of Australia's universities.

The Government 2.0 Taskforce Report recognises that information collected by, or for, the public sector is a national resource that should be managed to maximise public benefit.<sup>9</sup> Data.gov.au is an important initiative that is beginning to address the availability of government-funded data by providing access to public data from the Australian, state and territory governments.

Government investments in, or in support of, research data infrastructure have also been critical of the effectiveness of Australian researchers. All Australian, state and territory governments have significant ongoing investments in the collection and management of public-sector source data and

<sup>9</sup> Government 2.0 Taskforce, 2009, *Engage: Getting on with Government 2.0*  
[www.finance.gov.au/publications/gov20taskforcereport/index.html](http://www.finance.gov.au/publications/gov20taskforcereport/index.html).



related data that support a wide range of research. The health system, in particular, has substantial data holdings that are vital to health and human services research.

These factors combine to provide Australia with a globally competitive research data advantage.

#### **Box 8: International collaboration and coordination in world climate research**

International collaboration and coordination among over 20 modelling groups around the world is a key component in the World Climate Research Programme for the Coupled Model Intercomparison Project Phase 5 (CMIP5). CMIP5 provides a framework for coordinated climate change experiments and provides critical data for the Intergovernmental Panel on Climate Change Fifth Assessment Report. Australian researchers contribute results from the ACCESS modelling system (Box 2) to CMIP5. Critically, our data can be analysed in the context of results from overseas groups to identify common threats and opportunities associated with climate change. However, this places enormous pressures on our current and future research data infrastructure.

These, and other worldwide scientific data collections, are accessible through the Earth System Grid Federation (ESGF) gateway and data nodes for serving climate and environmental science data. Australia's ESG node has been established at the National Computational Infrastructure facility, Canberra. The ESGF has established a standard for international data publishing and data access services for scientific data collections.

Importantly CMIP5 and the ESGF are now directly involved in serving the climate science needs of an ever-increasing demand for climate change information. It has grown beyond the rationale of serving research needs and is already underpinning the serving of climate change information for the nation. This has proved a challenge. Providing a robust framework serving both the research and wider community remains a key challenge, as articulated in the 2012 document, *A plan for implementing climate change science in Australia*, ([www.climatechange.gov.au/sites/climatechange/files/documents/03\\_2013/plan-implementing-climate-change-science-australia.pdf](http://www.climatechange.gov.au/sites/climatechange/files/documents/03_2013/plan-implementing-climate-change-science-australia.pdf)).

This demand will only increase. Currently almost three petabytes of storage is needed to meet immediate requirements. By 2015–19 CMIP6 will be underway, with increasing model complexity and resolution and a deeper commitment to fulfil climate service requirements. A robust framework of high-performance computing, data storage in excess of 50 petabytes, and high-speed communications serving an interface to operational services will be required to meet such national needs.

#### **Future investment**

It should be noted that at the time of the release of this strategy, the 2011 Roadmap has not yet been funded and a stable funding environment for national research infrastructure has yet to be established. The strategy does not estimate the quantum of funding needed to support future investments in data-intensive research infrastructure. Existing mechanisms, such as regular research infrastructure roadmapping exercises, will serve as a foundation for future identification of detailed funding envelopes. This strategy, instead, emphasises the need for appropriate approaches to existing and future investments in data-intensive research infrastructure and proposes a foundation for how various parties can work together productively to enhance development of data infrastructure and benefits from data in years to come.

## 4. Principles

Australian Government investments in research data infrastructure should be guided by the principles set out in existing strategies—in particular the Strategic Framework for Research Infrastructure Investment principles, which appear in the *2011 Strategic Roadmap for Australian Research Infrastructure* (see Appendix C of this document).

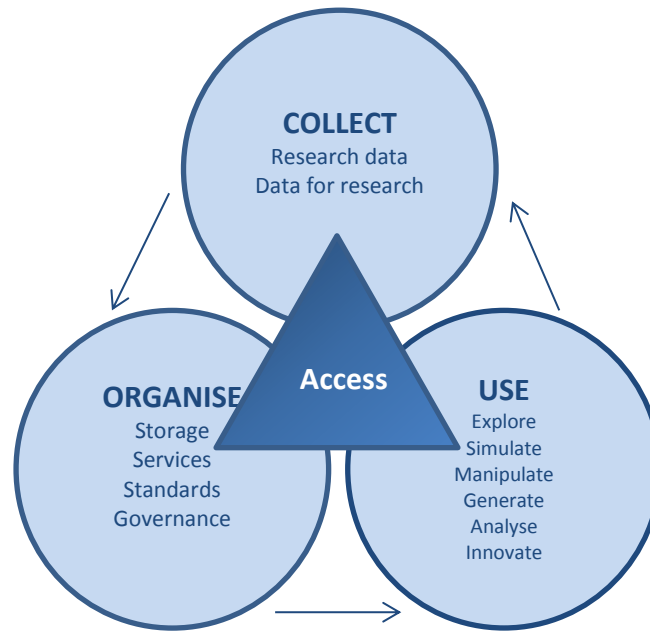
### Strategic Framework for Research Infrastructure Investment principles

Continuity of funding  
Holistic funding  
Prioritisation  
Excellence in research infrastructure  
Collaboration  
Co-investment  
Access and pricing for Australian-based infrastructure  
Access to overseas-based infrastructure  
Evaluation and monitoring

These agreed principles provide a basis for policy makers, investors, developers, operators and users to build and sustain an effective, holistic Australian research data infrastructure system, as displayed in Figure 2, that:

- *collects* data systematically and intentionally
- *organises* data and makes it discoverable and accessible
- *uses* data many times over and in as many ways as possible.

Access to research data, access to data for research, and access to enabling infrastructure critically supports this system.



*Figure 2: Components of a holistic Australian research data infrastructure system*

The Australian Research Data Infrastructure Strategy has been developed to be consistent with the principles in the 2011 Roadmap.

These two key Government strategies outline principles that apply broadly to all research infrastructure. The discussion below articulates how those principles (which have been emphasised in bold) apply to specific characteristics of research data infrastructure as an enabling capability that underpins all fields of research.

A national, **collaborative** approach to investment in research data infrastructure will reduce duplication, enhance economic and efficiency benefits, and optimise research outcomes. Appropriate access arrangements and agreed standards will facilitate collaboration, fostering multi-disciplinary research uses for existing data, enabling researchers to address emerging problems in new ways (for example, Box 9).

### Box 9: Tagged seals help solve 30-year mystery, serendipitously

Sensor-equipped southern elephant seals have helped scientists to discover a key source of cold, salty water that helps to regulate the earth's climate. Antarctic bottom water (AABW), which is dense and cold, was known to originate from three sources around the Antarctic coastline. For more than 30 years, scientists have speculated about the location of a fourth undiscovered source of AABW.

The Integrated Marine Observing System (IMOS) deploys satellite tags on southern elephant seals in order to incorporate key physical and biological information into spatial models designed to inform management strategies for areas of ecological significance within the Southern Ocean. This research continues to progress well.

However, because all IMOS data is openly accessible, an entirely different group of Australian and Japanese researchers were able to repurpose the seal tagging data and contribute to confirmation of the existence of a fourth stream of AABW coming from intense sea ice formation in the Cape Darnley Polynya, north-west of the Amery Ice Shelf. The research was published in *Nature Geoscience*<sup>1</sup>.

Because the seals went to an area of the Antarctic coastline that no ship was ever going to reach, particularly in the middle of winter, they measured the most extreme dense shelf water anywhere around Antarctica. Several of the seals foraged on the continental slope as far down as 2 kilometres, punching into a layer of dense Antarctic bottom water cascading down to the abyss. This data provided rare and valuable wintertime measurements of the AABW process, helping to solve a difficult oceanography puzzle using data derived from a research project on seal ecology. This function of IMOS provides a compelling demonstration of the value of systematic data collection and management, combined with a policy of open access.

<sup>1</sup> Ohshima, K, Fukamachi, Y, Williams, G, Nihashi, S, Roquet, F, Kitade, Y, Tamura, T, Hirano, D, Herraiz-Borreguero, L, Field, I, Hindell, M, Aoki, S & Wakatsuchi, M 2013, 'Antarctic bottom water production by intense sea-ice formation in the Cape Darnley polynya', *Nature Geoscience*, vol. 6, pp. 235–40 .

Policies and standards for **access for Australian-based research data infrastructure** should reduce barriers to the uptake and use of data across research fields and institutions. Access arrangements should optimise the use of infrastructure, increase the use of data, and support collaboration and international partnerships. **Access to overseas-based research data infrastructure** should be considered when it is cost effective. Open access will encourage international collaboration and co-funding and improve return on investment. Research data infrastructure should **support global quality and scale** through facilitating open data, observing and encouraging international better practice and contributing to the development of international standards.

Research data infrastructure should **increase the stock of knowledge** for use now and into the future, including through improving durability and discoverability, facilitating access and collaboration, and improving research skills, technical capability and digital literacy. Identification and **prioritisation** of the collection, curation and storage of data of lasting value and significance will help safeguard Australia's national data assets and sustain an effective, productive and transformative research environment which, in turn, supports a vibrant industry sector.

A joined-up research data environment will be a significant component of **a strong, cohesive research fabric** that will support basic and applied research across a broad range of disciplines

including the development of generic research data infrastructure, that enables rich connection of data across research fields so that new tools can be developed upon demand.

Two important additional principles strongly aligned with the 2011 Roadmap principles include the establishment of effective **coordination and governance mechanisms** and the **development and promotion of shared goals and standards** among research data stakeholders.

The establishment of robust coordination and governance mechanisms will support effective planning for a cohesive, enduring, and coherent research data environment. Good governance will encourage collaboration within and across research areas, nationally and internationally, and ensure the effective establishment, operation and management of research data infrastructure. Planning and coordination will encourage the development and implementation of technical standards.

Some of the most pressing problems for Australia require new ways of undertaking, sharing and harnessing the significant amount of research data that is available. The development of a broad coalition of stakeholders will strengthen the above principles to build and sustain an efficient, effective and flexible research data infrastructure environment. This coalition could include representatives from government agencies, the private research sectors, and citizen scientists, who collectively promote common standards for data sharing, discoverability, openness and re-use, while empowering users of research data and data for research.

