THE UNIVERSITY OF MELBOURNE

UNIVERSITY OF MELBOURNE SUBMISSION TO THE PRODUCTIVITY COMMISSION INQUIRY INTO

DATA AVAILABILITY AND USE

29 JULY 2016

Contents

Front page image used with permission, University of Melbourne Marketing and Communications, July 2016

Introduction

The Productivity Commission's Issues Paper has identified the critical considerations and barriers pertinent to improving availability and access to public and private data. These issues will be important to address if the benefits envisaged in the Australian Government's Public Data Policy Statement are to be realised.

The University of Melbourne ('the University') commends the emerging consensus and whole-of-government commitment to achieving greater cooperation and transparency on sharing digital information. Facilitating the safe and appropriate aggregation and linkage of data between different agencies and sectors will be critical.

Australia's future economic growth and social innovation depends on maintaining a robust research base. University researchers increasingly draw on large data sets to gain insights and discoveries that can drive innovation, efficiency, productivity, social wellbeing and economic growth for the nation.

With research expenditure second only to that of the CSIRO and the largest cohort of research students in Australia, the University is one of Australia's leading research-intensive institutions and host to a significant number of the nation's internationally renowned research institutes.

From the University perspective, there is outstanding research value in core datasets held by public agencies, particularly census and labour force longitudinal data, geo-coded National Address files, personal income and business tax data, health interventions and benefits data and social society data. However, these flagship public resources make up only part of the untapped data wealth that could translate to significant public benefit. Increasing technological ability to link disparate datasets further amplifies the potential impacts and benefits of data analytics, and places even greater value on policy reforms to enable efficient access to diverse data resources.

Universities and researchers are also pre-eminent data generators. The University generates data of significant value to national and international research communities, across biomedical, environmental, urban, economic, social, and historical data, which all have impact and value for public policy.

University academics maintain large and evolving datasets over the course of research. The University has well-established ethical and procedural arrangements governing the collection, storage and use of research data, with concurrent reference to applicable national codes, research sector requirements and/or parallel institutional ethical approval processes.

By collecting or utilising datasets as part of an evidence base, researchers contribute knowledge and findings that improve policy and practice, make better use of scarce resources and improve social services and outcomes. In health research and health care, data can be life-saving; historical big data can inform policy-makers on how to prevent suffering and premature loss of life.

Similarly, with a clearer legal framework and enabling conditions, data-driven experimentation and innovation on new technologies, digital services and other digital products has the potential to deliver transformational public benefits.

This submission includes a number of case studies and examples that demonstrate the **catalytic value of data** – particularly linked public or administrative data – in the research sector. This submission also identifies existing impediments and makes recommendations for essential next steps in order that Australia can advance its policy intent of providing more comprehensive, streamlined and efficient access to data.

Areas recommended for focus

The University strongly supports the goal of enhancing access and availability to data, particularly rich public data holdings. This submission supports the concept of enabling release of data for broad public purpose application, along with the development of appropriate protections to reflect the crucial importance of data protection, security, privacy, consent and consumer information. Key focus areas for the submission include:

- Whole-of-government, cross-jurisdictional collaboration facilitated by COAG leadership, in pursuit of a new model for accessing data such as establishment of a Data Release Authority or expansion of Trusted Access Models.

- Facilitate an authorising environment and public confidence on getting more public value from data.

- Build infrastructure for efficient collection of data, including through NCRIS, electronic medical records, and other key enabling platforms.

- Enable efficient and secure access to catalytic research data to trusted research communities.

- Update security and privacy protections to protect data integrity, including streamlining ethics approvals and national protocols for research undertakings relating to data collection and management.

For **further information** or to discuss our submission, Professor James McCluskey, Deputy Vice Chancellor (Research) can be contacted on dvc-research@unimelb.edu.au or (03) 8344 3238.

Summary of Recommendations

A) Whole-of-government, cross-jurisdictional collaboration on data integrity and release

Recommendation: Through the Council of Australian Governments progress foundation reforms, funding architecture and streamlined governance arrangements to support greater data availability and use across and within jurisdictions.

Recommendation: Progress a nationally harmonised legislative framework on data availability and use to resolve common jurisdictional duplication and overlap and facilitate wider data access in research projects with national implications.

Recommendation: Investigate the feasibility of an institutional 'data release authority' with a national remit to act as a central broker and manager of data integrity and conditions for release.

Recommendation: Investigate options for a 'trusted access domain' for secure access to public sector datasets by accredited researchers, with provision of linked administrative and survey data linked across time and sectors by accredited integration authorities and streamlined access to aggregate and individual data.

B) Facilitate an authorising environment and public confidence on getting more public value from data

Recommendation: With appropriate resourcing, public agencies such as the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) could provide effective strategic leadership on major data infrastructure and research sector advocacy on data access and availability.

Recommendation: Identify actions to be led by key agencies and research bodies to ensure that all entities and researchers seeking to benefit from greater data availability bear an appropriately rigorous onus to protect data integrity, potentially through establishing a single national accreditation or approvals process for highly-sensitive, linked administrative data.

Recommendation: Identify sites of responsibility for the creation of readily-accessible interfaces for individuals and organisations to check and manage information held about them across public and private databases.

C) Build infrastructure for efficient collection of data

Recommendation: Support critical tools, technologies and infrastructure, such as the National Collaborative Research Infrastructure Strategy, to enable data collection and usage, including escalated investment in electronic medical records in all health institutions and clinical settings.

Recommendation: Drive a nationwide introduction of an electronic medical records system (initially across all NHMRC Advanced Health Research and Translation Centres) at Healthcare Information Systems Society (HIMSS) Level 6 by 2020, and HIMSS Level 7 by 2022.[i]

Recommendation: Streamline national data linkage and access approval processes, including linkage of State and Commonwealth data linkage units, subject to comprehensive privacy and security protocols.

Recommendation: Establish incentives for public sector data custodians to conduct a formal process of consultation with researchers to explore future uses of the data to address health, technological, welfare and other social priorities as part of the design of public data holdings.

Recommendation: To keep pace with future workforce demands, government, universities and industry should recognise and invest in opportunities to train a nationally distributed and skilled workforce to supply best practice data management and data publishing expertise across all sectors.

D) Enable efficient and secure access to catalytic research data to trusted research communities

Recommendation: Enable free and secure access to public sector datasets and other datasets that enable research with public benefits, including electronic medical records, land use, population data, real-time healthcare data, registries (such as births, deaths and marriages) and other key datasets on social and economic activities.

Recommendation: Establish nationally consistent, fast, linked and secure digital systems across Australian States and Territories for storing historical and future births, deaths and marriages data.

Recommendation: Streamline and harmonise research ethics legislation and committee protocols on research ethics approvals across jurisdictions, sectors and institutions, particularly for clinical trials and other multi-centre human health research projects.

E) Update security and privacy protections to protect data integrity

Recommendation: Establish uniform national privacy legislation to clarify interests, protect data and enable release though safeguard conditions, public interest exemptions and other defences.

Recommendation: Establish a tailored defence for research use of data in relevant legislation.

Recommendation: Establish protocols to ensure data collected by non-government institutions for research purposes can only be used by government agencies for research purposes and policy development.

University of Melbourne submission

As a pioneering research-intensive university, the University is able to draw on deep experience and expertise across our campus about the direct relevance of data availability to research outcomes, and translation of those outcomes into community impact that can change – or save – people's lives.

A list of University of Melbourne academics and researchers who have contributed to or endorsed this University submission is included at **Appendix A**.

Data availability and access has a particular relevance to cutting-edge medical and biomedical research. The case studies and recommendations in this submission reflect the University's strong biomedical research capability and perspective. The University is a major partner with leading hospitals in the Melbourne Biomedical Precinct (MBP), an internationally significant and renowned aggregation of 25 health service, research and academic partners.

The MBP houses nationally important NCRIS research infrastructure including the 7 Tesla magnetic imaging system hosted at the Melbourne Brain Centre, as well as other research platforms based at the University's Bio21 Institute for Molecular Science and Biotechnology, the Victorian Comprehensive Cancer Centre and the Peter Doherty Institute for Infection and Immunity.

With 10,000 medical and scientific professionals conducting extensive clinical practice, research, teaching and research training through the MBP, the University and its partner institutions place significance on the future potential for scientific innovation, effective public policy and health interventions achievable through swift action towards greater data access and availability.

The University also hosts a number of multidisciplinary research activities that utilise data as a core resource. These include AURIN, Australia's NCRIS funded urban infrastructure network, a national collaboration delivering e-research infrastructure to empower better decisions for Australia's urban settlements; a Victorian node of the NCRIS funded RDS storage and access infrastructure for significant datasets; the Melbourne Networked Society Institute, which translates digital technologies and applications into services for the benefit of society; and the Centre for Epidemiology and Biostatistics. Some other examples are provided below.

Data with key relevance to the University

The Productivity Commission's Issues Paper observed the variety of data accumulating exponentially and the key categories pertinent to the inquiry. A wide variety of data categories are of particular relevance to the University researchers, including but not limited to:

- Digital data, such as survey data from purposive collections of research data in online environments.

- Data routinely generated through contact with health professionals, hospitalisations, vaccinations and social service providers and increasingly through the use of direct-to-consumer services such as pathology and genetic testing services.

- Self-generated 'life-logging' data (including metadata) emitted from mobile phones and other smart appliances, generated through educational and lifestyle applications such as fitness monitoring devices and web-based games, gambling, dating, and posts on social media.

- Transactional and geospatial data, including data generated from online records of retail purchases and the use of educational and financial services and roads and transport systems, as well as location sensing devices in public places.

- Administrative and legal data such as data about births, deaths, marriages, credit ratings, criminal convictions, incarceration rates, liquor and gambling licensing, immigration and customs records.[ii]

Over the last four years the University has run an institutional process to capture, store and publish data of significant value and impact. Although finalised data products and reference datasets are typically published, there are six broad categories of data forming the following taxonomy:

- Temporary or 'scratch' data – data that is generated on a path to working data.

