

Universities Australia Submission to the Productivity Commission Inquiry into Data Availability and Use

July 2016

Universities Australia (UA) welcomes the opportunity to provide a submission to the Productivity Commission on data availability and use. Universities play an essential role in driving Australia's open data agenda through their research programs, as producers and users of vast amounts of data in their own right, and as educators of Australia's data analysts and data scientists.

Data is now recognised as a key economic asset underpinning Australia's transformation into an innovation nation. At a time when other countries are investing in their data, Australia's capacity to remain competitive is contingent on its ability to harness the value of its data to drive innovation, productivity and economic growth. Australia is well-positioned to do so, with our globally renowned facilities and Australian researchers leading the way on innovative methods to link and integrate data.

A number of critical elements are needed to improve Australia's ability to grow the economy, improve decision-making and deliver data-driven innovation. The Government's Public Data Policy Statement¹ provides strong policy direction and support for optimising the use and reuse of public sector data. However, this will need to be supplemented with a focus on reducing legal and regulatory barriers, building a skilled workforce and ensuring that appropriate infrastructure is in place.

UA recommends that:

- (i) as a priority, the Productivity Commission consider how to create a legislative and regulatory environment that supports greater access to and use of data, particularly for research and decision-making purposes;*
- (ii) sufficient investment is provided to ensure that Australia's long-term data infrastructure needs are met, taking into account the lessons learnt from the National Research Collaborative Infrastructure Strategy;*
- (iii) the Productivity Commission consider strategies to harness existing talent and increase the knowledge, skills and abilities of the Australian workforce to grow a digital economy and develop an evidence-based and data-oriented culture; and*
- (iv) the Productivity Commission should examine the role of citizen science in increasing the availability and use of data and consider potential strategies to accelerate and scale up these initiatives.*

¹ Australian Government 2015, *Australian Government Public Data Policy Statement*, Commonwealth of Australia, Canberra.

Access to public sector data

Australian governments collect, hold and manage vast amounts of data with enormous potential value to researchers, industry and the broader community. Numerous reviews and studies have confirmed the long-term and ongoing social and economic benefits of open government data. The recent critical literature review by the Department of Communications and the Arts reports an economy-wide value of government data of between \$500 million to \$25 billion per year.²

UA welcomes the Australian Government's Public Data Policy Statement and recent initiatives to improve access to and use of public sector data. Cultural change and leadership is needed if we are to harness the value of Australia's data. The Government's Public Sector Data Management report provides a clear and cohesive roadmap to build on current successes, address barriers and improve practices. It is essential that the provision and curation of data be recognised as a core business activity of government departments and agencies, and that this work is resourced accordingly.

Lack of awareness of what data sets are available can act as a barrier to use. UA welcomes the development of a whole-of-government data catalogue that, in combination with the commitment to make non-sensitive public data open by default, will substantially improve the discoverability, use and value of aggregated or non-sensitive public data. Consideration should also be given to the standardisation of data sets, and strengthening engagement with end users to determine the demand for datasets by building on the [Open Data 500 Australia](#) project.

Research purposes

The kind of data most useful for research purposes is finely grained, individual unit record level data and data that is linked to other datasets. A linked approach is increasingly required to answer complex problems in research and in policy. The Population Health Research Network (PHRN), for example, has demonstrated the value of integrated datasets in health. One project linked 44 million hospital records with death data from four different state registrars over a five-year period. This meant that researchers could link hospital admissions to a death even if they took place at a different hospital or state. This has led to the identification of trends at a national level in ways that have never been possible before. A similarly strategic and considered approach is needed to support the use of other datasets, including linking public sector datasets to private sector data.

Differing requirements for access to data between government agencies can add another layer of complexity when research spans across areas. For example, protected information provisions in social security legislation sets a higher threshold for the use and reuse of data than privacy legislation. It is also vital that we are able to seamlessly link Commonwealth data to State and Territory data. There would be numerous benefits in harmonising privacy and security legislation across jurisdictions and agencies.

Australia is a world leader in developing innovative methods to link and integrate data in the research sector. Current privacy and confidentiality arrangements for public sector data have not kept pace with maturing tools and techniques to manage confidentiality. Researchers rarely need to know the identity of individuals to make use of data, and there are new ways to transform data to protect people's privacy

² Bureau of Communications Research 2016, *Open government data and why it matters*, Commonwealth of Australia, Canberra.

and still make pertinent information available for use. The PHRN, for example, is developing and testing leading-edge technology to ensure the safe and secure linking of data collections while working to protect individuals' identity and privacy. As a matter of priority, the Productivity Commission should consider reforms to the current legislative and regulatory regime to better recognise the maturity of Australia's infrastructure and data capabilities, and support the use of data.