By laying the right foundations through application of these broadly-agreed principles, research data infrastructure will help collect and generate the data to **enhance productivity growth** and address **Australia's key economic, social and environmental challenges**.

## 5. Needs

For Australia to increase productivity and address grand challenges over coming decades, it will require a national research capability that can deliver competitive advantage in a data-rich future. The research paradigm is shifting from one of protecting relatively limited amounts of sparsely distributed data, to one of exploiting ever-increasing amounts of data from multiple sources at local, regional, national and global scales.<sup>10</sup> Competitive research advantage is now gained from the smartest or fastest use of data and the ability to explore rich data deeply or broadly, as required. Individual researchers and institutions working in isolation cannot effectively exploit big data by themselves. A coordinated, appropriately scaled approach to data infrastructure will greatly increase the probability that Australian research and technology will deliver sustained benefit to current and future generations. The continued ability of Australian researchers to collaborate effectively internationally, to help resolve global and national problems while increasing their own opportunities to reap citations and career recognition, depends critically on the capacity to access, generate, manipulate, share and repurpose research data on an international stage.

Within this context, the Australian Government needs to sustain investment in high-performing research data infrastructure created over the last decade. Better coordination of future infrastructure investments—at the research sector level, within data-intensive research domains and across and between research institutions—will be another critical element of an outstanding research system. With sustained investment and enhanced coordination, we will be able to increase the benefits delivered through research data infrastructure by repositioning effort into known priority areas, taking full advantage of emergent opportunities, and laying a platform for addressing yet-to-be-identified problems.

In order to realise our research potential, Australian institutions and researchers must continue to collect nationally significant data—both data generated from research, and data for research sourced through government agencies, private industry, and citizen science. We must organise that data better by treating it as an appreciating asset, and by being systematic and methodical about its management, availability, and governance (for example, Box 10). We must also enable novel use of that data to deliver greater benefits to a broader range of beneficiaries, in a future that will require science and research to be increasingly multidisciplinary, cross-disciplinary and trans-disciplinary. Access to data lies at the heart of these needs. Research data infrastructure underpins solutions to these needs.

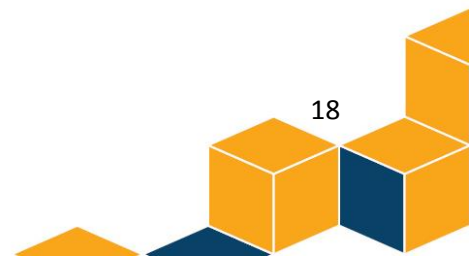
Australia's ongoing research success will therefore depend on investment in an underpinning national research data infrastructure. This will enable development and improved management of national and global research data environments through engagement of researchers and research institutions with a data-driven future for Australian research and technology.

Research data infrastructure investments under NCRIS and Super Science have made important progress in the areas of data management, data sharing, access and availability across and between research institutions, including dealing with the creation, capture, aggregation, transmission, storage, standards, access, re-use and curation of data. The significance and importance of research data to Australia's future are recognised throughout the 2011 Roadmap.

We now need to embed a national research data infrastructure framework within the Australian research system, in order to lock in the benefits of recent investments and position Australians to deliver ever-increasing value from research in a data-rich future.

---

<sup>10</sup> Hey, 2010; Hey et al., 2009.



Such a research data infrastructure framework must reflect the need to sustain investment in high-performing capabilities, better coordinate across investments at national, domain and institutional levels, and increase and improve delivery of benefit to stakeholders. It must articulate the requirement to collect research data systematically, organise research data as an appreciating asset, and enable its use in a future that will be increasingly data-driven, and non-deterministic.

The outcomes of a successful national research data infrastructure framework will be:

- sustained, priority research data collections and data generation and management infrastructure at national, research-domain and institutional levels
- an environment in which researchers can benefit from institutional and sector-wide data and infrastructure access arrangements that are clear and well-governed
- delivery of enhanced research outcomes in line with national priorities, as a result of better sharing and re-use of research data and data for research.

Drawing on these desiderata, the following framework (Figure 3) for an Australian research data infrastructure is proposed, from which the committee has derived 18 specific recommendations (Section 7).

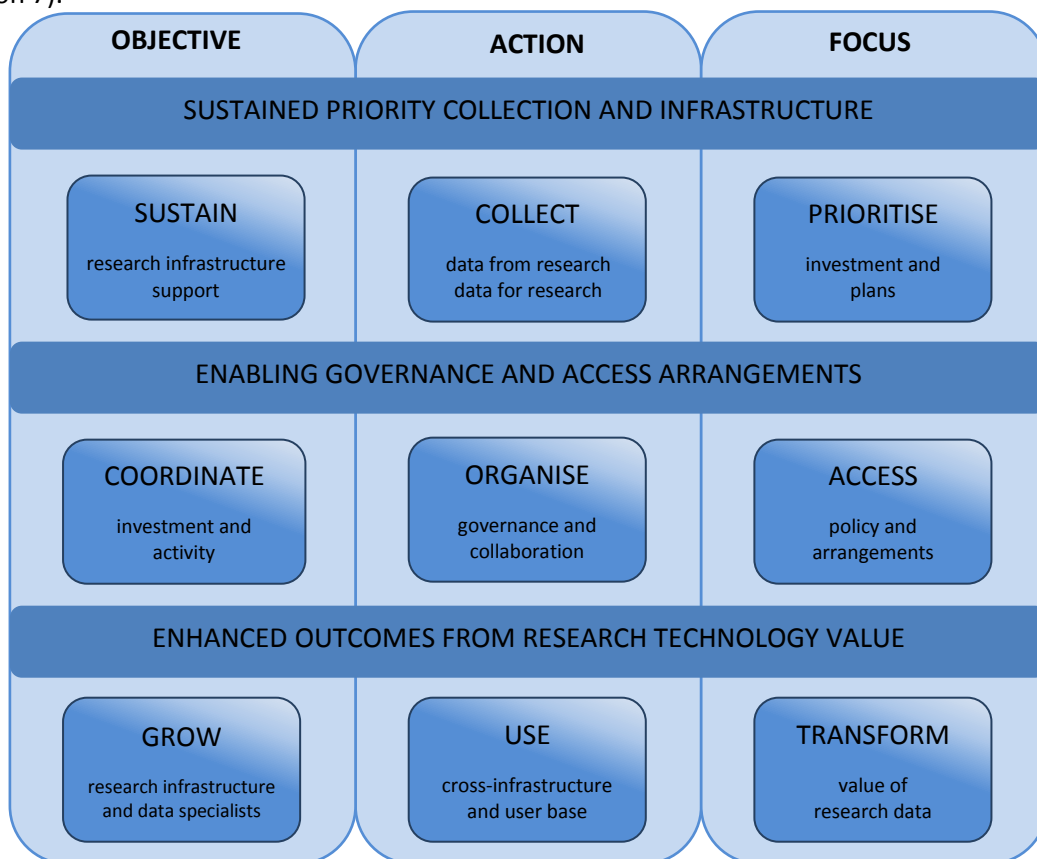


Figure 3: Proposed framework for an Australian research data infrastructure

Figure 3 shows that sustained priority collections and infrastructure entail sustained support of research infrastructure, the continued collection of data from research and data for research, and a level of prioritisation of investments in view of limited resources. To underpin such commitments, the appropriate enabling governance and access arrangements are needed for coordination and organisation of investment and activity and to ensure both effective collaboration and access to data by the users and researchers. Expected outcomes of such a framework include the growth of data

specialists and research infrastructures, a continued cross-linkage of infrastructures and users, and transformation of datasets through new added value.

## Where the future lies

Our aim must be to ensure that by default, research data is captured and discoverable at the time of creation, and that initiatives are in place to deal with already existing datasets, collections and publications.<sup>11</sup> It will be critical to recognise the foundational role of existing eResearch infrastructure investments, and develop the capability of research institutions in automated capture, publication and sharing of research data. Further, it will also be critical to ensure that the research sector as a whole can coordinate and, where necessary, require and reward the necessary changes in practice.

The 18 recommendations in Section 7 point the way to supporting the changes necessary to achieve a coordinated research data infrastructure environment.

The effectiveness of the recommendations and their successful implementation depends on the support and participation of key stakeholders across all levels of government, the research sector, industry, and other funders, operators and users of research data infrastructure.

A key recommendation is to establish a national research data infrastructure advisory committee (Recommendation 6), which will review, coordinate and provide coherence to research data infrastructure investments. The aim is an effectively coordinated national research system which fosters globally significant research.

One of the committee's first tasks will be to prioritise implementation of the 18 recommendations. The timescale for implementation will depend not on relative importance, but on availability of funding, capability, capacity and preparedness of all stakeholders, as well as perceived urgency and degree of complexity.

---

<sup>11</sup> Note that *discoverable* need not imply *accessible* at time of creation. But if data is created with discoverability in mind, it will be easier to ensure accessibility later.



### **Box 10: SURE: Demonstrating the importance of linked health data to population health research**

Population health research often requires access to large amounts of linked population health data. The National Collaborative Research Infrastructure Strategy (NCRIS) and Super Science Initiative have provided funding through the NCRIS-funded Population Health Research Network (PHRN) to significantly expand Australia's capacity to generate and access linked health data. Researchers use this data to investigate important critical public health issues, including hospital-related mortality, burden of injury, and childhood immunisation.

The Sax Institute in Sydney has received funding through PHRN to support this research. It has developed the Secure Unified Research Environment (SURE) that allows approved researchers to place research data in a secure laboratory space and remotely access and analyse it. A range of tools are available to assist the researchers and training is provided. While researchers can access unit record data within the laboratory, they are only permitted to remove summary data.

SURE is supporting collaborations across Australia so that researchers can now access the data from wherever they are based. SURE also helps protect linked population health information by ensuring that the research dataset is always held in a secure curated environment and only summary data can be removed.

## 6. Challenges and opportunities

Most areas of research are simply not feasible without research data infrastructure; for example, areas requiring detailed simulation and modelling, like astronomy and meteorological research, or remote data collection (see Boxes 6, 7, 8 and 11). This type of underpinning infrastructure is as fundamental to these researchers as a ship is to an oceanographer, or a telescope is to an astronomer. Non-delivery of research data infrastructure means that Australia chooses not to engage in some very major projects, and will fail to reap the full transformative research and technological benefits.

While the value of effective research data infrastructure and the opportunities that it creates are recognised widely, significant challenges exist and are considerations for this report.