- Working data – data that is kept and used to derive further data.

- Results data – primary data produced as a result of measurements and simulations.

- Observational data – curated data from instruments, such as telescopes.

- Published data – data associated with publications. Publishing data requires skills and expertise in data curation and data management. These are skills that are distinct from the research skill set and thus require a separate skilled workforce to conduct these sorts of activities.

- Reference data – data generated as reference datasets with the intention of being utilised by broad communities for further investigation.

All categories of data can be an important part of the research process, but two types that are specifically designed to be utilised by communities external to the research groups that generated the data are observational and reference data.

Data as a gateway to public good: Atlas of Living Australia

The Atlas of Living Australia (ALA) is an example of how centrally hosted reference data can be used to educate and support the interests of the wider community. The ALA is a collaborative, national project that aggregates biodiversity data from multiple sources and makes it available and usable online.

Supported as part of NCRIS, the ALA collects and hosts information on all the known species in Australia aggregated from a wide range of data providers: museums, herbaria, community groups, government departments, individuals and universities. The University of Melbourne is one of many contributors to the ALA.

The ALA supports researchers and members of the general public to undertake research, environmental monitoring, conservation planning, education, and biosecurity activities, and allows open learning about the biodiversity in Australia's diverse geographic area.

Existing data resources and sharing arrangements at the University

The University is demonstrably committed to finding ways to support data storage and sharing, including investing in platforms to support collaborative access to research data assets of national significance (including national reference collections).

The cultural change towards data openness that is occurring on the national and international stage is also occurring on campus. In some research disciplines more than others, such as climate studies, astrophysics and genomics, it is increasingly standard for datasets to be uploaded to data sharing infrastructure.

Exemplar data holdings or resources generated by, hosted by, or involving the University, are included below as illustration:

- **Australian Twins Registry**, a national registry that helps realise the full potential of research involving twins, to improve the health and wellbeing of all Australians.

- **Peter MacCallum Cancer Centre** Microscopy and Genomic Data repository.

- The **Melbourne Femur Collection**, a unique archive of human femoral bone tissue.

- The **Endocrine Genomics Virtual Laboratory** (EndoVL), which provides clinical/biomedical data resources supporting a wide range of clinical research groups across Australia and globally.

- The **Melbourne Academic Centre for Health** (**MACH**) is the largest NHMRC acknowledged **Advanced Health and Research Translation Centre** (AHRTC) in Australia. It is leading work on the standardisation of ethics and governance processes for multi-centre health research projects including clinical trials and is working closely on opt-in/opt-out strategies for patients to be involved in medical research. It also develops strategies for future research for its partners and advocates a strong data and EMR network for both healthcare and medical research.

  MACH is the representative body for 10 major Melbourne hospitals, eight medical research institutes and the University of Melbourne. It distinguishes itself by focusing on enhanced patient outcomes from a system perspective, rather than a focus on particular diseases. A key criterion for this recognition of the AHRTC by the NHMRC is a demonstrable record of collaboration and sharing of resources for research and access to patients, registries and databases.

- With support from the Department of Social Services, the **Melbourne Institute** has run the **Household, Income and Labour Dynamics in Australia (HILDA)** survey since 2001. The HILDA survey generates a standard dataset based on a household-based panel study and is available at a low cost to researchers. Iterations of the HILDA survey have led to several hundred publications, particularly in the social sciences.

- The University is lead agent for the **Australian Urban Research Infrastructure Network** (**AURIN**) an e-Infrastructure project supported by NCRIS. AURIN provides researchers, policy-makers and decision-makers with access to high value datasets and services through the AURIN Portal. AURIN links urban researchers, policy and decision makers to data, analytical tools and visualisation components to inform government policy settings and improve decision making within Australian Cities and regions.

  The University notes the separate submission made by AURIN to the Productivity Commission's present inquiry and endorses the recommendations contained therein.

- The University hosts the NCRIS funded **National eCollaboration Tools and Resources (NeCTAR)** project, that supports 14 virtual laboratories in diverse domains, each of which provide sophisticated tools for geographically distributed research communities to share data resources.

- The University works with the **Australian National Data Service (ANDS)** which is supporting the development of the 'Australian Research Data Commons', a cohesive collection of research data resources from all Australian research institutions, supported by NCRIS, to make better use of Australia's research data outputs.

- The University hosts the **Victorian Life Sciences Computing Initiative**, a supercomputer facility to help life scientists address some of the biggest challenges facing our health system, including questions associated with the analysis of data key data-generating research instruments in the precinct and beyond.

- **Community Indicators Victoria** (CIV) is located at the University and provides a comprehensive framework of community wellbeing measured by local level data derived from multiple administrative, survey and spatial data sources. CIV is a unique data resource that is not available in any other Australian state. CIV indicators are frequently used by Local, State and Australian Government departments, community organisations, schools and philanthropic organisations to measure societal progress, encourage evidence-based planning, monitor improvements, engage communities and inform the work of governance groups. For instance, CIV is one of the most frequently cited sources of evidence underpinning Victorian Municipal Public Health Plans. The CIV database can also be used by the general public to generate health and wellbeing reports for local areas.

- The University participates in the **Research Data Services Project (RDS),** a national infrastructure able to hold significant collections of research data. RDS is hosted at the University of Queensland and has a Victorian node, **VicNode**, co-hosted by the University of Melbourne and Monash University.

- With numerous partner organisations, the University participates in **med.data.edu.au**, a national facility to provide petabyte-scale research data storage and related high-speed networked computational services to Australian medical and health research organisations.

- A number of area-specific large data collections are hosted by RDS nodes, including **VicNode,** QLD nodes QCIF BNE and a NSW node, Intersect. These collections include the Australian Coordinated Characterisation Data Space (ACCDS); Australian National Medical Research Data Storage Facility (ANMRDSF); Murchison Widefield Array Data Archive (MWADA); National Environment Research Data Collection (NERDC); and the National Genomics Data Storage Facility (NGDSF).

These platforms, among others, are critical infrastructure for changing practices, research possibilities and data capabilities amongst the research profession.

Data sharing has been enabled by funding streams and platforms such as the NCRIS-funded NeCTAR Research Cloud and RDS VicNode storage services and the Australian National Data Services (ANDS). The University notes that the success of data hosting infrastructure is influenced by the level and continuity of resourcing, as well as by the regulatory environment.

High-value datasets

The Productivity Commission's Issues Paper sought feedback on the characteristics of high-value datasets. It is the University's view that a fixed categorisation of high value may not be helpful in the long term.

The University observes that most public sector datasets are likely to be of value to the research sector. However, the notion of data value is contingent on so many variables that it would be difficult to establish a lasting regime – whether operated by human agency or automated by algorithm – for assessing 'high value' and subsequent release imperatives. There is no single feature of a high-value dataset; rather, it is the combination of data features (quality, scope, granularity) and context (linkage to other data, intended application) that bestows value.

The perceived value of datasets is likely to be judged differently from different user perspectives; high-value datasets from the perspective of a research institution may be quite different to high-value from a commercial or retail standpoint. Different institutions already make value judgments on the intended purposes of the data usage; for instance, the Melbourne Institute opens its data for 'public good' use, but not for commercial use.

The use of data by researchers is also likely to produce improvements in data quality.. Data custodians must work closely with Australian researchers to identify opportunities to both exploit existing data and develop new resources that take into account international developments (for example, multinational collaborations to pool data).

Often the value of data arises from its application (i.e. the questions asked of the data) and the depth of linkage with other datasets (i.e. higher value from a more complete and empirical picture). Repurposing data with reference to administrative datasets or social media can transform the value proposition of a single dataset.

For instance, urban monitoring data may be of high value at three different scales, depending on the intended usage:

- a 'dump' of two weeks' urban monitoring data may be sufficient to provide a solid baseline for a study;

- several years' of archived urban monitoring data may be valuable for underpinning a longer-term analysis; or

- live data with real-time data stream, the largest and most expensive form of data, may be the necessary, and most costly, minimum requirement for app or service development.

The Productivity Commission's Issues Paper notes that Australian Governments collect, create and produce a wide variety of data including that which has significant potential for research use.

The Issues Paper also correctly observes that major public datasets are often made inaccessible to neutral third parties (such as academic researchers) by regulatory processes or contractual limitations.

Data as gateway to public good: understanding war veterans' disability support needs

In 2015, the University of Melbourne's Professor Philip Clarke and colleagues published research that was commissioned by the Department of Veterans' Affairs (DVA) in Australia to analyse the department's world-class, internally owned datasets. The main purpose of this research was to build an evidence base for the long term disability claims of Vietnam veterans.

The research team, including a specialist data linkage agency, was able to link datasets about veterans, including disability claims data to war-service information, war-service rolls, Medicare and PBS data, which were all held by DVA due to DVA's unique holistic care approach. The project linked the war-service records of more 60,000 personnel to subsequent disability claims over 50 year period. The data enabled a better understanding of the impacts of war service on disability, as well as highlighting the high level of claims (over 70 per cent of people who served in Vietnam have developed a war-related disability). Importantly the study demonstrates the need for long-term follow-up through data linkage as majority claims occurred more than 20 years after the end of Australia's involvement in the war. Common disabilities included post-traumatic stress disorder (PTSD), hearing loss and musculoskeletal diseases. The pension costs of treating and supporting veterans for these disabilities are likely to be more than AUD$1.5 million each for the remainder of veterans' lifetimes.

From a budget and a human welfare perspective, the data-based research highlighted the policy priority of finding effective strategies to prevent and treat PTSD. It also demonstrated patterns of disability and pension claims amongst war veterans, which assists policy-makers to anticipate the coming costs of other military engagements, such as the more recent mobilisation in Afghanistan.[iii]

Transformative data: specific datasets and data infrastructure of catalytic value to University researchers

Data is now the critical research infrastructure across the natural, physical, medical and social sciences, as well as for data-driven innovation on digital technologies, products and services. With the establishment of appropriately secure conditions and agile safeguards, there is significant – sometimes catalytic – public benefit to making large publically held data available for research use.