Planning and decision-making purposes

Public sector data provides critical information needed to assist decision-makers in universities, industry, communities and government to plan and respond to key social and economic challenges. National health workforce data, for example, has been collected over the years by a range of agencies including the Australian Institute of Health and Welfare (AIHW), the Australian Bureau of Statistics, the Department of Health and by the former Health Workforce Australia (HWA). Access to such data is useful to monitor trends, better link training opportunities to predicted workforce needs and assist in related education and workforce policy development, particularly at regional and discipline-specific levels. Harnessing this data would better inform Australia's national workforce planning and allow universities and other non-government institutions to play a more active role.

Charging for data services

Data collected in the course of usual government operations or business should be available free of charge to end users. There is greater acceptance for a cost recovery pricing model for specialised or bespoke data services where significant value has been added by government departments and agencies for the benefit of the end-user, such as additional data collection or data manipulation.

However, this approach would need to be carefully considered in relation to publicly-funded research. While UA acknowledges that much of the linked data used in research requires significant input from government agencies and departments, charging publicly-funded researchers would create unnecessary bureaucratic burden without generating any additional government revenue. The costs involved in making complex data available could potentially be offset by efficiency and productivity gains for departments and agencies with the creation of reusable, curated and linked datasets.

An alternative access regime to price is merit. Applications to access data for research purposes should go through a merit assessment before data is provided. Irrespective of the model that is chosen, the costing model needs to be reasonable and transparent enough to enable researchers to include the cost of accessing, archiving and storing data into competitive grants in a way that is acceptable to funders.

Access to private sector data

Governments, universities and businesses have identified Australia's low levels of collaboration between industry and researchers as a major barrier to making a successful transition to a productive and internationally competitive innovation nation. Australia ranks last of 26 OECD countries for the percentage of innovation-active businesses collaborating with universities and other research institutions. The measures introduced in the National Innovation and Science Agenda (NISA) are a vital first step, and other elements of the innovation system—such as data—would benefit from the same injection of long-term, strategic policy direction.

There is enormous potential for universities and industry to better engage and collaborate through data sharing and linking. Through the Yale Open Data Access Project, for example, the United States company, Medtronic provided Yale researchers with unprecedented access to clinical trial data for a new medical device as part of its commitment to patient safety and to address questions raised by the public and medical community about the company's data. Researchers completed an independent and critical review of the device and provided information to medical practitioners and patients to inform their decisions. UA encourages the Productivity Commission to consider the case for greater policy direction and support for private sector companies to share data with university researchers.

University initiatives to improve access to data

The university sector is a data-rich sector, and universities generate and use enormous volumes of data each day. Universities have long recognised the potential for data to further drive high quality education and teaching. Learning analytics are becoming increasingly sophisticated, enabling universities to benchmark their teaching performance against student satisfaction and to monitor student progress, performance and engagement so that intervention can occur when and as required.

Universities are at the forefront of Australia's response to new data challenges in an era where technologies are developing at an exponential rate and producing vast amounts of data on a daily basis. Universities are developing data management strategies to help contend with rapidly growing datasets, involving terabytes and petabytes of information sourced from different areas, and ensuring their quality and accessibility for business intelligence and data analytics applications. The sector has made significant inroads into curating these large data sets through projects such as those supported by the Australian National Data Service (ANDS). ANDS is a leader in facilitating the discovery and sharing of existing data, metadata, and contextual information from university researchers and their collaborating partners.

The sector has strongly supported the Quality Indicators for Learning and Teaching (QILT) initiative, which gives students, providers and the community more accessible, timely and reliable information on higher education performance. UA also welcomes the recent recommendation by Australian Council of Learned Academies (ACOLA) to expand QILT to provide further data and information for higher degree research students. The accessibility of institutional performance data through QILT will make the demand-driven university system significantly more effective by informing student choice, providing trust and confidence in public investment in higher education, and will facilitate better and faster decision-making by the sector.

Data sharing is critical to the research enterprise, and Australia's ongoing research success will depend on its capacity to optimise the use and re-use of research data. Australia, like its international counterparts, is taking steps towards open access publications and data. Both the National Health and Medical Research Council (NHMRC) and Australian Research Council (ARC) have introduced open access policies for publications, and researchers are required to consider and articulate a data management plan as part of ARC grant applications. Australia's universities and research sector have already been making significant changes to their internal policies and processes to improve the openness of research data and increase the flow of information and knowledge between the research sector and broader society.

Data infrastructure

While increasing the volume of available data is important, we also need to ensure that Australia has the infrastructure in place to allow for secure access, linking and analysis. Data must be well-curated and easily accessible, with supporting infrastructure and services in place to promote interoperability and collaboration across the public and private sectors. The growing volume and complexity of data requires cutting-edge tools, especially when data is drawn from and shared across different areas.

The National Collaboration Research Infrastructure Strategy (NCRIS) has pioneered a national-level, strategic and collaborative approach to infrastructure and has improved the excellence and impact of Australia's research data. The NCRIS facilities operate on a principle of collecting data once and re-using multiple times.