### Box 11: Challenges and opportunities for astronomy data collection on the Antarctic Plateau

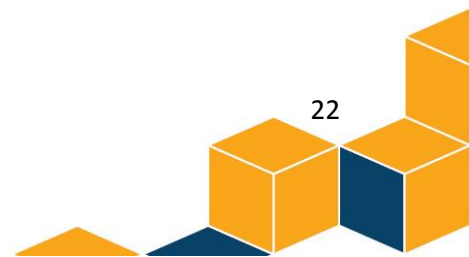
It is well-established that the Antarctic Plateau is an excellent site for an astronomical observatory. The high altitude and extremely dry, cold and stable atmosphere produce ideal conditions for optical, infra-red and sub-millimetre astronomy. Furthermore, the continuous temporal coverage over an Antarctic winter provides some unique scientific opportunities. Australian astronomers are involved in a number of projects across the Antarctic Plateau, including the Chinese-led Antarctic Schmidt Telescopes (AST3) project to erect three wide-field optical telescopes at Dome A, the highest point on the plateau.

A major limitation to Antarctic astronomy is the cost and low-bandwidth of the communications with the outside world. Currently, data is retrieved from the first of three remotely-operated AST3 telescopes using Iridium Openport. A single image from this telescope is 110 MB—about the equivalent of the monthly data allowance. In normal operation the telescope produces 1 to 2 terabytes of data per month, which far exceeds the capability of the Iridium OpenPort system.

To retrieve the data recorded over the Antarctic winter a traverse team is sent in, taking two weeks by tractor to travel 1,300 km to Dome A. The data is then returned on hard disks via an icebreaker to Fremantle, where a copy is made, and sent back to the University of New South Wales.

The scientific value of the data is significantly reduced by the delays in retrieval. In particular, discoveries of variable astronomical objects need to be followed up using other telescopes within a short period of time.

Developments such as Antarctic Broadband, supported under the Australian Government's Australian Space Research Program, could revolutionise the astronomy achievable from Dome A and significantly boost Australia's contribution to this collaborative project with China. A link bandwidth of between 100 gigabytes and 2 terabytes a month would allow the transmission of an important fraction of the data from the telescope, and would allow, for example, immediate detection and follow-up of transient astronomical objects—phenomena which can be observed for typically not more than a few days (for example, supernovae and gamma-ray bursts).



## Challenges

### Investment environment

Episodic funding presents a significant impediment to long-term planning for research data infrastructure. Long-term, sustainable funding is needed to develop, operate and sustain national-level research data infrastructure and to capitalise on current investment.

Government and industry are subject to global economic cycles, and in times of fiscal restraint funding may be limited or unavailable. The risks posed in such an environment include loss of expertise and the curtailing or ceasing of developments needed for the growth of the sector.

### National data assets

Australia has rich data resources developed over many decades. While these resources are often developed for a specific purpose and held by particular institutions or researchers within the scientific community, in many cases they acquire a broader audience or purpose that means they can validly be recognised as a national asset.

In these cases, it is important to ensure that appropriate infrastructure arrangements are in place to ensure future access is enabled. Where that data requires analysis and modelling to make it meaningful, appropriate supporting infrastructure is also needed to ensure the broadest possible benefits can be realised from Australia's research.

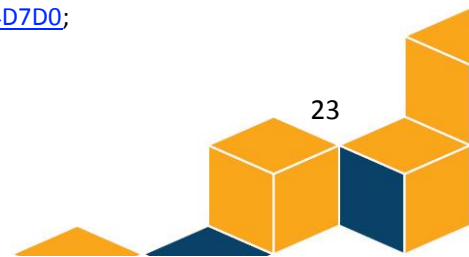
Coordinated, inter-organisational approaches will often be required to ensure data that has national research significance is adequately managed in appropriate research data infrastructure. Source data—for example, administrative data related to human service provision—that is used by researchers and drawn into the national research data infrastructure system, is frequently collected on an agency-by-agency basis using legacy systems that vary within and between organisations as well as across disciplines and sectors.

However, the datasets held by organisations will yield maximum value to researchers when they are organised into discoverable collections, stored in suitable and accessible infrastructure and made appropriately re-useable.

In an environment characterised by a diversity of organisational data management systems, achieving interoperability between these systems is a challenge, but the benefits are considerable. For instance, linking major research data and public data gateways (such as Research Data Australia and data.gov.au) or Australian data to international datasets—for example, the ALA is the Australian node of GBIF—provides significant advantages to researchers and policy makers.<sup>12</sup>

In cases where data is managed at the level of the individual researcher rather than the institution, it is particularly important that researchers are aware of, and have access to, appropriate infrastructure and the knowledge necessary to manage and share the data effectively.

<sup>12</sup> Research Data Australia, [researchdata.ands.org.au](https://researchdata.ands.org.au/); Data.gov.au, [www.data.gov.au/](https://www.data.gov.au/); GBIF, [www.data.gbif.org/welcome.htm?sessionId=F8CFF7D2E2C7EF000606346E99A4D7D0](https://www.data.gbif.org/welcome.htm?sessionId=F8CFF7D2E2C7EF000606346E99A4D7D0); ALA, [www.ala.org.au/data-sets/](https://www.ala.org.au/data-sets/)



## Data storage and access

Having identified data that constitutes a national research asset, barriers to securing appropriate storage resources and managing the data to ensure discoverability, accessibility and useability often still remain. Such barriers go beyond simply identifying a physical location for the data and can be policy-based, regulatory, cultural or technical. Regimes for governing access to data may be subject to a complex range of statutory and other regulations that can vary widely across institutions, jurisdictions and nations. Some data can only be shared or used when appropriate mediated access arrangements are in place.

There is currently a lack of incentives—both within institutions and across government policies and programmes—for researchers to publish and share data, and for research organisations to agree on coordinated frameworks and policies for data storage that facilitates access. Funding programmes often engender a competitive attitude between researchers, disciplines and institutions, thereby limiting opportunities for researchers to share and collaborate across institutional and disciplinary boundaries.

There continues to be a need to encourage and facilitate data sharing within and across disciplines and to promote a collaborative culture across the sector. This includes progressing efforts towards an environment of open data for research in Australia.

## Meeting researchers' needs

Once nationally significant research data has been stored in suitable research data infrastructure to ensure appropriate future access, it is imperative that the tools and processes are in place so that the widest range of researchers can make the best use of the data.

Discipline-specific research data infrastructure facilities often evolve independently from each other, presenting a challenge for researchers who may wish to bring the data from different research areas together to resolve certain questions. For example, questions of coastal research might be more easily addressed if the different infrastructure dealing with data in the marine and terrestrial environments were brought or developed together or in tandem, with common standards.

Researcher data requirements and use will continue to evolve, particularly as researchers grapple with more complex and extensive data enabled by technical developments. With petascale computing for research now available in Australia, the challenge presented by an accelerating need for processing and the analytical capability to handle such data cannot be ignored. New technology continually creates new opportunities and new solutions—which, in turn, create new questions and challenges and mean that the ongoing evolution of the tools and processes available is vital. Ongoing engagement with the research sector is essential to ensure research data infrastructure continues to meet researcher needs.

Australian researchers also face the challenge of acquiring the skills they need to analyse research data, particularly given the growing size and complexity of this data. People with the skills needed to maintain and develop next-generation research data infrastructure are also critical.

## Opportunities

### Investment leverage

In any funding environment it is imperative that the investment we make in research infrastructure is optimised for efficiency and impact. Opportunities exist to continue to develop research infrastructure and to improve its coordination. Research infrastructure developments should be strategic and collaborative rather than competitive, and should avoid unnecessary duplication. Improving cross-sectoral and cross-capability collaboration and leverage, including solution convergence, may help to reduce investment volatility. The roles that institutions can and do play in contributing to a joined-up national research data infrastructure system—through their own institutional investment and as participants and co-contributors to national initiatives—should be reinforced and leveraged further.

The time horizon for investment in Australian research data infrastructure should be decadal/multi-decadal reflecting the scale and complexity of required development as well as the importance of longitudinal (time-series) research data, although multi-decadal planning would need to take into account the speed of technological change.

### Coordinated research data generation and management

The development and implementation of common best-practice approaches to research data collection, generation, aggregation and management present significant opportunities. These can be independent of programmes, institutions and funding sources. A shared vision is important, as well as high-level policies to support coordinated practice. Collaborative approaches between the research sector, government agencies, private industry, and citizen science have great potential to bring benefits across the board and, from the researcher's point of view, particularly to open access to the types of data for research that may not be readily available at present.

Mechanisms for data quality management are of particular importance, including guiding principles and policy, and shared data practice. These mechanisms include roles and responsibilities of data stewards, standard concept definitions and dictionaries, information on file formats and coding standards.

There are significant opportunities to be gained from co-location of integral data services by ensuring the data is readily accessible to the computational capability, the tools, the storage and the high-bandwidth networks needed to move the data.

### Enabling governance

Australia has a powerful legacy from investments in national research infrastructure. While there has already been important cross-capability collaboration, there are additional benefits that will accrue from establishment of structures to foster and govern more systematic coordination between capabilities. Research data infrastructure plans at capability level should include reference to cross-capability collaboration.

### Data storage and access

As noted, while there are challenges associated with current policy and regulatory frameworks for storing and accessing research data, these also present opportunities for research data infrastructure. Adopting common approaches across institutions can bring consistency in widespread open access to, as well as more extensive and productive use of, research data.

It is clear that over the next decade, as the practice of research becomes increasingly collaborative and data-intensive, approaches must optimise the use of data for research, including through enabling and promoting open access to data.

An overall framework of open access to research data holds significant potential to operate as an organising mechanism, guiding mutually beneficial institutional responses to some of the challenges outlined, including data storage and access, investment leverage, coordinated research data management, and enabling governance.

Globally, there are moves towards open access to the outputs of publicly funded research, for both publications and data. Policies have been introduced or are currently being implemented in the United Kingdom, the United States, Canada and the European Union, and have been adopted internationally by agencies such as the World Bank and UNESCO, and by philanthropic funding bodies like the Wellcome Trust (examples are in Appendix A).<sup>13</sup> In Australia, efforts by research funding bodies towards an open publication policy are welcome signs of a shift towards greater access to research findings and may eventually encompass research data as well.