Responding to the question posed by the Issues Paper, this section highlights a small selection of top order data types that if created or released would drastically transform research capability and flow-on public benefits.

1. *Electronic medical records and integrated medical research data*

The nature of medical research has already changed significantly over the last decade. Larger research populations are needed to successfully investigate rare or complex combinations of diseases and deliver personalised medicine. Currently these large cohorts are recruited through many different participating hospitals and health care institutes.

If Australia is to participate in cutting-edge global research it must have records that can be interrogated readily and span multiple sites and large numbers of patients. For countries like Australia with relatively small populations this is particularly challenging.

Electronic medical records (EMR) would resolve a number of the barriers facing researchers seeking to work with large subject cohorts. Investment in the infrastructure to support EMR and the policy and regulatory environment for data sharing would mean researchers could progress past traditional inefficiencies in patient/subject recruitment.

Without widely-implemented EMR systems, individual researchers recruit patients for clinical trials with little or no crossover with the records systems of neighbouring medical institutes or other medical research programs. This patient recruitment demands time-consuming manual data collection. EMR and the subsequent data sharing enabled by EMR systems would give rise to far greater collaboration within Australian research communities as well as with international research consortia.

Internationally, the countries that lead the world's biomedical research are utilising EMR in all clinical settings. This is shown in the US and Europe, where EMR is swiftly emerging as the standard and pre-eminent clinical research source in medical sciences. For instance, in Sweden over 90 per cent of primary care providers have used electronic patient records for diagnostic data since 2009.[iv]

A genomic data surge is occurring around the world with the rise of EMR systems. A well-integrated data system that allows data collection from all types of devices, connected to a high-level EMR that

is suitable for biomedical research will provide a strong data backbone for Australian clinical and non-clinical biomedical research, consistent with our international competitors and consistent with our aim to retain our position as a leading biomedical research country.[v] The University notes recent investment in computational infrastructure platform med.data.edu.au to underpin research capability for the Australian health and medical research sector.[vi]

As well as enabling cutting-edge medical research, a forward-looking digital health data platform, supported by a clear strategy for safe and effective data linkage and release, would revolutionise Australia's health care.[vii] EMR systems assist hospitals and clinics to deliver safer care and more effective medication, access information quickly, support clinical decision making and obtain results efficiently. This increases efficiency by preventing duplication of services and improving patient outcomes through a 'one view' of the patient's treatment by several clinicians or movements between health services. EMR systems also enable a higher degree of patient ownership of their medical records, as patient files can readily follow the individual.

Electronic Medical Records – international comparisons

In the United States, about 35 per cent of approximately 1900 hospitals are working with an integrated EMR at HIMMS-EMRAM level 6 or 7, the highest levels of data integration and digitalisation of hospitals. Another 35 per cent of US hospitals are currently at level 5 and are rapidly approaching the highest levels.

In Europe, at least 80 academic and teaching hospitals are at level 6 or 7.

In Australia, only one hospital, St. Stephens in Queensland, has reached level 6 data integration and digitalisation. Australia has no hospitals at level 7. Most Australian hospitals use a number of different and potentially incompatible systems at levels 5 or below.

A small number of Melbourne hospitals are using EMRs. The Royal Children's Hospital launched its EMR in late April 2016. It is one of the first paediatric hospitals in Australia to replace paper-based medical records with a comprehensive state-of-the-art electronic record. Austin Health also uses a comprehensive EMR system and is participating in the Australian Digital Health Agency's MyHealthRecord system.

The University notes the oversight role taken up by the new Australian Commission for eHealth in July 2016 and endorses the important work underway to develop a national e-health records system.

2. *Population data*

The University views population data as a high-value dataset to the research sector. Historical population datasets provide rich social, economic, biological and medical data linked to time and place.

Population data can be used in conjunction with other data as an evidence base for scientific or medical investigations. As chronic disease is increasingly understood to have intergenerational dimensions, researchers need to track back to grandparents and great grandparents, their places of birth, socio-economic status and lifespan. Modern Australia is a nation of immigrants, bringing diversity of genetic, physiological and psychological traits reflecting the health and exposures of their ancestors. Today, with new populations from Asia and Africa bringing embodied histories of different diets, deprivation and malnutrition, as well as torture and war, there is greater potential and need to understand the biological and psychological legacies of the past in order to plan effective health and welfare services.

As data sharing technology matures over coming years, it will be feasible to link historical data to current clinical records, Medicare records, the Cancer Registry, the Twin Registry and other similar data registries. With appropriate social licence and regulatory permissions, it would be possible that Victoria's cord-blood collection could be linked with other key datasets.[viii]

Data analytics can reveal the causes and therefore, possible prevention and response pathways, for health problems or disorders. Population data linkage is already occurring for such purposes in other countries:

- Sweden is linking its deep historical demographic database to contemporary medical and population data. This is essential to support the clinical application of genomics. The historically deep Swedish data has revealed the effects of maternal malnutrition and stress on the survival of grand-daughters, and of stress on pre-pubescent boys on the life span of their grandsons.

- In Canada, the Quebecois population registers have placed Montreal at the centre of population genetics and the study of inherited mutations.

- Studies of the Dutch Famine (in World War II) and other research conducted in China have revealed the increase in psychiatric illness in those exposed in utero to extreme stress and calorie restriction. In time, ongoing data analysis may show whether the third generations of these populations have suffered ill-effects as well.

Australia is home to many civilian war and terror survivors whose children and therefore grandchildren as foetal eggs, were exposed in utero to famine and extreme stress. Similarly, these datasets could help answer significant scientific questions and in doing so, fundamentally influence Australia's health policy.

Population data also has a high-value application in social sciences-based research and policy. Demography – the study of population structure and change – is central to all government and business planning. Fertility, survival, poverty and wealth can all be inherited and of course altered by historical exposures and experience. Population data could have catalytic social policy impact in a number of ways:

- Economic historians could use it to understand how wealth has been accumulated or lost by individuals over time and transferred through generations.

- Geographic Information Systems (GIS) could be used to map health and disease, while sophisticated, historically specific socio-economic markers on residence and occupation would illuminate social factors.

- Diseases, communities and groups at risk could be mapped and clustered, in order to inform most effective use of public interventions and funding.

- Probate records from the Victorian Public Records Office could be linked to deaths, enabling economists to undertake long-run analysis of intergenerational wealth similar to the work undertaken by Thomas Piketty in France.

Data as a gateway to public good: biologic medicines

Biologic medicines are an increasing proportion of new drugs coming to market. These are highly complex treatments targeted to a much more defined population of patients. As yet researchers have limited data on the impacts of many aspects of their use, such as the impact of switching between agents. Wider availability of this data, held by industry and regulators, could play an important role in driving substantial and much needed improvements to Australia's pharmacovigilance systems, to support the safe and effective use of these medicines.[ix]

3. *Real-time health care data*

Real-time clinical data is of high value to researchers and clinicians working in health care. To enable this, Australia needs a twenty-first century, patient-centred digital health system. The University supports the concept of 'digital health' as an enabler of a fluid, virtual, multidisciplinary healthcare that is seamlessly wrapped around the patient.[x]

A linked digital health system offers potential to drive the collection and use of real time data on clinically meaningful outcomes. The ability to have more rapid evaluation of practise and outcomes – both for individuals and populations, and whether the outcomes are positive or otherwise – will lead to a range of opportunities for early intervention and improvement. Data linkage and digitised records could transform approaches to continuing quality improvement in health outcomes.[xi]

Data as a gateway to public good: diverse applications

AusHEART: This study collected clinical information assessing cardiovascular risk of patients in general practice. The University of Melbourne's Professor Philip Clarke was able to supplement that data by asking the patients to give their permission to link it to MBS and PBS data through a protocol developed by Medicare Australia. That linkage provided new insights into the effect of co-payments on usage of drugs such as statins. It also gave researchers new insights into the distribution of Medicare benefits across different income groups. Recruitment of patients for this study is noteworthy given the high level of voluntary participation. Patients were recruited from GP clinics and were asked to sign a linkage consent form to remain current for five years. Around 70 per cent of patients agreed to the linkage of their data for the purposes stated.

Planning and placing localised health services: The Australian Urban Research Infrastructure Network (AURIN) hosted by the University of Melbourne is using government open data programs to help health authorities and urban planners identify places with greatest need for extra health services. The project is an example of how data can enable governments, agencies, researchers and businesses to collaborate for the benefit of the community.

The Urban Health Data Connectivity Project will make health data available through the online AURIN Portal, where it can be analysed alongside a host of urban and demographic data, such as population density, income, social disadvantage and other indicators of liveability and equity.

This will enable urban researchers, planners and policy-makers to identify where future hospitals, GPs and even dentists will be most needed as baby boomers hit retirement age. This endeavour is also a good example of data linkage with public policy application, as the project gathers data from a variety of sources including ANDS, med.data.edu.au, National Health Services Directory, NSW fall-related injury hospitalisations data, and the 'MyHospitals' dataset.[xii]

*Data Hubs***:** AURIN has set up a series of Data Hubs across Australia that work as a single point of access allowing users to search, access and use a variety of data. The Data Hub allows users and data custodians to contribute data to the hub and provides information about the data (metadata).

Collaboration and interaction between a community of data users and data owners differentiates the AURIN Data Hub concept from the data warehouse and geoportal concepts. This collaboration has direct benefits for data custodians, and addresses the issues of discovery, access, and format diversity of research data. This includes access to private or restricted data, not just open or free data.[xiii]

Impediments to data access and availability

As noted in the Issues Paper, there are a number of issues surrounding access and use of public sector data, which serve as barriers to data availability for research purposes. This section identifies some of the key impediments to access from the University research perspective.

