Submission
2016 National Research Infrastructure Roadmap
Capability Issues Paper

Name	Dr John Beaton
Title/role	Executive Director
Organisation	Academy of the Social Sciences in Australia

Answers to some of the below questions are addressed in the Comments section and attachments.

Question 1: Are there other capability areas that should be considered?

Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.

Question 3: Should national research infrastructure investment assist with access to international facilities?

Question 4: What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?

Question 5: Should research workforce skills be considered a research infrastructure issue?

Question 6: How can national research infrastructure assist in training and skills development?

Question 7: What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

Question 9: What should the criteria and funding arrangements for defunding or decommissioning look like?

Question 10: What financing models should the Government consider to support investment in national research infrastructure?

Question 11: When should capabilities be expected to address standard and accreditation requirements?

Question 12: Are there international or global models that represent best practice for national research infrastructure that could be considered?

Question 13: In considering whole of life investment including decommissioning or defunding for national research infrastructure are there examples domestic or international that should be examined?

Question 14: Are there alternative financing options, including international models that the Government could consider to support investment in national research infrastructure?

Health and Medical Sciences

Question 15: Are the identified emerging directions and research infrastructure capabilities for Health and Medical Sciences right? Are there any missing or additional needed?

Question 16: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 17: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Health and Medical Sciences capability area?

Environment and Natural Resource Management

Question 18: Are the identified emerging directions and research infrastructure capabilities for Environment and Natural Resource Management right? Are there any missing or additional needed?

Question 19: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 20: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Environment and Natural Resource Management capability area?

Advanced Physics, Chemistry, Mathematics and Materials

Question 21: Are the identified emerging directions and research infrastructure capabilities for Advanced Physics, Chemistry, Mathematics and Materials right? Are there any missing or additional needed?

Question 22: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 23: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Advanced Physics, Chemistry, Mathematics and Materials capability area?

Understanding Cultures and Communities

Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?

Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

National Security

Question 27: Are the identified emerging directions and research infrastructure capabilities for National Security right? Are there any missing or additional needed?

Question 28: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 29: Is there anything else that needs to be included or considered in the 2016 Roadmap for the National Security capability area?

Underpinning Research Infrastructure

Question 30: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed?

Question 31: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 32: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Underpinning Research Infrastructure capability area?

Data for Research and Discoverability

Question 33   Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed?

Question 34:   Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 35:   Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?

Other comments

If you believe that there are issues not addressed in this Issues Paper or the associated questions, please provide your comments under this heading noting the overall 20 page limit of submissions.

Summary

In order to maximise the contributions of the social sciences and other research to the advancement of the national interest, this Academy strongly recommends an investment in an integrated national data infrastructure. Such a system, through extensive linkages, would support the collection, analysis, curation, and utilisation of Australian data and, at relatively modest cost, enhance evidence-based policy handsomely. It would bridge all data users and providers including academics, government, business, industry, NFPs, community organisations and even individuals.

The essential elements of the development and strengthening of data infrastructure are

- *improved access and data protection.*
- *improved integration of collections.*
- *reliable and up-to-date information to conduct realistic studies to make credible recommendations.*
- *efficiency, productivity, and scientific quality across disciplines as a result of data comparability[1].*

Similarly, effective training and education programs will strengthen the capability of data infrastructure.

Introduction

Industry, business, government, and community organisations all have the need and opportunities to use information to advance their missions through knowledge. Modern computing technology provides for the rapid management of vast amounts of data to facilitate this. Each of these stakeholders use data that has been provided largely by social scientists from the disciplines of management, demography, statistics, sociology, psychology, economics, political science, etc. who routinely collect and analyse data relating to the behaviour of the Australian people and their institutions. The estimated economic value of open data in Australia is estimated to be $64 billion, with education, transport and consumer products the largest benefiting sectors[2].

Research and analyses by the social sciences can therefore serve the needs of industry and business, but a principal benefit is also the public good through the provision of knowledge to policy development. These sentiments were recognized in the 2011 NCRIS Strategic Roadmap which identified the core strength of Australian social science research as:

> *"Australia's humanities, arts and social science researchers are recognised internationally for delivering solutions to challenging questions facing society and for bringing their expertise to bear on complex issues, such as in health, the environment, social cohesion and security. Their work encompasses law, society, education, identity, economics, business, governance, history, culture, language and creativity." [3]*

The ability of Australian social scientists to provide knowledge-based advice to government is often dependent upon their capacity to study long term trends and developments. Those studies are critically dependent upon access to high quality, nationally representative, data to describe and compare the social, political, economic and environmental characteristics of Australia and Australians.