A particularly significant international step towards open access to research data was a statement signed by the G8 Science Ministers in June 2013 (Appendix B), which proposes new areas for collaboration and agreement, including open scientific research data and improved access to the peer-reviewed, published results of scientific data.

In support of the principle regarding open scientific research data, the G8 stated that:

*To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.*<sup>14</sup>

There have also been moves by international funding bodies and programmes to adopt policies of open access to research data, which may have significant implications for research collaborations. These include the (United States) National Science Foundation<sup>15</sup>, and United States bodies subject to the directive of the Government Office of Science and Technology Policy<sup>16</sup> that federal agencies with a budget of \$100 million or more develop plans to make the results of their research publicly accessible, including datasets. A related development is the inclusion of an Open Research Data pilot programme in Horizon 2020, the European Union's Research and Innovation funding programme for 2014–20, which has allocated €24 billion to science and another €30 billion to research into major European concerns<sup>17</sup>.

Australian Government agencies and funding bodies are also regulating to improve access to publicly funded data. For example, the Australian Research Council (ARC) recently amended its rules relating to the management of data and publications arising from ARC-funded projects from 2014, with the objective 'to ensure the widest possible dissemination of the research supported by ARC funding, in

<sup>13</sup> For example, Wellcome Trust, 2010, *Policy on data management and sharing*, [www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm](http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm).

<sup>14</sup> G8 Science Ministers Statement London UK, 12 June 2013, [www.gov.uk/gove16rnment/publications/g8-science-ministers-statement-london-12-june-2013](http://www.gov.uk/gove16rnment/publications/g8-science-ministers-statement-london-12-june-2013)

<sup>15</sup> [www.nsf.gov/bfa/dias/policy/dmp.jsp](http://www.nsf.gov/bfa/dias/policy/dmp.jsp)

<sup>16</sup> [www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

<sup>17</sup> [www.coar-repositories.org/files/Horizon\\_2020\\_Open\\_Data\\_Pilot\\_20130703\\_final.pdf](http://www.coar-repositories.org/files/Horizon_2020_Open_Data_Pilot_20130703_final.pdf)

the most effective manner and at the earliest opportunity'. Projects must outline how data arising from an ARC-funded project has been made publicly available where appropriate. Moving beyond open publication, the ARC now 'strongly encourages' funded projects to deposit data arising from a project in an appropriate publicly accessible repository.<sup>18</sup>

It is timely, therefore, for Australia to consider supporting the set of principles on open scientific research data developed by the G8. Such a step would signal Australia's willingness to international partners to remain engaged on research data policy matters, and would initiate the development of an open access framework that best positions Australian researchers and research institutions to operate in the data-intensive research future.

Nationally and globally, Australia stands to benefit significantly from achieving an environment in which well-managed research data is made quickly and easily discoverable, accessible and re-useable. This type of environment can improve the efficiency with which research is carried out; improve the overall quality of research data through subjecting it to greater scrutiny; and increase the potential for collaboration around data with international and private sector partners, and, perhaps most importantly for smaller research organisations, ensure the costs associated with securing the data their researchers need are minimised.

To achieve the widely recognised benefits accrued by open access arrangements, a number of pressing considerations remain, and addressing these carefully but expeditiously should be a high priority. These include consideration of privacy, commercial or security issues; the ongoing development of metadata standards; institutional readiness; the development of researcher skills; and the development of incentives that will drive cultural change.

It is impossible to consider the uptake of open access policies without also ensuring the appropriate underpinning research data infrastructure is in place. Australia has positioned itself well to implement open access arrangements by investing in a range of infrastructures to support data creation and generation, management, storage, and dissemination. We now have a suite of state-of-the-art national capabilities that enable us to participate in the global move towards open access. In turn, this boosts Australia's ability to conduct outstanding research on an interconnected international stage, to collaborate internationally, and to attract the best researchers from around the world.

---

<sup>18</sup> Australian Research Council, 2012, ARC Open Access Policy, [www.arc.gov.au/pdf/ARC%20Open%20Access%20Policy\\_print\\_version.pdf](http://www.arc.gov.au/pdf/ARC%20Open%20Access%20Policy_print_version.pdf). For examples in funding rules for funding commencement in 2014, see Discovery Early Career Researcher Award [www.arc.gov.au/pdf/DECRA14/DE14\\_funding%20rules.pdf](http://www.arc.gov.au/pdf/DECRA14/DE14_funding%20rules.pdf) and Linkage Projects [www.arc.gov.au/pdf/LP14/LP14%20Funding%20Rules.pdf](http://www.arc.gov.au/pdf/LP14/LP14%20Funding%20Rules.pdf).



## 7. Recommendations

RDIC has developed the following recommendations as part of the Australian Research Data Infrastructure Strategy, in line with the framework proposed in Section 6. The recommendations are consistent with principles articulated in Chapter 5, and address challenges and opportunities identified in Section 6.

Collectively, these recommendations promote and achieve the outcomes stated elsewhere in this strategy (Section 5) and encapsulated in the framework diagram in Figure 2:

- sustained, priority research data collections and data generation and management infrastructures, at national, research domain and institutional levels
- an environment in which researchers can benefit from institutional and sector-wide data and infrastructure access arrangements that are clear and well-governed
- delivery of enhanced research outcomes in line with national priorities, as a result of better sharing and re-use of research data and data for research.

They are the proposed means through which future investments in research data infrastructure can be made and used more productively, and provide a framework of conditionality for any future investments in research data infrastructure.

These recommendations are directed to the Australian Government—particularly those in government responsible for science and research investments and strategy—to decision-makers, planners and implementers of research data infrastructure, and to its users. The recommendations will also assist parties involved in the collection and generation of research data and data that may be relevant to research, including at all levels of government.

The recommendation to establish a national research data infrastructure advisory committee (Recommendation 6) will ensure the advice to government is effectively prioritised, consistent with the prevailing environment.

### Sustained priority collection and infrastructure

For Australian researchers to have any capacity to undertake world-class research, Australia's current and future research data infrastructure requires sustained investment to ensure Australian researchers can undertake outstanding research on an international stage. National research data infrastructure is a basic necessity to enable researchers to address Australia's national challenges.

Significant research data collections—social, economic, environmental and others—are required. These collections are best developed—and in some cases, can only be developed—and maintained by the Australian research community. Internationally significant reference data collections are also vital to Australian research, and Australia can contribute significantly where it maintains a globally competitive advantage.

Peaks and troughs must be avoided in funding longitudinal and reference data collections, as they present a risk of lack of continuity for these collections. Such breaks in continuity reduce the value of these collections for research.



Data is a key national asset, and should be managed to maximise research return on investment over its life cycle. Regular review and evaluation should be undertaken to ensure that investments remain well positioned to address future challenges and opportunities. Australia's reliance on international data and our potential to contribute to such datasets must be recognised at all stages of infrastructure investment and development.

**Recommendation 1:** Recognise the ongoing value of Australia's research data assets, and safeguard prior and future investment in them by developing a sustained approach to investment in research data infrastructure.

**Recommendation 2:** Adopt a rolling decadal planning and prioritisation approach to investment in research data and research data infrastructure. Planning and prioritisation should reference the data infrastructures described in the *2011 Strategic Roadmap for Australian Research Infrastructure*.

**Recommendation 3:** Require best practice in discoverability, accessibility and usability of Australian research data from all future investments in research data infrastructure, in line with leading data management policies and procedures. Set minimum data management requirements for nationally funded research facilities, to be incorporated into funding guidelines and agreements.

**Recommendation 4:** Design national research data infrastructure investments to enhance and complement capability at the institutional and international level, encourage institutional commitment, and foster sustained institutional capacity in data-intensive research.

**Recommendation 5:** Coordinate research data infrastructure investments to drive collaboration between, and commitment from, research institutions, government agencies, industry participants and community stakeholders.

## Enabling governance and access arrangements

Research data assets need to be accessible and governed effectively, if we are to realise the full potential and value from Australia's investment in research data infrastructure. International progress toward open access to research data (open data) requires a parallel policy in Australia if our researchers are to build and retain a collaboration advantage (see Box 12 for example). Consequently, the development of our national response will need to be cultivated carefully.

Efficiency can be derived from enhanced organisation and coordination (including technical interoperability):

- within domains, such as disciplines and strategic research priority areas (through formal planning)
- among research data infrastructure investments (through coordination and joint governance)
- between research data infrastructure investments and broader government data infrastructure investments (through liaison).

Maximising access to, and re-use of, well-managed research data leads to a greater return on the investment in research and infrastructure as well as improved scientific integrity and new data-centric methods. Of growing importance is how researchers gain access to government or public data that may be relevant to their research. Encouragement from the research sector will be necessary if this is to occur, in conjunction with growing moves towards open government.

Such moves toward opening access can be promoted in part via research data infrastructure development and operation.

### Box 12: Terra Nova: Establishing a collaboration advantage for Australian researchers and policy makers

Terra Nova is an information hub for climate change adaptation research which simplifies the discovery, sharing and use of climate change adaptation information for all users. It is also a failsafe to ensure this vital information is preserved and made accessible. Terra Nova is an initiative of Griffith University, the Queensland Cyber Infrastructure Foundation (QCIF), and the Australian National Data Service (ANDS).

Terra Nova has great potential to bridge the gap between researchers and planners. For example, the site includes a visualisation tool that can be used as a standard interface to climate change models created for specific purposes. The tool not only makes it easy to see and understand climate change modelling data, it also allows data generated for one particular purpose to be re-used for many other purposes, and this re-use extends the reach of the original research investment in collecting the data.

Terra Nova works collaboratively with the Australian Government to act as the repository for research funded under the \$8 million National Resource Management (NRM) Climate Change Impacts and Adaptation Research initiative. Under this programme, nine project teams from research institutions are working with regional NRM organisations across Australia to deliver climate change information and provide guidance on the effective use of that information in resource management planning processes. This is an example of the role Terra Nova plays in helping to make research data and information more securely stored, readily discoverable, and frequently re-used.

QCloud, the Research Data Storage Infrastructure (RDSI) data storage node operated by QCIF, is offering the hundreds of terabytes of storage required to host Terra Nova's extensive and expanding data collections and to make them widely available to researchers, governments and other interested parties. The infrastructure provided by QCloud and RDSI makes projects such as Terra Nova possible. Without economical and straightforward access to these capital-intensive facilities, it would be easier for universities to put these projects off. Instead, the storage infrastructure provides Queensland and Australian researchers with a significant collaboration advantage through Terra Nova.