Trust – Increased data openness carries a risk of revealing information that is sensitive to individuals or organisations. As noted by the Issues Paper, trust is essential to maintaining confidence in the collection, storage and usage of data. The public commitment to releasing data should be considered within the context of enhancing the social compact that Australians have already made with each other for their mutual and societal benefit.

Lack of data integration – From a health and medical research perspective, the lack of integration of patient controlled records systems is an impediment to data access and usage. For instance, MyHealthRecord.gov.au, a secure online summary of Australians' health information hosted by the Australian Government's Australian Digital Health Agency, is under-used by the Australian population and clinicians as it is not integrated with hospital systems. This represents a significant lost opportunity to draw together patient data. An integrated data approach is needed to keep Australia competitive in a rapidly digitalising world where automated data collection is essential and consumers are more informed and enfranchised about the use of their data.[xiv]

Unclear and complex governance arrangements – Obtaining data from public sources for research purposes is often difficult, lengthy and requires negotiating with a number of different agencies. The Issues Paper correctly canvasses a number of factors which have historically contributed to inertia in government agencies releasing public sector data.

In the experience of University researchers, obstacles emerge most frequently at jurisdictional junctures between State/Territory and Federal agencies. A complicated web of approvals, regulation and authority exists in each jurisdiction and there is no integrated authority for release. For example, cancer registries are State-based, and national data on cancer-related deaths must also be approved by each State. These arrangements mean collection of national cancer data for a research project can take years.

Expertise and skills – Publishing data requires a unique mix of skills: an understanding of the research; familiarity with the data structure; ability to engage with the community who will utilise it; and understanding how the data will be used. This mix of skills, or teams with the appropriate mix of skills, are rare.

To keep pace with future workforce demands, government, universities and industry should recognise and invest in opportunities to train a nationally distributed and skilled workforce to supply best practice data management and data publishing expertise across all sectors.

Lack of practical fit with clinical workflows – Efforts to collect and share data are often hindered by lack of professional uptake. The usability or usefulness of data based platforms in health is significantly affected by how well it supports the clinical workflow. If the use of a data collection system or tool impedes the provision of care to patients or adversely impacts clinicians or researcher's work it will not be taken up.[xv]

Insufficient variables in datasets – At present, lost opportunities occur where new national surveys omit key items that would add significant value for health researchers at negligible additional cost. As part of planning for any publicly funded data collections, data custodians should be required to conduct a formal process of consultation with researchers to explore future uses of the data to address health, technological, welfare and other social priorities, which may be brought into the design of the data collection

High costs of access to data – Accessing data is prohibitively expensive and consumes a disproportionate amount of research funding. For example, one of the University's researchers in demographics and health reports allocating a significant proportion of ARC grant funds on accessing the base data for her research; this includes $60,000 in a single transaction to the Victorian Registry for access to 6000 births, deaths and marriages certificates. Over the past 17 years, the same researcher has allocated more than $250,000 in research funds towards purchasing registration certificates from Victoria and NSW that are vital to her research work.

Inefficient or inaccessible format – In some instances, public data is available publically at the front end of the web but withheld from open or specific provision in a raw database format. This means researchers are forced to 'scrape' the web for data and laboriously compile their own research databases, which replicate the existing public databases sitting behind the public facing page.

For example, the rich data on school outcomes and performance underpinning the MySchool website is publically available as individual school records on MySchool, yet the same information as a single database format has been withheld from researchers due to concerns that it could be linked to other databases.

State-based examples of data linkage in Australia

Western Australia (WA) was the first state to initiate data linkage in the 1970s, but its use was initially limited in scope and purpose. Over the years, specific health databases, such as mortality and births were developed, giving rise to small clusters of linked data for specific populations.

In 1995, using an infrastructure grant, WA established a full-population-based data linkage system. The Data Linkage WA collection now links to eight 'core' administrative datasets (births, cancer, and deaths registries, electoral roll, emergency department records, hospital morbidity, mental health records and midwives notifications) and over 40 'infrastructure' dataset linkages for a range of approved projects.

The Data Linkage Branch in the Department of Health WA oversees a system of linkages connecting data about health events across all individuals in Western Australia. These linkages are created and maintained using rigorous, internationally accepted privacy preserving protocols, probabilistic matching and extensive clerical review. Health data can be requested for ethically approved research, planning and evaluation projects which aim to improve the health of Western Australians.

New South Wales (NSW) followed suit with the establishment of the Centre for Health Record Linkage in 2006, providing linkage services for NSW and the Australian Capital Territory. The linkage of NSW hospital data to a large-scale cohort study of individuals aged 45 years and over has been particularly valuable.

Data linkage and barriers to linkage

The Issues Paper observes that the ability to link databases could generate greatly enhanced research outcomes. This has been well demonstrated by the outcomes of past research undertakings based on data linkage. In the University's experience, data linkage projects have helped identify and prevent new hazards such as birth defects caused by certain medicines in pregnancy, blood clots following long haul air travel and cancers following CT scans in childhood.

Notwithstanding the security and privacy considerations outlined later in this submission, 'data linkage enables longitudinal health studies, epidemiological surveillance on rare diseases, the examination of variations in healthcare and the capture and comparison of healthcare costs across populations and over time. It also allows researchers to assess health outcomes on specific cohorts, such as Indigenous populations and children, who are generally difficult to recruit'.[xvi]

Data linkage circumvents loss to follow-up problems common in longitudinal studies because patients are tracked retrospectively and long-term outcomes, such as survival, can be assessed via linkage to mortality databases.[xvii]

The University notes that all Australian states have established data linkage units with the capacity to provide population-wide linked data, supported by the establishment in 2009 of the Population Health Research Network.

However, despite progress significant barriers remain that impede effective and efficient data linkage. Data is collected and linked by different organisations or units in each jurisdiction, prolonging approvals to link and access databases due to the numerous data custodians and legislative protocols.

Replication or absence of high impact research due to poor data access

Longitudinal studies on ageing: The current situation of restricted or poor access to publicly funded data in Australia necessitates researchers to undertake ad-hoc studies to gather data which should be readily available through publicly-collected sources.

For example, the Sax Institute, including leading University of Melbourne researchers, have undertaken the '45 and Up Study', which recruited more than a quarter of a million people aged over 45. The 45 and Up Study is the largest ongoing study of healthy ageing in the Southern Hemisphere and is an example of the type of long-term, wide-ranging health research that could be more readily undertaken through more efficient access to linked public health data.

Cardiovascular disease (CVD): CVD is the most costly and burdensome disease in Australia, but there is currently no overall national picture of the distribution of the disease and no way of seeing geographic, socioeconomic, or ethno-racial trends in CVD. The Australian Institute of Health and Welfare emphasised the lack of CVD data in this area in its 2010 Report 'Women and heart disease: cardiovascular profile of women in Australia'. There is a clear public benefit in compiling CVD data to enable research to predict, prevent and effectively treat heart disease.

Possible models and features to support greater access to data

The University supports efforts to enable researchers to have streamlined and secure access to public data, and to other sorts of data where appropriate, through options such as a mandated release authority or through trusted access models.

Release Authority

A 'release authority' with a national remit to act as a central broker and manager of data integrity and release is a worthwhile model for investigation. With a strong legislative foundation and appropriate powers, a release authority could negotiate requests for data with data custodians and

facilitate data release, set timelines, define security and privacy terms and work across jurisdictional boundaries to enforce protections and ensure accountability in the release of data.

The release authority could also facilitate knowledgeable, or curated, connections between the data custodian and the releasing entity to ensure that someone with hands-on knowledge about the data is involved in approving the data release.

Trusted access communities

The University notes the model outlined in the Public Sector Data Management report in 2015 of a 'linked Commonwealth dataset with a domain for secure access by government and researchers, providing a collection of linked administrative and survey data linked across time and sectors, data linkage by accredited integration authorities, and streamlined access to aggregate and individual data, while maintaining appropriate protection of individuals' records'.[xviii] There is merit in this high-level outline of a trusted access model.

The trusted access model can be distinguished from current open data mechanisms being implemented in Australia and other countries, which encourage general public access to and reuse of public data. Trusted access would enable approved researchers access to sensitive and linked public datasets using safe and agreed platforms, enabled by mutual commitment and support from participating institutions and public agencies.

There is precedent for the model of allowing researchers to apply for and use data under strict conditions within a secure data site. The United States allows such access through a set of Federal Statistical Research Data Centers, which are overseen by their equivalent of the Australian Bureau of Statistics.[xix] To access data through a RDC, researchers must apply, agree to conditions of use, pass a background check, swear under the penalty of law to maintain the confidentiality of the data, and in some cases pay for the costs of using the RDC.

A trusted access model would require shared principles and clear responsibilities and accountabilities on the parts of government bodies and researchers. Its design should aim to achieve a balance of flexible, dynamic and efficient accessibility for data users along with adequate protections for data owners and data subjects.

Other features to consider incorporating into a model for enabling data access and availability are:

Limited sample access as a guaranteed minimum access – Based on the HILDA approach (discussed further below), a standard 10 per cent sample could be created and linked to as many administrative data sets as possible. In this model, researchers would apply for that standard dataset and then have the opportunity to request more detailed data under stricter conditions.

Responsiveness to data de-aggregation – A model could be designed to enable researchers to access de-aggregated, sensitive levels of data such as line listings and data with spatial referencing. The ability to access data at greater levels of granularity would significantly affect the research potential

and impact of data. Presently, it is a common practice for data to be provided at a highly aggregated level as a means of protecting the data. A trusted access model could assist in addressing security concerns and allow release of data at a finer scale (resolution or granularity) as is necessary for research purposes.

Opt-in and opt-out processes – A data access model could have opt-out processes that ensure access by researchers is implied unless permission is withdrawn. For more sensitive data, the model might feature opt-in processes, which ensures access to data holdings is not granted unless explicitly authorised by the individual to whom the data relates.

Enabling legislation – Any data access model should be supported by nationally harmonised legislation and protocols on data availability and use, as discussed below in relation to the merits of the AIHW model. Enabling legislation could include guaranteed access to data for research uses in particular contexts, such as in pandemics or epidemics, which would reflect and maintain the public good intention of the data access framework.