ANDS provides training on data management policy and planning, and collecting and sharing best practice. Other eResearch investments like the Research Data Storage Infrastructure (RDSI) provide national storage infrastructure to securely store and access datasets of lasting value. The Australian Access Federation supports data access and authentication, while other critical facilities such as AARNet provide the tools needed to link, transmit and share large data sets across the country and internationally. Australia's supercomputing facilities in the National Computational Infrastructure and the Pawsey Supercomputing Centre provide peak capacity for extremely large and complex datasets, such as those relating to climate science, geoscience and radioastronomy. Data infrastructure and related services offer wide-ranging benefits beyond providing access to data itself. Reviews of data centres have found that they provide additional incentives for researchers to improve the quality of their data, and by collating a wide range of datasets in one place, researchers are encouraged to explore linkages to related work.³

This combination of infrastructure and resources has provided Australian researchers with opportunities for improved collaboration and has greatly improved the ease and level of access to research data by industry. The Integrated Marine Observing System (IMOS), for example, is a multi-institutional, multi-disciplinary facility that operates observation equipment throughout Australia's coastal and open oceans, and makes all of its data available to the Australian marine and climate science community and its international collaborators. The Great Australian Bight research program uses IMOS data to examine how this important marine asset can be developed sustainably. The project is a collaboration between BP, CSIRO, the South Australian Research and Development Institute, the University of Adelaide and Flinders University.

Many of the technical challenges in opening access to data now faced by government departments and agencies have been resolved within the research sector and there is untapped potential to deepen linkages and share lessons across government, researchers and industry. The Department of Social Services, for example, is investigating the use of the Secure Unified Research Environment, provided by the Sax Institute as part of the NCRIS-funded PHRN, as a means of securely sharing unit-level social security datasets with credentialed individuals.⁴

³ Houghton, J & Gruen, N 2014, *Open Research Data*, Report to the Australian National Data Service.

⁴ APH Senate Select Committee on Health 2016, *Sixth Interim Report (Big health data: Australia's big potential)*, Commonwealth of Australia, Canberra.

UA welcomes the Government's commitment of ongoing funding to maintain Australia's world-class data infrastructure and the development of a National Research Infrastructure Roadmap to guide future investment. New capital investment, in addition to the operational funding allocated to current NCRIS facilities, will be critical in generating new jobs and driving improvements in the access to and use of data in both the public and private sectors.

Citizen science

A key focus of the Productivity Commission's Issues Paper is on the impact of open data, data sharing and linking on individuals as subjects of data. Increasingly though, individuals are also active contributors, co-creators and collaborators of data, helping to build and describe research datasets through citizen science initiatives.

Citizen scientists have played a crucial role in discovering new species, making medical breakthroughs and identifying distant galaxies. In Australia, 10,000 people have collected 10 million records for the Atlas of Australian Birds database. More than 9,600 people analysed 330,000 photos of marine habitats in a single week.⁵ [GovHack](#) is another example of governments providing members of the public with greater access to data to solve public problems. Providing opportunities for members of the public to contribute to research and policy projects is a key mechanism for building Australia's pipeline of much needed data and analytical skills, and improving scientific literacy within the community.

However, more could be done to enhance the effectiveness of individual projects and better harness the ingenuity and skills of the public. In the US, government agencies have made significant investments in citizen science, launching CitizenScience.gov as an online government-wide listing of citizen science projects and appointing citizen science agency coordinators. The recent creation of the Australian Citizen Science Association provides the opportunity to take a more strategic approach. UA recommends the Productivity Commission examine the more active role that individuals play in accessing and using data, and consider potential strategies to accelerate and scale up citizen science initiatives.

Data workforce

Australia's ability to harness the full value of its data relies on building a workforce that is digitally literate and equipped with the skills for data analysis, interpretation and research translation. The introduction of the demand-driven system for university places has helped to meet the growing employment need for highly-skilled graduates. It is important that Australia's students are equipped to exploit technological and digital developments, particularly those that rely on a range of information and communication technology and data analysis. Recent initiatives to improve the take up of STEM in schools and opportunities for women to stay in STEM careers will also help meet the workforce needs of the digital economy.

Increasing the number of employees with research and data skills in government agencies and industry is critical to grow our analytical and innovative capacity in the public and private sectors. Unlike many other nations, the majority of Australia's researchers are employed in the higher education sector. It is of vital

⁵ Pecl, G, Gillies, C, Sbrocchi, C & Roetman, P 2015, 'Building Australia through Citizen Science' in Office of the Chief Scientist Occasional Paper Series 11, Commonwealth of Australia, Canberra.

importance that our research system provides the flexible learning options and skills needed to ensure graduates can succeed in whatever profession they choose. A national-scale industry placement scheme for research students, as recommended by the recent reviews of the research training system and research policy and funding arrangements, should be considered further.

Incorporating additional requirements under the current policy and funding arrangements is challenging. For example, there is concern that the current arrangements create barriers for potential students already working in government or industry to complete a PhD part-time and improve their data and analytical skills. The changes to the funding for research training announced as part of the NISA is a positive step, but much will rely on the implementation of the new program. UA looks forward to working with industry and Government to develop an approach that maximises the flexibility in the system and will meet Australia's future workforce needs.