Source: Research Data Storage Infrastructure project, *RDSI Story—Adapting to climate change: Terra Nova* [www.rdsi.uq.edu.au/rdsi-story-terra-nova](http://www.rdsi.uq.edu.au/rdsi-story-terra-nova)

**Recommendation 6:** Establish a national research data infrastructure advisory committee to review, coordinate and provide coherence to the implementation of research data infrastructure investments, including through assisting national research data infrastructure facilities to generate and build partnerships and collaborations. The committee will have particular regard to the recommendations in this strategy in delivering its terms of reference.<sup>19</sup>

**Recommendation 7:** With reference to Recommendations 1 to 5 above, require funded capability areas under the *2011 Strategic Roadmap for Australian Research Infrastructure* to include research data infrastructure plans consistent with this strategy. These plans should include national data strategies to coordinate collection, generation, manipulation and management of research data

<sup>19</sup> Draft terms of reference are in Appendix D.

across institutions, programmes, sectors, and funding sources so as to maximise access, use and re-use.

**Recommendation 8:** Encourage and support custodians of government data collections to make sure that data is available, accessible and useable for research purposes, consistent with the principles of open government.

**Recommendation 9:** Ensure that research data infrastructure facilities are configured to enable appropriate access for data outputs for publicly funded research, consistent with the growing national and international policy momentum toward open access to research data.

**Recommendation 10:** Enable Australian researchers to access national and international research data, and data for research, through effective development and use of research infrastructure consistent with common and standard data approaches.

**Recommendation 11:** Encourage the adoption of appropriate national services that ensure open access to data for a broad spectrum of researchers, and require custodians of publicly funded research data to make that data available, accessible and useable.

## Delivery of enhanced research outcomes

Investments in analysis tools that enable effective, efficient use and re-use of data of lasting value and importance will assist in transforming Australian research outcomes, including addressing national and global research questions across disciplinary and jurisdictional boundaries.

A skilled workforce, including researchers who can develop, implement and use these tools, and skilled technical implementation and support staff, is essential. Building the data literacy of the Australian research community is an urgent priority to improve Australia's global competitiveness and ensure that Australia is not left behind in the emerging environment of data-driven and data-intensive research.

Australian researchers need to be provided with state-of-the-art tools to access, manipulate, analyse, simulate, model, share, and use research data and data for research effectively, irrespective of distance, discipline, research problem, institution, or engagement in international partnerships. These tools need to be open access, version controlled, documented and hosted via an appropriate forum, to ensure results are genuinely reproducible. Australian research communities should be able to access Australian and international research data with data tools that are created once but used many times. Access to the combination of data and tools used to process the data goes a long way to ensuring its reproducibility.

Researcher and research community-focused analysis tools, data products, virtual labs, and other such technical capability are expected outputs from investments in research data infrastructure. Similarly, suitably skilled research and support workforces are needed to develop, implement, operate and support the data infrastructure. The challenge of suitable career pathways for such skilled personnel must be addressed to enable an effective and efficient use of the infrastructure, as well as retention of valuable human capability and knowledge to support researchers.

Australia's data assets should be optimised to provide a collaboration advantage for Australian researchers and enable international researchers to work with Australian researchers on our national challenges.

**Recommendation 12:** Plan investments in research data tools strategically, with due consideration to existing investments, so that they are non-duplicative, and at a scale that is nationally and internationally relevant and can be supported.

**Recommendation 13:** Position research data infrastructure investments globally so as to provide Australian researchers with an environment in which they are seen as partners of choice in data-intensive research internationally.

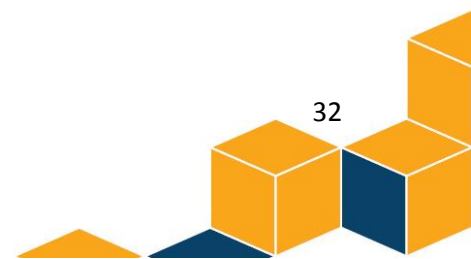
**Recommendation 14:** Ensure data services are flexible and adaptable, using standardised and open architectures, to help researchers explore research data assets effectively.

**Recommendation 15:** Urgently build the digital literacy of the research community through increased emphasis on use of research data infrastructure in research training, and links with higher education programmes to create a level of data competency in future generations of Australian scientists, researchers and academics that is globally competitive.

**Recommendation 16:** Provide incentives for researchers to contribute to and make use of effective research data infrastructure, by linking adoption to increased citations, emerging data impact metrics such as data citation, greater career recognition and further access to institutional support.

**Recommendation 17:** Build and maintain the human capacity—including that of support staff, technical staff and data specialists—to realise optimal benefit from investments in research data infrastructure. The competencies and skills required, the role of communities of practice, and career development paths should be emphasised. These efforts should occur in conjunction with and within the parameters of existing government activity to support research workforces.

**Recommendation 18:** Maintain and develop a coherent national infrastructure to aggregate and richly integrate information across domains, sectors and facilities, thus enabling researchers to navigate, explore and exploit Australia's evolving data assets.



## Appendix A: General references

Accessibility, sustainability, excellence: how to expand access to research publications, Research Information Network, Report of the working group on expanding access to published research findings (Finch Report)	<a href="http://www.researchinfonet.org/publish/finch/">www.researchinfonet.org/publish/finch/</a>
ARC Open Access Policy	<a href="http://www.arc.gov.au/applicants/open_access.htm">www.arc.gov.au/applicants/open_access.htm</a>
Australian Public Service Information and Communications Strategy 2012–2015	<a href="http://www.finance.gov.au/publications/ict_strategy_2012_2015/">www.finance.gov.au/publications/ict_strategy_2012_2015/</a>
Manyika, J, Chui, M, Brown, B, Bughin, J, Dobbs, R, Roxburgh, C & Hung Byers, A 2011, <i>Big Data: The next frontier for innovation, competition and productivity</i> , McKinsey & Company	<a href="http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation">www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation</a>
Digital Agenda for Europe (DAE) The Digital Agenda website, European Commission	<a href="http://www.eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:HTML">www.eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:HTML</a>
Directive on the re-use of public sector information (PSI Directive), The Digital Agenda website, European Commission	<a href="http://www.ec.europa.eu/information_society/policy/psi/rules/eu/index_en.htm">www.ec.europa.eu/information_society/policy/psi/rules/eu/index_en.htm</a>
European Open Government Data Initiative (OGDI)	<a href="http://www.epsiplatform.eu/content/microsoft-announces-open-government-data-initiative-ogdi">www.epsiplatform.eu/content/microsoft-announces-open-government-data-initiative-ogdi</a>
Global Earth Observation system of Systems (GEOSS) Data sharing principles, Group on Earth Observations website	<a href="http://www.earthobservations.org/geoss_ta_da_tar.shtml">www.earthobservations.org/geoss_ta_da_tar.shtml</a>
Horizon 2020—Outline of pilot for Open Research Data, Horizon 2020 website, European Commission	<a href="http://www.ec.europa.eu/programmes/horizon2020/en/what-horizon-2020">www.ec.europa.eu/programmes/horizon2020/en/what-horizon-2020</a>
O'Reilly, T, Steele, J, Loukides, M & Hill, C 2012, <i>Solving the Wanamaker problem for health care</i> O'Reilly website	<a href="http://www.strata.oreilly.com/">www.strata.oreilly.com/</a>
Ad-hoc Strategic Coordinating Committee on Information and Data (SCCID Report), Interim Report to the ICSU Committee on Scientific Planning and Review, International Council for Science website	<a href="http://www.icsu.org/publications/reports-and-reviews/strategic-coordinating-committee-on-information-and-data-report/">www.icsu.org/publications/reports-and-reviews/strategic-coordinating-committee-on-information-and-data-report/</a> Other reports on data and information: <a href="http://www.icsu.org/publications">www.icsu.org/publications</a>
<i>Mapping the data landscape: report of the 2011 Canadian Research Data Summit</i> , Produced by the Research Data Strategy Working Group, Research Data Canada website	<a href="http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/events/data_summit_2011.html">http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/events/data_summit_2011.html</a>
National Science Foundation (US) Data Sharing Policy, The National Science Foundation website	<a href="http://www.nsf.gov/bfa/dias/policy/dmp.jsp">www.nsf.gov/bfa/dias/policy/dmp.jsp</a>
NHMRC Policy statement on Open Access,	<a href="http://www.nhmrc.gov.au/grants/policy/dissemination-">www.nhmrc.gov.au/grants/policy/dissemination-</a>

Australian Government, National Health and Medical Research Council website	<a href="#">research-findings</a>
OECD Declaration on Access to Research Data from Public Funding—30 January 2004—C(2004)31/REV1, Decisions, Recommendations and other Instruments of the Organisation for Economic Co-operation and Development website	<a href="http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157">http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157</a>
OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, OECD website	<a href="http://www.oecd.org/internet/interneteconomy/oecdguidelinesonthe protection of privacy and transborder flows of personal data.htm">www.oecd.org/internet/interneteconomy/oecdguidelinesonthe protection of privacy and transborder flows of personal data.htm</a>
OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD website	<a href="http://www.oecd.org/science/scienceandtechnologypolicy">www.oecd.org/science/scienceandtechnologypolicy</a>
Open data strategy for Europe, Europa.eu website	<a href="http://www.europa.eu/rapid/press-release_IP-11-1524_en.htm?locale=en">www.europa.eu/rapid/press-release_IP-11-1524_en.htm?locale=en</a> <a href="http://www.europa.eu/rapid/press-release_MEMO-11-891_en.htm">www.europa.eu/rapid/press-release_MEMO-11-891_en.htm</a>
Recommendation for enhanced access and more effective use of public sector information (PSI), OECD website	<a href="http://www.oecd.org/sti/oecdrecommendationonpublicsectorinformationpsi.htm">www.oecd.org/sti/oecdrecommendationonpublicsectorinformationpsi.htm</a>
Research Councils UK (RCUK) Policy on access to research outputs, Research Councils UK website	<a href="http://www.rcuk.ac.uk">www.rcuk.ac.uk</a>
<i>Riding the wave: How Europe can gain from the rising tide of scientific data</i> , Final report of the High Level Expert Group on Scientific Data, A submission to the European Commission)	<a href="#">Riding the wave</a>
Strategic report for translational systems biology and bioinformatics in the European Union, (EU 7th Framework Vision) INBIOMED vision project, European Commission	<a href="http://www.inbiomedvision.eu/index.html">www.inbiomedvision.eu/index.html</a>
Australian Government 2013, <i>Strategic review of health and medical research—better health through research</i> (McKeon Review)	<a href="http://www.mckeonreview.org.au/">www.mckeonreview.org.au/</a>
The ANDS Technical Working Group 2007, <i>Towards an Australian Data Commons</i> , Australian Government	<a href="http://www.ands.org.au/">www.ands.org.au/</a>
UK Government response to the Finch Group Report, ( <i>Accessibility, sustainability, excellence: how to expand access to research publications</i> ) GOV.UK website	<a href="http://www.bis.gov.uk/assets/biscore/science/docs/l/12-975-letter-government-response-to-finch-report-research-publications.pdf">www.bis.gov.uk/assets/biscore/science/docs/l/12-975-letter-government-response-to-finch-report-research-publications.pdf</a>
HM Government 2012, UK Open Data White Paper: <i>Unleashing the Potential</i>	<a href="#">Open Data White Paper</a>
US Office of Science and Technology Policy, Memorandum on Increasing Access to the Results of Federally Funded Scientific Research,	<a href="http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf">www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf</a>