Flexibility and streamlining of permissions – A data access model would contribute greater efficiency to research processes and outcomes if it is designed to accommodate flexible and evolving research inquiries and applications. A model based on one-off 'data dumps' would risk replicating some of the impediments already encountered by researchers. Currently researchers report it can take up to four years to re-do permission requests and access processes in States and Territories on account of changing research queries in a project utilising the same previously approved dataset.

Clear responsibilities – All participants in a data access model would need clearly defined roles and accountabilities. For example, clarity would be necessary on who has responsibility to keep data safe, coordinate access, inform any breaches, and develop interfaces to establish an automated and easy access Freedom of Information framework to accommodate individuals seeking to access information about the data resource.

Resourced and mandated leading agencies – The University recognises that making public data more accessible may require all sectors, including Universities, to alter local processes and practices within their respective institutions. With appropriate resourcing public agencies, such as the ARC and the NHMRC, could provide effective strategic leadership on major infrastructure and research sector advocacy on data access and availability.

Data as a gateway to public good: childhood CT scans

In 2000, there was increasing theoretical concern about the possibility of increased cancer risks following childhood exposures to computer tomography (CT) scans. These concerns were largely based on dose extrapolations from other exposures, such as those in survivors of the Japanese atomic bombs. In order to determine whether this theoretical concern was valid, a study looking at those who had actually been exposed to ionising radiation from CT scans was required. University of Melbourne Professorial Fellow John Mathews led the study.

The study proposed linking Medicare records of CT scan exposures to the national register of cancer notifications (administered by the Australian Institute of Health and Welfare (AIHW) on behalf of the states and territories, which provide the data). In 2008, the study received NHMCR funding and approval from all states and territories to access cancer data through AIHW. However, it took two years to obtain Commonwealth Department of Health approval to access de-identified Medicare data.

In 2011, the study began linking data from Medicare with the AIHW cancer records, using a unique but anonymised patient number provided by the Department of Health. The final research file combined both sets of data, but only in de-identified form. The study found cancer risk was increased by an average of 24% for persons exposed to CT scans before the age of 20 years.

However the increase was as large as 200% for brain cancers in the small number of children with CT exposures at very young ages. This ground breaking finding attracted a great deal of interest and the research publication became the most highly cited of any BMJ (formerly British Medical Journal) article in 2014.

In 2013-14 the researchers of the study advised the Commonwealth about the evidence on cancer risks and CT scan doses, especially in children, and helped develop educational materials. Professional advice also encouraged radiological practice to reduce CT procedures in children, and to reduce radiation doses for those CT procedures that were fully justified on clinical grounds. This case study highlights the difficulties researchers face without access to a strategic and responsive national data system. Had access to the data been approved sooner and the study proceeded at an earlier date, the study would have generated results sooner.

Existing data access models in Australia and overseas

A number of institutions and collaborative research frameworks already exist in Australia with transferable features for future models for expanding data availability and access.

The Australian Bureau of Statistics (ABS) and the Australian Institute of Health and Welfare (AIHW) are longstanding institutions that consolidate State and Territory data at a national level to provide

routine reports on mortality, economic and social indicators, cancer incidence and other health conditions. These agencies are designated as trusted authorities to manage linkage and de-identification where necessary. The University recommends the Productivity Commission look to these agencies as exemplars for broader data access and availability.

ABS   The ABS holds or manages a significant proportion of statistical data, including census data, Confidentialised Unit Record Files (CURFs) microdata and reference information of interest to University researchers. The University notes that the ABS is currently working collaboratively across the Australian Government to improve access to public sector data holdings and develop statistical solutions that maintain confidentiality but unlock the value of information.

The ABS works hard to make data accessible for researchers, but some of its routine privacy protection procedures may benefit from modification to enable more effective access for research purposes. For example, the ABS requires that in frequency tabulations, they will not report the exact number of individuals when the number is small, but rather give a randomised number. From the research perspective, this rule wastes important information and resources, because for many analytic purposes the most useful information comes from the estimation of effects that are influenced by cells with small numbers of individuals. In most contexts, reporting of small cell numbers will not be a threat to individual privacy.

In the contexts where risk of identification or disclosure is deemed to be high, the ABS could release the exact numbers to researchers for purposes of statistical analysis, with the requirement that the numbers not be published in that (potentially identifiable) form. As discussed in other parts of this submission, individual privacy in such instances could be further protected by appropriate obligations and professional sanctions for any attempt at re-identification or breach of privacy by end-users.

The University recognises and supports efforts underway to reduce constraints such as reducing the cost burden of access to specific data, and enhancing the malleability of ABS datasets to accommodate requests for data.

AIHW   The AIHW is a government supported national agency that provides reliable, regular and relevant information and statistics on Australia's health and welfare.

In the University's view, the AIHW working model of data management and use benefits from a strong legislative framework that resolves many of the common jurisdictional obstacles that face joint Commonwealth and States endeavours.

AIHW is well situated to consolidate State and Territory data relating to health and welfare issues, and to report relevant statistics for the nation as a whole. AIHW has recently been authorised to act as a trusted agency in providing access to and linkage with Medicare and other Commonwealth datasets.

There are a number of remaining constraints to data access in the AIHW model, which would be important to resolve if that model were to be expanded or replicated:

- Complex processes for approval when projects need access to national data. It is still necessary to obtain separate ethical and access approvals from every State and Territory. An efficient means of streamlining approvals could be to enable approvals to be granted by a release authority with national remit or trusted institution, with States and Territories retaining a power of veto, with requirement to show cause if and when veto is exercised.

- Requirement of destruction of linkage keys. A past requirement by Commonwealth Health and Medicare was that linkage keys should be destroyed at the completion of every linkage activity. This link and destroy requirement means that for longitudinal studies, the resource and time intensive linkage has to be repeated each time the evidence-base is revisited in the course of research follow-up.

- AIHW provides an important and heavily utilised national service. In addition to adequate funding to support the data release activity, the AIHW's efficiency would be greatly increased if linkage keys could be retained for potential re-use in an appropriately high-security environment.

BioGrid Australia

Along with other Melbourne Biomedical Precinct institutions and numerous research centres around Australia, the University is a collaborator and member of BioGrid Australia. BioGrid is an innovative health informatics research platform that facilitates privacy-protected research across many hospitals and medical research institutes. BioGrid provides an online Access Request System by which researchers can apply for access to specific health databases, such as cancer, diabetes, epilepsy and rare tumours.

While BioGrid manages applications for data access, data custodians retain complete control over access to their data. A Scientific Expert Review Committee assesses the proposed investigation and ethics approval from a properly constituted Human Research Ethics Committee is required before BioGrid will provide data access. Once authorised, researchers can access de-identified data for specified research.

Grhanite Health Informatics Unit (HIU)

Grhanite HIU is a major initiative of the University of Melbourne Faculty of Medicine, Dentistry and Health Sciences. Grhanite is a technological solution for gathering and hosting national healthcare data collection for audit, research and health surveillance. It has built-in capability to source data from almost all modern database technologies, including Medicare and all major GP systems used across Australia.

The Grhanite model places an emphasis on anonymity and privacy-protected record linkage as the default format for data extraction for research purposes. It also manages and documents consent processes to ensure the system can only transmit data where the appropriate permissions are in-place. This is achieved by recording opt-in consent, implementing opt-out consent mechanisms, or enabling protected data extraction using waivers of consent granted by ethics committees.[xx]

Centres for Research Excellence (CREs)

Existing research networks and collaborative research structures could be used as a base for expanding data access and availability. To an extent, this is already occurring in an ad hoc way.

For example, the NHMRC recently announced a tender for the establishment of a new Centre for Research Excellence in Infectious Disease Emergency Response Research, to provide the foundations for ongoing, coordinated, national effort against infectious diseases. The new CRE will support a multidisciplinary team which will in turn support a collaborative network of research groups and institutions across Australia undertaking infectious disease emergency response research in the health system.

The CRE selection requirements also require, amongst other criteria, the strengthened *national* capacity for infectious disease emergency response research, the creation of a strong evidence base on health actions and improved national collaborative research. Professionalised collaborative research platforms such as CREs, with secure and peer managed access arrangements, could be a useful model for unlocking sensitive public data amongst trusted committees of researchers for valuable public purpose.

HILDA   HILDA collects information about economic and subjective well-being, labour market dynamics and family dynamics. The experience and practices of the HILDA survey team at the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute) offer a proven model of how essential concerns regarding subjects' privacy and confidentiality can be balanced with researchers' desires to work with sensitive data.

The HILDA survey asks and records answers to many highly-sensitive questions, including people's incomes, earnings, drug and alcohol use, psychological problems, victimisation, and sexual orientation, under a guarantee to subjects that their responses will be kept confidential. Researchers can access several versions of the data. A 'general release' version of the data is available to researchers inside and outside Australia who sign deeds with the Department of Social Services that commit them to storing, handling, and using the data securely and to destroying the data after use.

Even with these safeguards, the general version withholds many details – such as detailed geographic codes and precise birth dates that might indirectly identify subjects. An

'unconfidentialised' version of the data that has more detailed information (but still omits direct personal identifiers) is also available to researchers who must abide by even stricter data security procedures. Thus, 'confidentialising' data along with strict licensing requirements to maintain data security provides one approach for data access.

Outside Australia, there are a number of worthwhile models for data holding and access which the University commends to the Productivity Commission for further investigation as applicable models. These include:

Canada – Cancer Care Ontario

The Cancer Care Ontario data holdings store information collected from healthcare service providers across the province. This information enables public health policy-makers and clinicians to plan and fund cancer and other healthcare services, develop guidelines, and monitor and manage the performance of the cancer and renal care systems.