---

[1] Samara, M. & Scott, H. (2014). *Research Infrastructures in Social Sciences and Humanities. Net4Society.eu.*
https://www.ffg.at/sites/default/files/downloads/program_line/n4s3_ris_in_ssh.pdf
[2] *Open government data and why it matters now.* (2016). *Communications.gov.au.*
https://www.communications.gov.au/departmental-news/open-government-data-and-why-it-matters-now
[3] *2011 strategic roadmap for Australian research infrastructure.* (2011). *Department of Innovation, Industry, Science and Research.*
https://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2011.pdf

The types of data used in Australian research include fundamentally crucial basic data, such as:

- The Australian population census and other important data sources.
- Longitudinal studies of individuals (such as the Household Income and Labour Dynamics in Australia, Longitudinal Study of Australian Children and Longitudinal Surveys of Australian Youth, funded by various Australian government departments).
- Nationally representative cross-sectional surveys such as the Australian Election Study and Australian Survey of Social Attitudes, and periodic surveys by the Australian Bureau of Statistics.
- Time series of economic and social phenomena such as unemployment, labour force participation and other statistics produced from the ABS Labour Force Survey.
- Administrative data collections such as the Research and Evaluation Database from the Department of Employment and the Mortality data register from the Australian Institute of Health and Welfare, and open data sources such as data.gov.au.
- Business data collections such as the historical records of firms and other organisations held in locations such as the Butlin Archives and recent data from business activity and records.
- Qualitative data collections such as the sub-studies from the Australian Longitudinal Study on Women's Health.
- Linked data such as that provided through the Population Health Research Network and Australian Urban Research Infrastructure Network, and the Australian Census Longitudinal Dataset.

These major forms of traditional data collection can now be complemented by incidental data from social media sources such as Twitter and Facebook and from business analysis of transactions conducted.

Small and mid-sized institutional users of data are finding free-access becoming limited in supply and the commodification of big data and deep analytics tools being affordable only by very large corporations and governments. This will only heighten the need for Australian research infrastructure to accommodate the needs of academic institutions and their researchers.

Crucial infrastructure to support Australian data collection has been established through existing programs of government agencies such as the Australian Bureau of Statistics and Department of Social Services, and funding programs such as ARC, NHMRC and NCRIS.

Indeed Australia has a global reputation in many data areas. The issue is to take this area of national comparative advantage for research and build on and enhance its contribution even further through greater infrastructure support.

In particular there are still many gaps and deficiencies. Importantly, we are concerned that each of these sources is subject to significant risks, including inconsistent funding and even data loss. Data sources that should enable long-term evaluation of changes in Australian

public, industrial, business, and social structures are either funded on a short-term basis, or subject to cut-backs in government budgets.

In terms of open data access, Australia is ranked third in the G20 open data ranking, its progress has been less strong than the UK and US, which have each been vigorous and consistent innovators[4]. In addition, a significant proportion of Australian data cannot be made available through an open data environment such as data.gov.au at the unit record level needed for much academic and policy research due to privacy and confidentiality requirements.

What is needed?

In order to maximise the contributions of the social sciences and others, we recommend an investment in an integrated national data infrastructure. Such a system would support collection, analysis, curation, and the utilisation of Australian data.

As noted in the 2016 National Research Infrastructure Capability (NRIC) Issues Paper, a foundation for such a national data infrastructure does exist within the Australian Data Archive (ADA)[5]. The ADA has provided an archival service to Australian social science researchers since 1981. It provides a useful model of some of the breadth, capacity, and access of an integrated structure. The ADA is one of only three existing research infrastructures (of the forty-two identified in the NRIC process) as "not currently supported by national research infrastructure programs".

Without like support the ADA has a greatly reduced capacity to provide the kind of integrated data services that are required to meet the manifold needs of industry, business, government, and community organisations.

The essential elements of the development and strengthening of infrastructure are

- *improved access and data protection.*
- *improved integration of collections.*
- *reliable and up-to-date information to conduct realistic studies to make credible recommendations.*
- *efficiency, productivity, and scientific quality across disciplines as a result of data comparability.* [6]

Similarly, effective training and education programs will strengthen the capability of data infrastructure.

Changes in data collection practices and analyses will require commensurate innovation in research methods to ensure they are consistent with the ongoing needs of Australian researchers contributing to public policy and wider development.

There is also a need for consideration of data's provenance, and the perspectives and context of its development, as fundamental to assessing its validity for use. Given that the

[4] *Open for Business: How Open Data Can Help Achieve the G20 Growth Target.* (2014). *Lateral Economics.* https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf
[5] *Australian Data Archive.* (2016). http://ada.edu.au
[6] Samara, M. & Scott, H. (2014). *Research Infrastructures in Social Sciences and Humanities.Net4Society.eu.* https://www.ffg.at/sites/default/files/downloads/program_line/n4s3_ris_in_ssh.pdf

growth, storage and availability of data are central themes of this review, it is imperative that the capability for rich, critical analysis be cultivated at the same time as usage access grows.

Data providers and analysts require a number of infrastructure pathways including:

1. A stable long-term preservation and curation environment.
2. Effective access for bona-fide researchers, including capabilities for open data download, secure access, and machine-to-machine data access.
3. An integrated data management environment, having the following attributes:
   a. Data storage facilities (such as National Computational Infrastructure and Research Data Services) providing high availability and large-volume data storage and back-up facilities.
   b. Secure access facilities (such as the Population Health Research Network and the ABS Microdata Laboratory).
   c. Data linkage (through Statistical Data Integration Authorities), and data integration and interrogation environments (such as the Australian Urban Research Infrastructure Network).
   d. Data discovery services (such as Australian National Data Service) to enable data discovery through shared platforms and services such as Digital Object Identifiers.
   e. Sufficient data protection and security.
4. Soft infrastructure in support of innovation in research methods, such as the use of probability-based sampling methods for online surveys, and experimental methods for behavioural economics to enable new forms of research data collection and analysis to be established and supported.
5. National protocols for data access and utilisation.