The White House website	
The Royal Society 2012, <i>Science as an open enterprise</i> , The Royal Society Science Centre report 02/12. The Royal Society website includes terms of reference, case studies of data use and data repositories	<a href="http://www.royalsociety.org/policy/projects/science-public-enterprise/report/">www.royalsociety.org/policy/projects/science-public-enterprise/report/</a>
Wellcome Trust Open access policy: a position statement in support of open and unrestricted access to published research, Wellcome Trust website	<a href="http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm">www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm</a>  2012 <a href="#">update</a>
Wellcome Trust Policy on data management and sharing, Wellcome Trust website	<a href="http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm">www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm</a>
White House Open Government Directive, The White House website	<a href="#">White House Open Government Directive</a>
White House Memorandum on Transparency and Open Government, The White House website	<a href="#">White House Memorandum on Transparency and Open Government</a>

## Appendix B: Addendum on open data

*Joint statement from the first meeting of G8 Science Ministers in London on 12 June 2013*

### Introduction

We, the G8 Science Ministers met in London on Wednesday 12 June with Presidents of our respective national science academies, as part of the UK's G8 Presidency. At this unique meeting we discussed how our nations could lead efforts to improve the transparency, coherence and coordination of the global scientific research enterprise in order to address global challenges and maximise the social and economic benefits of research.

So today, recognising the role that science has to play in securing present and future sustainable growth, we approved a statement which proposes to the G8 for consideration new areas for collaboration and agreement on global challenges, global research infrastructure, open scientific research data, and increasing access to the peer-reviewed, published results of scientific research.

### 1. Global Challenges

Global challenges refer to humanity's most pressing concerns, transcending national boundaries and posing significant threats to societies and ecosystems. They require international cooperation because of the nature and magnitude of their potential consequences. Science plays a pivotal role in addressing these challenges. We discussed new and emerging global challenges that require concerted G8 cooperation amongst the science community. We highlighted the importance of taking a cross-disciplinary and cross-sectoral approach, across the social, human, natural, life and environmental sciences. We accepted that there was value in enhanced collaboration to share insights from national activities to systematically identify and examine potential threats and challenges that can be addressed through science.

We acknowledged that there is a broad range of global challenges requiring our attention in the near future. We highlighted in particular, urbanisation, pollution, energy security, climate change, biodiversity, ocean acidification, youth unemployment, inequality, how to translate basic science to personalised and regenerative medicine, the ageing population and neurodegenerative diseases during our discussions—whilst recognising that this was not a definitive list.

At the meeting we focused on antimicrobial (i.e. anti-viral, anti-bacterial, anti-fungal and anti-parasite) drug resistance as a major health security challenge of the twenty first century. We decided to act concertedly on developing the scientific input necessary to reduce antimicrobial resistance working with existing agencies such as the World Health Organisation and by taking into consideration other activities, to:

- i. preserve the efficacy of existing antimicrobial agents, in part by avoiding misuse and optimising prescribing practices in our respective countries (for human, veterinary and aquaculture use);
- ii. prevent the emergence of antimicrobial/drug resistance, in part through developing rapid diagnostics to inform antimicrobial drugs use;
- iii. support the development of new antimicrobial agents and interventions to treat microbial infection, in part by stimulating the antimicrobial drug research and development pipeline;
- iv. support the development of new diagnostics to improve early diagnosis of antimicrobial resistant infections and improve treatment efficacy;



- v. support international cooperation and sharing of surveillance data to improve global understanding of the spread of antimicrobial drugs resistance;
- vi. support theoretical and applied research to better understand the origin, spread, evolution and development of resistance in microorganisms (including viruses, bacteria and parasites) and the role of the innate immune system.

## **2. Global Research Infrastructure**

Research infrastructures (RI) are key elements in research and innovation policies. In some cases, their complexity as well as high development, construction and operation costs, requirement for a critical mass of highly qualified human resources, or simply the global nature of the scientific challenge addressed, makes it impossible for one country or region alone to build and operate these facilities. In such cases it becomes crucial to make concerted efforts at the international level for the realisation of 'global research infrastructures' (GRI). We recognise the potential for increased international cooperation on global research infrastructures and the benefits that arise from a shared outlook on frontier research and collaboration and good governance.

We recognise the work of the Group of Senior Officials (GSO) formed after the G8 Ministerial meeting in Okinawa in 2008 to take stock and explore cooperation on GRIs.

- i. We have decided to adopt the Framework for GRIs decided by the GSO as the principles and reference terms under which G8 countries consider cooperation on GRIs, and encourage other nations to adopt them.
- ii. We approve a new mandate for the GSO to enable it to fulfil the areas of its original mandate that have not yet been addressed including to:
  - Promote the Framework and continue to exchange information on potential future research infrastructure that may present opportunities for international collaboration, noting especially the needs of the Global Challenges; and
  - Share information on national research infrastructure priorities and prioritization processes; identify areas of potential benefit that could be achieved through sharing of best practices.
  - Create a representative list of GRIs open to global cooperation of interest to new partners.
  - We invite the GSO to report in 2015 on their progress.

## **3. Open Scientific Research Data**

Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and its results communicated. It can provide society with the necessary information to solve global challenges. We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of scientific research are as widely available as practical, enable transparency in science and engage the public in the scientific process. We have decided to support the set of principles for open scientific research data outlined below as a basis for further discussions.

- i. To the greatest extent and with the fewest constraints possible publicly funded scientific research data should be open, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.

- ii. Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.
- iii. To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.
- iv. To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including recognition of researchers fulfilling these principles, and appropriate digital infrastructure.

We decide to build on the existing work to coordinate and enable international data collaboration.

#### **4. Expanding Access to Scientific Research Results**

We recognise that effective global scientific research and public understanding of science and commercial innovation by enterprises is supported by free and rapid public access to published, publicly funded research. The generation, sharing and exploitation of scientific knowledge are integral to the creation of wealth and the enhancement of our quality of life. We recognise that G8 nations have an important opportunity and responsibility to promote policies that increase access to the results of publicly funded research results to spur scientific discovery, enable better international collaboration and coordination of research, enhance the engagement of society and help support economic prosperity.

- i. We endorse the principle that increasing access to the peer-reviewed, published results of publicly funded published research will accelerate research, drive innovation, and benefit the economy.
- ii. We recognise the importance of peer review and the valuable role played by publishers, including Learned Societies. Increasing free access to peer-reviewed, published research results will require sustainable solutions.
- iii. We recognise the potential benefits of immediate global access to and unrestricted use of published peer-reviewed, publicly funded research results in line with the necessity of IP protection.
- iv. We recognise that there are different routes to open access (green, gold and other innovative models) which need to be explored and potentially developed in a complementary way.
- v. We recognise that the long-term preservation of the increasingly digitized body of scientific publications and data requires careful consideration at the national and international levels to ensure that the scientific results of our time will be available also to future generations.
- vi. We recognise that further work is required to optimise increasing public access to peer-reviewed, publicly funded published research and its underlying data and that international coordination and cooperation will provide for an efficient transition to “open access”.
- vii. We share the intention, therefore, to continue our cooperative efforts and will consider how best to address the global promotion of increasing public access to the results of publicly funded published research including to peer-reviewed published research and research data.

We recognise the role of our national science academies and research organisations across these important agendas, working regionally, nationally and globally through their respective networks.

Signed by G8 Science Ministers 12 June 2013

**David Willetts**

Minister for Universities and Science UK

**Dmitry Livanov**

Minister of Education and Science of the Russian Federation

**Georg Schütte**

State Secretary for Education and Research, Germany

**Yuko Harayama**

Executive member, Council for Science and Technology Policy, Japan

**Maria Chiara Carrozza**

Minister of Education, University and Research, Italy

**Gary Goodyear**

Minister of State for Science & Technology, Canada

**Geneviève Fioraso**

Minister for Higher Education and Research, France

**Patricia Falcone**

Associate Director of the Office of Science and Technology Policy, USA

**Máire Geoghegan-Quinn**

European Commissioner for Research, Innovation and Science

[www.gov.uk/government/news/g8-science-ministers-statement](http://www.gov.uk/government/news/g8-science-ministers-statement)

## Appendix C: Principles from the Strategic Framework for Research Infrastructure Investment

### *Extract of principles from the Strategic Framework for Research Infrastructure Investment*

#### Definition of Research Infrastructure

Research infrastructure comprises the assets, facilities and services which support research across the innovation system and which maintain the capacity of researchers to undertake excellent research and deliver innovation outcomes.

#### Principles for Research Infrastructure Investment

##### Continuity of Funding

- Research infrastructure funding programs should be ongoing and predictable, to achieve optimal use of funds.
- Infrastructure that continues to be a priority should be able to access funding for ongoing operations.

##### *Guiding considerations*

- Ongoing and predictable funding programs support a more strategic, collaborative and planned approach to research infrastructure investment.
- Ongoing operational funding for priority national and landmark research infrastructure assists in maximising the benefit from the original investment.