Canada – Institute for Clinical Evaluative Sciences (ICES)

ICES is a not-for-profit research institute based at Sonnybrook Health Sciences Centre in Toronto, encompassing a community of research, data and clinical experts, and a secure and accessible array of Ontario's health-related data. ICES leads cutting-edge studies that evaluate health care delivery and outcomes. Through ICES researchers can access a vast and secure array of Ontario's health-related data, including population-based health surveys, anonymous patient records, as well as clinical and administrative databases.

United Kingdom – National Health Commission in Cambridge

The National Health Commission in Cambridge has a public committee that scrutinises applications for the release of data and decides what should be in the public domain.

United Kingdom – Administrative Data Research Network (ADRN)

Launched in 2014 following the report of the Administrative Data Taskforce published in 2012, the ADRN is a public data linkage program undertaken by a UK-wide partnership of universities, government departments and agencies, national statistics authorities, the third sector, funders and researchers.

The ADRN provides approvals for data access and facilitates a safe environment for accredited researchers to carry out social and economic research using linked, de-identified administrative data.[xxi] With coordination provided by the Administrative Data Service, the ADRN hosts a node in each of the four countries in the UK and receives funding through the UK Economic and Social Research Council.

Norway – Data Inspection Authority

Norway and other Nordic countries established administrative registers in the mid-1960s. Unique personal identifiers were adopted for all administrative registers, registers covering education and labour market participation, and registers for linking members of families together. Today, Norway has established ways for research communities to access this rich administrative data source, including the establishment of a national Data Inspection Authority.

For each research project, both the administrative units who administer the data as well as the national Data Inspection Authority must consent to the use and matching of registers based on a detailed application procedure. Since the registers contain sensitive and very private information, the national statistical offices link the registers for research purposes and provide identification numbers for researchers that are different from the actual administrative identification numbers.[xxii]

The data access arrangements in Norway have enabled extensive research with policy application. A small selection of examples in the education policy space include: greater understanding of children's education and labour market outcomes through child care availability; the effect of parental income on in-school education outcomes; and the effect of class size on students' academic achievement.[xxiii]

Security and standards

It is incontestable that data security must be a central concern in designing and implementing new policy and regulatory architecture for enhanced data access and availability. The risks posed to citizens and institutions by data insecurity are significant.

Public sector data which is gathered about people, but not necessarily voluntarily submitted by people for research purposes, can be seen to be even more sensitive to the risk of exposure or malpractice. All entities seeking to benefit from wider data availability, including research institutions, must be prepared to adopt practices and measures to secure data appropriately.

Med.data.edu.au provides a summary of the threshold security components that IT systems require in order to securely store and handle personal health information (including health data collected for research purposes).[xxiv]

Ethics committees and data management at the University

University researchers generate a significant volume of data in the course of their work that is often based on trials on, or information about, human and animal subjects. The University has well-established policies and processes for the collection, storage, use, re-use, access and retention of research data and records associated with their research program, including confidential research data and records. The way data is handled by the researcher and institution is a key part of ethical approval processes conducted at the outset of a research endeavour.

Researchers are required to establish and document clear procedures for the collection, ownership and storage of research data and records when involved in a joint research project, collaborative research or research undertaken in accordance with a contractual agreement.

Researchers at all universities are subject to Australian frameworks for researcher accountability and responsibilities, including the Australian Code for the Responsible Conduct of Research (ACRCR) which is set by the NHMRC in partnership with the Australian Research Council and Universities Australia. A complex raft of State and Federal legislation also covers ethics approvals as well as individual hospital or research centre ethics processes (see **Appendix B**).

The University's compliance and advice hub, the Office of Research Ethics and Integrity (OREI), is the central manager of ethics committees and processes. The OREI is currently leading a review to identify opportunities to make approval processes more agile and responsive to research timelines. The OREI is also responsible for building and maintaining an accessible library of best practice and tools for guiding research applicants through the ethics approval system.

Ethics committees processes at the University

The University has a number of ethics committees who oversee how research is conducted on human and animal subjects, and how the resulting evidence-base is stored and used.

Applications for ethics approval for research projects are assessed by institutional and/or departmental-level ethics committees comprised of a mix of senior researchers, professionals (such as veterinarians), animal or human welfare representatives and lay persons. The split roles and responsibilities between different levels of committees functions both as an efficiency mechanism and as a review mechanism for decisions. The University's Animal Welfare and Ethics Committee and Central Human Research Ethics Committee also play key roles in setting policies for ethics approvals, general standards and post-research conduct.

Ethics committees at the University seek to process approvals in as timely and efficient manner as possible. A recent analysis of the University's processes showed that applications for research on animals takes on average 45 days and applications for human research take on average 30 days. For

higher risk studies on human subjects, applications can take approximately three months to be approved.

In relation to data management, ethics committees' processes are intended to ensure that the integrity and security of research data and records is maintained, and that the research data is stored in a retrievable way. The recent analysis of ethics approval process at the University showed that data release is a consideration in approximately 20% of applications that are returned to the researcher with a query from an ethics committee.

The University's research ethics policies also take into account confidentiality restrictions, ongoing custodial responsibilities after the project is finished or the researcher has moved institutions, confidentiality agreements, copyright, privacy and intellectual property.

Areas of focus for achieving efficient research outcomes

The rapid pace of technological change and the increasingly interdisciplinary nature of research present many challenges to researchers and ethics committee members.

- *Training and information about research planning and ethics processes* – There is a key responsibility for supervisors to advise higher research degree students on the preparation of a research plan and ethics application that is low risk and will progress through the approvals process efficiently. Universities must play a key role in ensuring the information is available to students and researchers to ensure an optimal route through approvals processes, and adequate technical or advisory support from supervisors to avoid systemic inefficiencies.

- *Strong onus on researchers regarding data protection and usage* – There is already a strong onus on researchers to abide by and respect security and privacy protections. The ethical parameters for use of data in research are managed by experienced ethics committees and other governance processes (as discussed below).

  Significantly enhanced access and availability to sensitive public data, such as large-scale administrative data with non-excludable risk of disclosure or personal identification, may require a more specific onus on researchers to protect the security and integrity of data. In this case, researchers and institutions should be supported by a new national framework for managing accreditation, standards and educational materials. Additional obligations on should be matched by concurrent investment in agreed standards, mechanisms and resources for data protection.

  Requirements on researchers should come from a principled position that data security and protection of privacy is a mutual responsibility for government, researcher and institution alike.

- *Single national accreditation or access approval process* – In circumstances of public data sensitivity, a secure environment for accessing linked and anonymised administrative data could be achieved through researcher accreditation. Drawing on the features of the UK Administrative Data Research Network (discussed earlier), this might entail establishment of a single national accreditation process in Australia through which approved researchers can access linked and de-identified administrative data.

  As in the UK, the accreditation process could encompass nationally consistent training and refresher courses, 'trusted researcher' definitions, approval checks and panel membership, designation of responsibilities, disclosure protocols and sanctions recommendations.[xxv] The accreditation process could also accommodate any essential additional requirements for research access to specific administrative datasets.[xxvi]

- *Streamlining ethics approvals between multiple committees and jurisdictions, particularly for clinical trials* – Governance arrangements have been known to add complications to ethics approval processes. Despite ongoing efforts to streamline processes and leadership by the NHMRC, obstacles can arise between jurisdictions, and between institutions, particularly in relation to clinical trials.

  o *Jurisdictional* – Differing approval regimes across States and Territories can mean researchers must make repeated applications. For example, wildlife studies often cross State boundaries, meaning that researchers must make applications with licencing bodies in each relevant jurisdiction. The Australia-wide code of conduct is interpreted and applied to activities by State decision-makers, with notable divergence in the level of assurances required by different States/Territories. This means animal ethics committees at the University must scrutinise the ethical considerations of a single research application to the same level/s as necessary to satisfy the unique requirements of each State or Territory jurisdiction in which the research will be physically conducted.

  o *Institutions* – Researchers at the University often work in conjunction with partnering institutions, such as hospitals or medical clinics. While the University may be hosting the researcher, the supervision and assessment of the ethics process may be better placed in the hospital where procedures and testing will occur.

    Parallel ethics approval processes across institutions can lead to inefficiencies and delays in processing research applications. However, any streamlining must be done carefully given the exposure to risk for both the human subject and the institution relying on the external approval.

    To reduce ethics approval complications, last year the University, together with twelve other institutional members of the Melbourne Academic Centre for Health, signed a Memorandum of Understanding to implement common, streamlined and

harmonised ethical and governance review of multi-centre human health research projects. Outcomes from this ongoing process to date include: single ethical review for a large number of multi-site human research projects, including clinical trials; sharing and uptake of best practice between institutions; streamlining and harmonising of administrative forms and processes, retrospective mutual acceptance of prior scientific and ethical reviews.

- o *Agencies* – Similarly, replicated research permissions can arise where research crosses into the jurisdiction of government departments and agencies. For example, if research relies on education data concerning school-aged children, the researcher is likely to have to conduct a parallel approvals application through the Department as well as the supervising academic institution. There are also separate ethics processes situated with the NHMRC, which are relied upon in an ad hoc manner by the University when a research proposal is particularly sensitive. There is scope to streamline ethics approvals processes in this regard.

Data as a gateway to public good: discovering the link between deep vein thrombosis and long-haul air travel

Although Western Australia (WA) had been a leader for many years in linking its own health data, it had had difficulty in obtaining approval to link with Commonwealth data. In 2000, the international community was faced with growing concern about reports of deep venous (vein) thrombosis and pulmonary embolism in those who had just experienced a long-haul air-trip.

This presented an opportunity for Commonwealth officials and WA researchers to devise a study to link details of passengers arriving from incoming international flights into Perth airport (using Department of Transport data), with WA hospital admissions for deep venous thrombosis and pulmonary embolism. After access was granted, the research team was able to link the de-identified data sets collected by both the State and Commonwealth Government.

Access to this data helped reveal how the estimated absolute risk from a long haul flight was small, and the increased risk following a flight was over within 14-21 days of arrival. The study highlights the benefits of data-linkage and how access to data can help demystify public health issues. Despite the success of the study, the small linkage unit within the Commonwealth Department of Health was subsequently closed.