There is also compelling and growing demand in higher education institutions worldwide to provide education and training in New Analytics. These courses will help to serve the needs of industry and business to have work-ready graduates who are informed and capable in areas of data analysis appropriate to massive data sets and emerging technologies.

Why is this important?

Institutions including those of industry, business, government, and community organisations all require a sophisticated understanding of data relating to their constituent parts and the populations they serve. There are a number of reasons as to why national effective and sustained national funding of research infrastructure is crucial for all data users. In addition:

1. There are legal requirements across all sectors for the secure storage of data relating to their activities.
2. Research results across academic sectors require effective curation, as do institutions and individuals on public funding, including for purposes of replication and accountability.
3. Regulatory agencies require substantiated records of research activities, methods, and outcomes.
4. The maintenance of an integrated system of data management is a highly efficient means of providing generations of researchers with comparative evidence to assist them in the production of sophisticated and useful analysis.

5. An integrated national data infrastructure system can maximise data security mechanisms and regulate effective access.
6. Australia will do well in its international partnerships to provide a best-practice data management environment where trust is an imperative. For instance, another example of an integrated data system is the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is harmonised with the U.S. Health and Retirement Study and the English Longitudinal Study of Ageing and has become a role model for several ageing surveys worldwide.

Failure to make appropriate use of available data leads directly to waste. For instance, in the crucial area of education outlays and outcomes, the Productivity Commission has recently noted that:

> *"Spending on education in Australia has increased yet standards have stagnated, because data is not being used effectively to improve student outcomes There has been a 14% increase in spending per student over the last 10 years, yet student performance remains broadly unchanged and, in some areas, has actually decreased. Adopting and applying an evidence-based approach to education policy and teaching practices is what drives a better allocation of resources and improved outcomes. The new report recommends that governments take "a shared approach to developing a high-quality and relevant national education evidence base."[7]*

Comments such as these from the Productivity Commission indicate the importance of data in identifying the source of problems that can be addressed in targeted ways. The further examination of Big Data in train by the Productivity Commission will provide enhanced impetus for this understanding and should equally provide impetus for the infrastructure needed.

Likewise, the Commonwealth Government has recently announced a new Australian Public Service Data Skills and Capability Framework to enhance data skills, data capacity and data literacy. According to Assistant Minister Angus Taylor:

> *"..there was a global undersupply of data-related skills, with the McKinsey Global Institute estimating that by 2018 the US alone would face a shortage of approximately 140,000 to 190,000 people with complex analytical skills and a shortage of 1.5 million managers and analysts to analyse big data and make decisions based on their findings".[8]*

The Government in Australia wishes to ensure that likewise policy here employs the best data and skills. And accessible, attested, comprehensive and deep social science data, alongside other statistics, is at the core of this.

We need only think of the pay-offs to the nation from data intensive social science impact such as education loans, medibank development, HIV prevention, national competition policy, water management, alcohol and tobacco policies, superannuation policies, HIV prevention and more.

---

[7] *Draft report - Education Evidence Base Productivity Commission.* (2016). *Pc.gov.au.*
http://www.pc.gov.au/inquiries/current/education-evidence/draft
[8] *Data skills and capability in the Australia Public Service.* (2016). *Department of Prime Minister and Cabinet.*
https://www.dpmc.gov.au/sites/default/files/publications/data-skills-capability.pdf

Conclusion & Recommendations

This Academy welcomes Government's interest in providing for critical national research infrastructure. It affirms that integrated national data infrastructure must be part of advancing the public interest. Such infrastructure will pay for itself many times over in national advancement.

The Academy recommends investment in integrated national social science data infrastructure to establish a hub for a data repository, building upon the Australian Data Archive, which leverages and harmonises multiple data collection platforms and institutional partners. This hub must ensure sufficient access, data protection and security, well-advertised data sets, efficient regulation and sufficient training, and any other considerations that would support the development of:

- a stable long-term preservation and curation environment
- effective access for bona-fide researchers
- an integrated data management environment
- soft infrastructure in support of innovation in research methods
- clear protocols for data access and utilisation
- sufficient graduate education and training in New Analytics relevant to Big Data
- employment pathways to bridge academic, business, industry and NFP enterprises

This Academy further recommends that any research social science data infrastructure hubs be extended internationally much like the SHARE model of the EU providing unprecedented efficiency in world-wide collaborations.

The Academy is available at any time to further discuss this submission.

Academy of the Social Sciences in Australia
26 Balmain Crescent, Acton ACT 2601
GPO Box 1956, Canberra ACT 2601
P: +61 2 6249 1788
John.Beaton@assa.edu.au