##### Holistic Funding

- Funding required to support research infrastructure will vary between elements, including capital costs, governance, skilled technical support staff and operations and maintenance. Support should be available to cover these key elements.
- Funding programs should allow some funding for project development costs, either for a facilitation-based process or for project development and scoping activities.
- In the context where not all national and landmark infrastructure would necessarily be replaced, depreciation for these facilities should not be funded by Australian Government funding programs.

##### *Guiding considerations*

- The ability to invest in human capital and operating costs results in superior service delivery and more efficient, productive and viable research infrastructure facilities.
- Funding for specialist staff assists in developing and maintaining the highly skilled workforce required for the efficient operation of sophisticated facilities.
- Rigorous, consultative project planning is a key input to developing excellent research infrastructure facilities, particularly at the national and landmark scale.

## Prioritisation

- Any proposed research infrastructure investment should align with and support Australia's research, innovation and infrastructure priorities.
- Funding for Australia's research infrastructure should focus on areas where Australia:
  - undertakes world-leading research or innovation;
  - has demonstrated a particular strength in international terms; or
  - has reasons to seek to strengthen capacity in an area of research or innovation.
- Prioritisation of investment in research infrastructure is necessary to ensure appropriate, effective and efficient investment; to support strategic decision making with regard to national and landmark infrastructure; and to ensure Australia achieves the maximum outcome for the money invested.
- Processes for funding research infrastructure should be transparent, provide effective use of funds and clearly target intended outcomes.

### *Guiding considerations*

- With finite resources, Australia needs to choose where to target its investments in research infrastructure.
- Australia needs to consider its priorities in both a national and an international context.
- Transparent processes to determine priorities will lead to better informed and more widely supported outcomes.
- The strategic identification of capabilities and priorities should be through a consultative roadmapping process every three years.

## Excellence in research infrastructure

- Proposals for investment in all scales of research infrastructure should be evaluated on the basis of their ability to create excellent infrastructure.
- Governance structures should be robust and fit for purpose to ensure the delivery of excellence in research infrastructure.

### *Guiding considerations*

- Excellence in research infrastructure is essential to ensuring Australia is able to continue to compete internationally and contributes to a strong innovation system.

## Collaboration

- Funding should favour investments that demonstrate collaborative approaches for the creation and development of research infrastructure and that foster and facilitate a collaborative research culture.

### *Guiding considerations*

- Collaboration is a key driver of innovation and is critical to ensuring the research community can deliver the outcomes Australia needs.

- There are often economic and efficiency benefits from taking a collaborative approach to establishing and operating research infrastructure.

### **Co-investment**

- Co-investment in research infrastructure is desirable as it demonstrates a commitment by the investing party/ies to the project. Any program requirements for co-investment should be flexible to leverage maximum support.

#### *Guiding considerations*

- Flexibility and transparency in co-investment requirements can lead to greater overall leverage and improves the ability of States and Territories to coordinate support for research infrastructure with the Australian Government.
- Opportunities for industry co-investment in research infrastructure facilities should be clear and encouraged as a basis for closer research collaboration.

### **Access and Pricing for Australian-based infrastructure**

- Research infrastructure at the national and landmark scale should be made widely accessible to publicly funded researchers.
- Research infrastructure at the local scale should be made accessible to the extent possible in order to maximise use and support collaboration between institutions.
- Pricing policies for research infrastructure should be clear and transparent and allow for flexibility in the charging model, while still maximising the public benefit.
- Access to and pricing of finite research infrastructure resources should be based on a combination of factors including merit, co-investment, the role of the host institution, opportunities for early career researchers, and supporting collaborative research.

#### *Guiding considerations*

- An effective access regime ensures that research infrastructure is put to optimum use and fosters collaboration both nationally and internationally.
- An effective pricing policy for publicly funded research infrastructure ensures that meritorious research is not priced out of the market.
- Clear and transparent pricing policies allow for access costs to be built into research funding proposals.

### **Access to overseas-based infrastructure**

- Research infrastructure funding programs should consider Australian membership of, or contribution to the construction of, overseas facilities as the development of infrastructure in Australia is not always the most cost effective solution to providing research infrastructure.
- Research funding programs should consider requests for funding Australian researcher access to overseas facilities.
- Where possible Australian research infrastructure facilities should be encouraged to provide access to International researchers to foster international links and collaborations and build local skills.

### *Guiding considerations*

- Funding access to overseas-based research infrastructure ensures Australian researchers can utilise the best infrastructure available and furthers Australia's engagement with the global research community.

### **Evaluation and Monitoring**

- Research infrastructure funding programs should incorporate procedures for regular and rigorous monitoring and evaluation to ensure the effective use of public funds.

### *Guiding considerations*

- Evaluation and monitoring is essential to determine whether the research infrastructure has delivered its desired outcomes and achieved its objectives over the short and medium term, as well as over its whole life-cycle.
- Consideration of whether the research infrastructure continues to be a national priority is assisted through rigorous evaluation.

## Appendix D: Recommendation 6—Proposed draft terms of reference

### Draft terms of reference for an Australian research data infrastructure advisory committee

#### Background

The proposal for a new committee focused on national research data infrastructure arises in the Australian Research Data Infrastructure Strategy report of the Research Data Infrastructure Committee (RDIC) established by the then Department of Industry, Innovation, Science, Research and Tertiary Education in 2012. RDIC drew on the combined experience of builders and operators of Australia's national research data infrastructure over the last decade.

The specific recommendation of RDIC is as follows:

#### **Recommendation 6:**

Establish a national research data infrastructure advisory committee to review, coordinate, and provide coherence to implementation of research data infrastructure investments, including through assisting national research data infrastructure facilities to generate and build partnerships and collaborations. The committee will have particular regard to the recommendations in this strategy in delivering its terms of reference.

The deliberations of such a committee would take a cross-domain, as well as cross-institutional, approach to research data infrastructure, and include an international perspective.

#### Terms of reference

##### *Purpose*

The proposed committee will:

- provide advice on coordination and alignment of investments and infrastructure funded within the science and research portfolio for the purposes of research data collection, organisation and use
- champion research data infrastructure policy and implementation to improve the quality and quantity of data available for research, including data developed and managed by other portfolios and industry
- provide a point of contact within the research domain for strategic activities related to data and related infrastructure across government, such as the implementation of 'open data' policies and strategies to improve digital skills and literacy.

##### *Functions*

The committee will:

- advise on:
  - coordination of research data infrastructure
  - mechanisms to improve the long term availability, accessibility and usability of research data



- the further development of Australia's research data infrastructure in the global context
- identify:
  - partnerships and other arrangements that would improve research data infrastructure and data outcomes
  - capabilities where additional data investments may be directed
  - strategies to encourage the building and maintenance of human capacity
- monitor and report on progress and outcomes of efforts to improve national, collaborative research data infrastructure.

### *Composition*

The committee is expected to include a balance of relevant views and experience drawn from the research sector and across government and industry.

Areas of expertise for the committee to have or to have access to include:

- data infrastructure policy and planning across research, government and industry
- use of data infrastructure across research domains and sectors
- data and data technology
- operation of research data infrastructure, covering data collection and generation, organisation and use.

The chair of the committee would be appointed by the department and may be a member of the body the committee reports to.

### *Issues to be addressed*

The challenges posed by data and the resulting requirements addressed by the committee are long term and can be expected to endure for the foreseeable future.

- Overall progress should be reported on a regular basis and the committee arrangements, including composition, should be reviewed every three years.
- Research data infrastructure has and will continue to have system-wide impact and long-term effects.
- The committee should report as part of the most senior advisory structure supported by the science and research portfolio with oversight of research and research infrastructure.

## Appendix E: Glossary

**2011 Strategic Roadmap for Australian Research Infrastructure:** Articulates the priority research infrastructure areas of a national scale (capability areas) to develop Australia's research capacity and enhance research outcomes over the subsequent five to 10 years.

**Australian Research Data Commons:** The result of an Australian Government allocation to the Australian National Data Service (ANDS) \$48 million from the Super Science Initiative to create and develop an Australian Research Data Commons research infrastructure to support the discovery of, and access to, research data held in Australian universities, publicly funded research agencies and government organisations for the use of research.

**Autonomous profiling floats:** A development in oceanography that enables global broadscale ocean observations of temperature, salinity, velocity, and additional variables; these small ocean-borne robotic probes float as deep as 2 kilometres, providing real-time data for use in climate, weather, oceanographic and fisheries research.

**Big data:** Data collections so large and complex that they are difficult to process using traditional data processing applications. Challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualisation.

**Capability (NCRIS):** 16 capability areas (priorities for investment in national-scale research infrastructure) identified in the 2006 NCRIS Roadmap by the National Collaborative Research Infrastructure Strategy (NCRIS) Committee. In 2008 and 2011, the capability areas were reviewed and refined, and new areas were identified. The *2011 Strategic Roadmap for Australian Research Infrastructure* identifies current capability areas.

**Characterisation (NCRIS):** Facilities and equipment designed to characterise the basic/fundamental physical, chemical and structural attributes of matter including (as examples) neutron scattering, X-rays, microscopy.

**Cloud computing (or the cloud):** Used to describe a number of different types of computing concepts that involve a large number of computers connected through a real-time communication network (typically the internet) and usually with the advantage of scalability (of service, storage etc).

**Collection:** A flexible concept for any groupings (for example, aggregations) of objects or digital data (for example, 'a data collection or dataset'. Collection can also refer to: the act of bringing together data with intent to create a value-added aggregation (for example, 'the creation, collection, manipulation, etc of data'); or to data capture (for example, 'automated data collection from sensors').

**Curation:** Curation involves ways of organising, displaying, and repurposing preserved data.

**Data:** Facts and statistics including measurements, observations, images and numbers generated and collected together for reference or analysis.

**Data cube:** A three- (or higher) dimensional array of values, commonly used to describe a time series of image data.

**Data driven (research):** A concept where data obtained from experiments or trials leads to the development of knowledge and theories.

**Data for research:** Data which is generated or collected by government agencies and private enterprises which is potentially useful for researchers.

**Data intensive research:** Loosely, used to describe a class of research, including but not limited to **eResearch**, where data is one of the primary elements; often involves large, complex or multiple datasets.

**Data management:** A process whereby data is organised to conduct research efficiently and to provide a foundation for the retention and continuity of access to the data for as long as required.