Privacy

As previously stated in this submission, the University and its researchers welcome the momentum towards greater access and availability of datasets, particularly public and administrative data. As the Productivity Commission Issues Paper also recognises, policy-makers need to take into account the privacy of those whose information may make up the datasets and provide the basis for an appropriate balance of all the interests involved. Ideally the starting point should be an acknowledgment that the rights of individuals are engaged.[xxvii]

Securing public trust in the use of private or publicly-held data to advance the public good is a key consideration (Deloitte, 2016). This is essential to enable better access for researchers to such data. In the research setting, privacy and security concerns become even more pressing due to the challenges of navigating different governance environments and risk scenarios created by multiple research partners who may be geographically dispersed, within and between countries.[xxviii]

Privacy laws should ideally reflect broader community standards and expectations of trust. As a minimum, individuals should have some personal control over their personal information, whether characterised in terms of 'privacy', 'data protection' or 'data privacy'. This is also the current legal position in most countries of the world, including in Australia.

As noted in the Issues Paper, Australia faces the challenge of having myriad privacy legislation, including the *Privacy Act* 1988 (Cth) at the federal level as well as diverse state and territory legislation, whose scope and limits are inconsistent and sometimes unclear (see ALRC, 2008). In addition there are common law protections of privacy, including the equitable action for breach of confidence where unauthorised use is made of confidential information disclosed or collected from an individual or group of individuals (see ALRC 2014). A recent summary of legislation, codes, policies and best practice on securing and protecting data collected from individuals for research purposes is available at med.data.edu.au.[xxix]

Personal information

If the data is not 'personal information' then the various Privacy Acts do not apply. This might be helpful for certain types of research work where anonymity may be maintained, such as where data is aggregated, as in the case of urban analytics data collected by local government and released in the form of aggregated data, with time scales for hourly parking movements or pedestrian movements. This data is aggregated in a way that means the individual cannot automatically be identified, but the data can still be used to optimise transport or service delivery.

However it may be risky simply to rely on the possibility of anonymity as a way around the requirements of the privacy statutes. Even with so-called anonymised information a person may be reasonably identifiable, for example through cross-matching across databases (see OAIC, 2015) and

thus the risks of hacking and negligent losses cannot be ruled out. In this way, information may be deemed (or become) 'personal' in the altered circumstances of linkage and usage.

Common privacy obstacles

The inconsistent and complex legislative regimes covering privacy and data-derived research go some way towards explaining why Australia has less data-sharing than elsewhere in the world, including Europe where there are exceptions for research uses in the current EU Data Protection Directive (e.g. the qualifying language in Article 6(b) attached for reference at **Appendix C**).[xxx]

The EU Data Protection Directive may be usefully compared to Australian data sharing provisions, such as the provision made for health research in sections 95 and 95A *Privacy Act* (with approved NHMRC Guidelines), and the rather narrowly focused research exceptions in other Acts such as the *Health Records Act* 2001 (Vic) (HPPs 1.1(e) and 2.2(g)) and the *Privacy and Data Protection Act* 2014 (Vic) (IPPs 2.1(c) and 10.2)).

The scope of the available defence to breach of confidence is also unclear in Australia (see, for instance, the discussion of whether the true scope of the defence is the 'public interest' or rather more particularly 'iniquity' in *Australian Football League v The Age* (2006)).

Further, while consent may be a way to comply with many of the requirements of the Acts, and to obviate an action for breach of confidence, this can be an onerous (albeit not necessarily impossible) standard where many individuals are involved. The OAIC's APP Guidelines provide that consent must be clear and unequivocal, should not be bundled, and that for sensitive information 'generally express consent should be sought' (2014, pp 8-11).

Addressing the challenges: balancing data availability with privacy protections

National uniform privacy laws – Privacy laws in Australia are inconsistent. A preferable model would see uniform and coherent national and State and Territory legislation. As noted in the Issues paper, the significant complexity caused by having multiple pieces of legislation (plus common law) governing privacy was observed by the Australian Law Reform Commission's review of privacy laws in Australia in 2008 and 2014. At the same time there is still scope to update the legislation to provide for stronger transparency and control mechanisms appropriate to modern technologies and practices.

Tailored defence for research use – A tailored defence for research use in the relevant Acts would be helpful for researchers and their institutions. This could include certain safeguards such as reasonable attempts at anonymisation or pseudonymisation combined with some other standards

that apply regardless of such attempts and bearing in mind that there may be a range of scenarios where individuals may be harmed by data breaches (see Tene and Polonetsky, 2016).

Similarly, the legal framework could establish 'safe harbour' provisions to facilitate data-centric innovation in specified secure circumstances so new ideas can be tested and experimentation in data-driven products and services may be undertaken within the bounds of the law.

A possible model here might be article 89 of the new EU Data Protection Regulation 2016 (framed in more elaborated and up-to-date terms than the current Directive), which will come into force in 2018.[xxxi]

Complementary public interest defences – At the same time the relevant Acts could provide for an equivalent research defence to breach of confidence or a public interest defence drafted in sufficiently broad terms to encompass research uses. Thought might also be given to the available defences in the *Copyright Act* 1968 (Cth) (for instance, whether s 40 covering 'research' is sufficient: see ALRC (2013), ch 7 and generally ALRC (2014); and whether there should be a defence for data mining: see ALRC (2014), ch 11).

Restrictions on appropriate use of data by all potential users – Much of the concern with protecting privacy and confidentiality of data focuses on personal data maintained by the government being disclosed to others. However, there is also a reciprocal concern about sensitive data collected by research organisations, such as the HILDA collected by the Melbourne Institute, being disclosed to the government. To address these concerns, especially in the context of data linkages, there should be restrictions that these types of data either will not be used by government agencies or will only be used for research purposes with strict conditional requirements about privacy and sanctions for misuse.

Data and the wider community

Engaging the public – Fundamental to the concept of consent is the expectation that research participants are able to make informed and voluntary decisions about their research participation. The conditions for consent are well-established in research ethics, however there are increasing challenges facing researchers regarding consent specific to using digital data.

The significant growth in the collection of health-related data by a range of stakeholders emphasises the need for consumers to be able to have oversight and control of who can see, update and use their information.

The University supports robust and transparent processes to support consumer information, ownership of their health data and ability to make choices that effect their health and healthcare.

This is vital throughout the lifecycle of patient care, from effective preventative strategies to palliative care.[xxxii]

One means of doing this may be the creation of readily-accessible interfaces for individuals and organisations to check information held about them across public and private databases.

Engaging professionals – Clinical engagement is well-documented as pivotal in bringing about change in health care and practise, including driving successful eHealth initiatives. Support for data reforms will be contingent on wide consultation and engagement across the clinical specialities.

Building skills for a data-capable research workforce – Education and professional development about the benefits and safe practices of data-based research should go hand-in-hand with the maturing of the technology that supports greater data access and availability.

The University has a key role in training PhD students and academics with respect for the integrity and security of data, as well as data curation and collection skills. The University is committed to playing its role in fostering a research culture that embeds and recognises the importance of security, ethical and professional standards in relation to data availability and access.

Notes

[i] Submission by the Melbourne Academic Centre for Health, 'The AHRTCs and a Digital Backbone: the lifeline for biomedical research and healthcare', Australian Medical Research and Innovation Five Year Strategy Review.

[ii] Karin Clark et al, Guidelines for the Ethical Use of Digital Data in Human Research, Carlton Connect Initiative, University of Melbourne 2015, available online at http://www.carltonconnect.com.au/wp-content/uploads/2015/06/Ethical-Use-of-Digital-Data.pdf

[iii] https://theconversation.com/bringing-the-war-home-the-rising-disability-claims-of-afghanistan-war-vets-56021

[iv] The Commonwealth Fund, *International Health Care System Profiles* available at http://international.commonwealthfund.org/features/ehrs/.

[v] Submission by the Melbourne Academic Centre for Health.

[vi] For more detail, see http://med.data.edu.au/about/ and in particular, the *Discussion Paper on Legal, Best Practice and Security Frameworks for consideration in operation of the Australian National Medical Research Data Storage Facility* available at http://med.data.edu.au/wp-content/uploads/2016/01/Med.Data_Legal_Security_Framework_Jan2016.pdf.

[vii] Royal Australasian College of Physicians, Submission on the draft National Digital Health Strategy for Australia, May 2016.

[viii] Cord blood is the blood left behind in the placenta and umbilical cord after birth and a rich source of blood-forming stem cells.

[ix] Submission by the Royal Australasian College of Physicians.

[x] Submission by the Royal Australasian College of Physicians.

[xi] Submission by the Royal Australasian College of Physicians.

xii See http://aurin.org.au/blog/2016/06/20/are-we-planning-adequate-healthcare-for-the-coming-ageing-boom/.

xiii Examples of Data Hub projects can be accessed here: http://aurin.org.au/projects/data-hubs/.

xiv Submission by the Melbourne Academic Centre for Health.

xv Submission by the Royal Australasian College of Physicians.

xvi Tew, Dalziel, Petrie and Clarke, *Growth of linked hospital data use in Australia: a systematic review*, Australian Health Review (unpublished: received 5 Feb 2016, accepted for publishing 5 June 2016).

xvii Tew, Dalziel, Petrie and Clarke, *Growth of linked hospital data use in Australia: a systematic review.*

xviii Australian Government Department of the Prime Minister and Cabinet, Public Sector Data Management report 2015, available online at https://www.dpmc.gov.au/sites/default/files/publications/public_sector_data_mgt_project.pdf.

xix See more on United States RDCs, http://www.census.gov/about/adrm/fsrdc/about.html

xx More information about Grhanite is available at http://www.grhanite.com/.

xxi For further information on the Administrative Data Resource Network see https://adrn.ac.uk/; see also *The UK Administrative Data Research Network: Improving Access for Research and Policy,* Report from the Administrative Data Taskforce (Economic and Social Research Council in collaboration with the Medical Research Council and the Wellcome Trust) December, 2012 (see pages 20-23); UK Government's response available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206873/bis-13-920-government-response-administrative-data-taskforce.pdf

xxii David N. Figlio, Krzysztof Karbownik, Kjell G. Salvanes. *Education Research and Administrative Data*; Working Paper 21592, National Bureau of Economic Research, September 2015.

xxiii David N. Figlio, Krzysztof Karbownik, Kjell G. Salvanes. *Education Research and Administrative Data*.

xxiv Available at http://med.data.edu.au/personal-health-data-and-research/#it_system_requirements.

xxv *The UK Administrative Data Research Network: Improving Access for Research and Policy*.

xxvi For further information on UK ADRN 'trusted researcher' definition and approval process see https://adrn.ac.uk/protecting-privacy/trusted-researchers; for further information on approvals panels and process see https://adrn.ac.uk/application-process/approvals-panel.

xxvii See Goold 2015; Richardson 2015.

xxviii Karin Clark et al, *Guidelines for the Ethical Use of Digital Data in Human Research*.

xxix See resources and summary tables at http://med.data.edu.au/personal-health-data-and-research/#legislation.

xxx See also the more elaborated provision in the *Data Protection Act* 1998 (UK) section 33.

xxxi See Maldoff, 2016.

xxxii Submission by the Royal Australasian College of Physicians.