**Data rich (research):** Usually refers to research involving or requiring many and/or large and/or complex datasets.

**Data storage (computer):** Technology consisting of computer components and recording media used to retain digital data.

**Data transfer (digital):** The movement of digital data via media including as examples, copper wire, optical fibre, wireless communication and storage media.

**Digital object identifiers (DOI):** Character string (a digital identifier) which uniquely identifies an object such as an electronic document or dataset. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for an object is permanent, whereas its location and other metadata may change. DOIs are used to facilitate continuity of access and citation tracking of data.

**Discoverability (data):** Making data discoverable by adding interpretable metadata to increase data reuse and decrease unnecessary duplication.

**eResearch:** eResearch is the use of advanced ICT to support existing and new forms of research.

**Flux tower:** Flux towers measure carbon dioxide, methane, and carbon monoxide, providing information specific to a single ecosystem type or condition. They can be used to measure net ecosystem carbon dioxide exchange. Data from flux towers located in contrasting environments can increase understanding of the effects of different activities or vegetation gradients, and help test models of carbon exchange.

**Generation (data):** Functions, routes and procedures by which data come into existence, usually as part of a workflow by which data ultimately reach a database.

**High-performance computer:** See **Supercomputer**.

**Integration (data):** Both a concept and analytical techniques whereby heterogeneous data can be combined and analysed meaningfully.

**Interoperability:** The ability of diverse systems and organizations to work together (inter-operate), allowing for information exchange.

**Metadata:** Data about data. It falls into a number of categories, which are functionally quite different. Structural metadata refers to the design and specification of data structures, and might be referred to as 'data about the containers of data', like identifiers. Descriptive metadata refers to details about the content of data: 'data about the content of data', for instance, the type of data usually found in library catalogues. Other categories of metadata might include administrative metadata, preservation metadata, provenance metadata, rights management metadata, etc.

**NCRIS:** Through the National Collaborative Research Infrastructure Strategy (NCRIS), the Australian Government provided \$542 million from 2004–5 to 2010–11 for national-scale collaborative research infrastructure. In its May 2013 Budget, the Government announced additional NCRIS funding of \$185.9 million over two years (2013–14 to 2014–15) to support the operation and maintenance of the most critical projects established under NCRIS and the Super Science Initiative.

**Open access:** The ideal, concept and practice of providing unrestricted access via the Internet to peer-reviewed scholarly research publications often based on the principle that publicly funded information should be publicly accessible. Initially applied only to research publications, the term is now also commonly used in the research world for data outputs of research, where data in scope are easily discoverable and re-usable, free, machine-readable, transformable, and based on open standards.

**Petascale:** A computer system capable of reaching performance in excess of one petaflops, i.e. one quadrillion Floating Point Operations Per Second (FLOPS).

**Polynya:** An area of open water surrounded by sea ice.

**Publicly funded research:** Research partly or wholly funded with public money.

**Publicly Funded Research Agencies (PFRAs) (Australia):** A subset of the research agencies, including CSIRO, the Australian Nuclear Science and Technology Organisation, and the Australian Institute of Marine Science (AIMS).

**Research data:** Data generated by, or for, the research process.

**Research Data Alliance:** An international collaboration that aims to accelerate and facilitate the sharing and exchange of research data. Australia, the United States and the European Union are foundation members. Australia's participation is facilitated by ANDS.

**Research data infrastructure:** A range of facilities, equipment, or tools that serve research through data generation, data manipulation, and data access. These facilities, equipment and tools include the data itself. Underpinning infrastructure includes eResearch infrastructure (for instance, data storage or tools) and data collecting and generating infrastructure that encompass large or systemic research infrastructure installations (such as high-performance computers, telescopes and marine observation systems, among others).

**Research fabric:** A comprehensive investment framework as conceived in NRIP that underpins Australia's basic and applied research capacity across the disciplines, while allowing research investment to be focused in a strategic way that addresses national challenges and contributes to increased productivity.

**Research institutions:** Terms commonly applied to an institution where research is conducted as one of its primary functions; includes, as examples, CSIRO, medical research institutes, and universities.

**Rich connection (data):** The links to contextual information that make data more valuable (such as re-use information, grant information, researcher information, location information, etc).

**Standards (data/metadata):** Open, documented, common and interoperable schemas, protocols and formats for data and metadata.

**Super Science Initiative:** In the May 2009 Budget, the Australian Government announced \$1.1 billion for the Super Science Initiative, funded through the Education Investment Fund, to build new large-scale collaborative research infrastructure.

**Supercomputer:** A computer at the frontline of contemporary processing capacity; in particular, speed of calculation.

**Tools/smart tools:** Software applications (for example, for manipulating data) in the context of a research (or eResearch) problem (for example, visualisation, integration, annotation, etc).

**Virtual laboratories:** Generic term to describe linked research facilities, data repositories and computational tools, often using cloud technology to maintain links.

## Appendix F: Research Data Infrastructure Committee membership, role and terms of reference

### Background

Australian Government investment in research infrastructure to date, most recently through the Super Science Initiative and the National Collaborative Research Infrastructure Strategy (NCRIS), has led to significant data generating capabilities across the research spectrum.

In parallel with these investments, research institutions and other bodies, including state governments, have themselves been making substantial investments in research data and facilities.

The rate of data growth produced by these capabilities makes continued efforts to develop and coordinate Australia's capacity to manage, store, share and access data within and across research disciplines integral to ensuring Australia's ongoing research success.

As noted in the *2011 Strategic Roadmap for Australian Research Infrastructure* (2011 Roadmap), recent initiatives such as NCRIS and Super Science have positioned Australia well globally, but there continues to be a need to support a broader approach towards the creation, management, storage and re-use of research data, particularly as Australian and international research data outputs continue to grow to monumental size and complexity.

### Overview of committee

Recognising the elevated importance of effective infrastructure to support research data management, including as articulated in the 2011 Roadmap, the Department of Industry, Innovation, Science, Research and Tertiary Education<sup>20</sup> is establishing a Research Data Infrastructure Committee (the committee).

The committee will undertake a review and analysis of the current national research data landscape, to form the basis of advice to the Australian Government regarding how to optimise existing and future investments in research data infrastructure.

In this context, the committee will provide advice regarding high-level, strategic issues relating to data management, such as governance, the relative roles of government, institutions and research facilities, and how to manage these roles effectively to encourage a collaborative approach to research data management in Australia.

This advice is expected to inform future deliberations of the Australian Government's Australian Research Committee (ARCom) in relation to the further development of the National Research Investment Plan.

### Composition

The Committee will comprise individuals with specialist knowledge of current national research data investments both at a high level, for example the Australian eResearch Infrastructure Council (AeRIC) and the Australian National Data Service (ANDS); and at the level of data-intensive research

---

<sup>20</sup> Dissolved 25 March 2013

infrastructures, for example the Integrated Marine Observing System (IMOS) and the Terrestrial Ecosystem Research Network (TERN). The proposed members below have been identified along these lines.

## Chair

Dr Ron Sandland BSc (Hons) (University of Sydney), PhD (University of NSW), FTSE, AM

Ron Sandland was previously CSIRO's Deputy Chief Executive. He joined CSIRO's Division of Mathematics and Statistics in 1969 and became chief of the division in 1988, later managing its merger with the Division of Information Technology into what then became CSIRO Mathematical and Information Sciences. In 1999 he became Deputy Chief Executive of CSIRO and led the Flagship Initiative. This involved initially six major cross-disciplinary research programmes and was aimed at addressing problems of a national priority. Dr Sandland's research interests were in applying statistics to solving challenging real-world problems. He was made an Honorary Life Member of the Statistical Society of Australia in 1998 and won the CSIRO Medal for Lifetime Achievement in 2006. He is a Fellow of the Australian Academy of Technological Sciences and Engineering, and was made a member of the Order of Australia in 2007. Dr Sandland is currently Chair of the Australian National Data Service Steering Committee, the boards of the Australian Centre of Excellence for Risk Analysis and of the Australian Mathematical Sciences Institute and member of the Australian eResearch Infrastructure Council (AeRIC).

## Membership

The individuals proposed below may nominate alternative representatives. It is expected that consultation with the broader research community will occur during the development of the final advice to government.

Name	Organisation
<b>Dr Ron Sandland (Chair)</b>	<b>Australian eResearch Infrastructure Council (AeRIC)</b>
Dr Rhys Francis	Australian eResearch Infrastructure Council (AeRIC)
Professor Gerard Goggin	Australian eResearch Infrastructure Council (AeRIC)
Professor Andy Pitman	Australian eResearch Infrastructure Council (AeRIC) ARC Centre of Excellence for Climate System Science
Professor Attila Brungs	Chair, 2011 eResearch Expert Working Group
Dr Ross Wilkinson	Australian National Data Service (ANDS)
Dr Adrian Burton	Australian National Data Service (ANDS)
Ms Clare McLaughlin	Department of Industry, Innovation, Climate Change, Science, Research and Tertiary Education (DIICCS RTE)
Mr Warwick McDonald	Bureau of Meteorology (BoM)
Mr Peter Hicks	Research Data Storage Infrastructure (RDSI)
Dr Stuart Minchin	Geoscience Australia (GA)

Mr Tim Moltmann	Integrated Marine Observing System (IMOS)
Dr John La Salle	Atlas of Living Australia (ALA)
Professor Tim Clancy	Terrestrial Ecosystem Research Network (TERN)
Dr Merran Smith	Population Health Research Network (PHRN)
Mr Mark McAuley (Mr McAuley delegated Dr Yeshe Fenner to attend)	Astronomy Australia Limited (AAL)
Mr Andrew Gilbert	Bioplatforms Australia (BPA)

## Logistics

It is expected that the committee will hold its first meeting in August 2012 and will meet as required to provide final advice to government by early 2013.

Secretariat support for the committee will be provided by Research Infrastructure Branch, DIICCSRTE.

## Terms of reference

The committee will:

- provide a forum for consultation between government, data generation and management research infrastructure capabilities, and the broader research sector to inform future planning for research data infrastructure in Australia
- develop a framework view of future requirements for research data management in Australia
- advise on the interrelationships between the data infrastructure investments and the relative roles that government, institutions and research facilities can play in optimising these
- advise on the means by which the effectiveness and efficiency of the relationships between data generating investments and the data infrastructure investments could be optimised in future implementation
- propose options for governance models that would optimise the outcomes from future national research data infrastructure investment.