Bibliography

Australian Government Department of the Prime Minister and Cabinet, *Public Sector Data Management report 2015*, available online at https://www.dpmc.gov.au/sites/default/files/publications/public_sector_data_mgt_project.pdf.

Australian Law Reform Commission, *For Your Information: Australian Privacy Law and Practice* (ALRC Report 108), available online at http://www.alrc.gov.au/publications/report-108

Australian Law Reform Commission, *Copyright and the Digital Economy* (DP 79), 2013, available online at http://www.alrc.gov.au/publications/copyright-and-digital-economy-dp-79.

Australian Law Reform Commission, *Copyright and the Digital Economy* (Report 122), 2014, available online at http://www.alrc.gov.au/publications/copyright-report-122.

Australian Law Reform Commission, *Serious Invasions of Privacy in the Digital Era* (Report 123), 2014, available online at https://www.alrc.gov.au/publications/serious-invasions-privacy-digital-era-alrc-report-123.

Karin Clark et al, *Guidelines for the Ethical Use of Digital Data in Human Research*, Carlton Connect Initiative, University of Melbourne 2015, available online at http://www.carltonconnect.com.au/wp-content/uploads/2015/06/Ethical-Use-of-Digital-Data.pdf

Philip Clarke, 'Bringing the war home: the rising disability claims of Afghanistan war veterans', *The Conversation*, published March 17 2016, available online https://theconversation.com/bringing-the-war-home-the-rising-disability-claims-of-afghanistan-war-vets-56021

Deloitte Australian Privacy Index 2016, available online at http://www2.deloitte.com/au/en/pages/risk/articles/deloitte-australian-privacy-index-2016.html.

David N. Figlio, Krzysztof Karbownik, Kjell G. Salvanes. *Education Research and Administrative Data*; Working Paper 21592, National Bureau of Economic Research, September 2015.

Ben Goold, 'Ticking Boxes and Telling Stories: Privacy, Data Sharing, and the Construction of Identity', public lecture, Melbourne Law School, 2015 (copy on file and can be made available on request).

Gabe Maldoff, 'How GDPR Changes the Rules for Research', *The Privacy Adviser* April 19 2016, available online at https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/.

National Health and Medical Research Council Guidelines approved by the Office of the Australian Information Commissioner under section 95 of the Privacy Act, available online at https://www.oaic.gov.au/agencies-and-organisations/legally-binding-guidelines-and-rules/guidelines-under-section-95-of-the-privacy-act-1988-2014 and https://www.oaic.gov.au/agencies-and-organisations/legally-binding-guidelines-and-rules/guidelines-under-section-95a-of-the-privacy-act-1988-2014.

Office of the Australian Information Commissioner, Australian Privacy Principles Guidelines, 2014, available online at https://www.oaic.gov.au/images/documents/privacy/applying-privacy-law/app-guidelines/APP-guidelines-combined-set-v1.pdf.

Office of the Australian Information Commissioner, *Consultation draft: Guide to big data and the Australian Privacy Principles*, May 2016, available online at https://www.oaic.gov.au/engage-with-us/consultations/guide-to-big-data-and-the-australian-privacy-principles/consultation-draft-guide-to-big-data-and-the-australian-privacy-principles.

Regulation (EU) 2016/… of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), available online at http://data.consilium.europa.eu/doc/document/ST-5419-2016-INIT/en/pdf.

Med.data.edu.au, *Discussion Paper on Legal, Best Practice and Security Frameworks for consideration in operation of the Australian National Medical Research Data Storage Facility*, January 2016 available at http://med.data.edu.au/wp-content/uploads/2016/01/Med.Data_Legal_Security_Framework_Jan2016.pdf.

Megan Richardson, 'The Battle for Rights: Getting Data Protection Cases to Court' (2015) 2 *Oslo Law Review* 23, available online at https://www.journals.uio.no/index.php/oslawreview/article/view/2350

Royal Australasian College of Physicians, Submission on the draft National Digital Health Strategy for Australia, May 2016.

Omar Tene and Jules Polonetsky, 'Beyond IRBs: Ethical Guidelines for Data Research' (2016) 72 *Washington & Lee Law Review* , available online at http://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1044&context=wlulr-online.

Tew, Dalziel, Petrie and Clarke, *Growth of linked hospital data use in Australia: a systematic review*, Australian Health Review (unpublished: received 5 Feb 2016, accepted for publishing 5 June 2016).

Melbourne Academic Centre for Health, 'The AHRTCs and a Digital Backbone: the lifeline for biomedical research and healthcare', Submission to the Australian Medical Research and Innovation Five Year Strategy Review (unpublished).

UK Administrative Data Taskforce (Economic and Social Research Council in collaboration with the Medical Research Council and the Wellcome Trust), *The UK Administrative Data Research Network: Improving Access for Research and Policy*, report from December, 2012.

Appendix A – Submission contributors

In alphabetical order (by surname)

Dr. Gideon Aschwanden, Melbourne School of Design

Dan Barr, Acting Director, Office of Research Ethics and Integrity, University of Melbourne

Prof. Jim Bishop, Professor and Executive Director, Victorian Comprehensive Cancer Centre

Prof. Lorena Brown, Department of Microbiology and Immunology, School of Biomedical Sciences, Faculty of Medicine, Dentistry and Health Sciences

Karin Clark, Senior Fellow, Melbourne Law Masters

Prof. Philip Clarke, Professor of Health Economics, Centre for Health Policy, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences

Dr. Serryn Eagleson, Data Hubs Leader, AURIN

Kelly Farrow, Advisor, Policy and Government Relations

Prof. Lynn Gillam, Chair, Central Human Research Ethics Committee

Anne Kelly, Research Collaboration Project Officer

Prof. Lars Kulik, Department of Computing & Information Systems, Melbourne School of Engineering

Prof. Marcia Langton, Foundation Chair in Australian Indigenous Studies, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences

Prof. Chris Leckie, Department of Computing & Information Systems

Dr. Steven Manos, Research Computation Strategy and Digital Research Platforms

Prof. John Mathews, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences

Prof. Janet McCalman, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences

Prof. James McCluskey, Deputy Vice-Chancellor Research

Prof. Jodie McVernon, Professor and Director of Doherty Epidemiology, The Peter Doherty Institute for Infection and Immunity, The University of Melbourne & Royal Melbourne Hospital, Head, Modelling & Simulation Unit, Centre for Epidemiology & Biostatistics, Melbourne School of Population & Global Health and Murdoch Children's Research Institute

A\Prof. Hans Meij, Executive Director, Melbourne Academic Centre for Health

Prof. Thas Nirmalathas, Melbourne Networked Society Institute

Prof. Terry Nolan, Melbourne School of Population & Global Health, Faculty of Medicine, Dentistry and Health Sciences

Prof. David Ribar, Professorial Research Fellow, Acting Director of the Melbourne Institute & Director of the Economic and Social Disadvantage Program, Melbourne Institute of Applied Economic and Social Research

Prof. Megan Richardson, Melbourne Law School

Dr. Rebecca Ritte, Research Fellow (Epidemiology), Onemda VicHealth Koori Health Unit, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences

Prof. Tony Scott, Melbourne Institute of Applied Economic and Social Research

Prof. Arthur Shulkes, Chair, Animal Welfare and Ethics Committee

Prof. Richard Sinnott, Director of e-Research (University of Melbourne) and Professor of Applied Computing Systems (Melbourne School of Engineering)

Prof. Liz Sonenberg, Pro Vice-Chancellor Research Collaboration & Infrastructure

Emma Williams, Australian Urban Research Infrastructure Network (AURIN)

Carlene Wilson, Director, Policy and Government Relations

Appendix B – Legislation and codes on research ethics

Legislation and Codes of Conduct relating to research at the University of Melbourne include:

University of Melbourne

- University Code of Conduct for Research
- Animal Care and Use Standards
- Animal Use and Welfare Guidelines
- University of Melbourne's Animal Ethics Committee Terms of Reference
- Human Ethics Review Structure at the University of Melbourne

State (Victorian)

- Victorian Acts and Regulations relevant to animal welfare, including the *Prevention of Cruelty to Animals Act (and Regulations)*, the *National Parks Act*, the *Wildlife Act* and the *Fisheries Act.*
- Victorian Codes of Practice for Animal Welfare

National

- Australian Code for the Responsible Conduct of Research
- Australian Code for the care and use of animals for scientific purposes 8th edition (2013)
- National Statement on Ethical Conduct in Human Research
- Australian Acts and Regulations including the *Gene Technology Act* 2000 and *Gene Technology Regulations* 2001

Appendix C – Legislative privacy protection examples

DIRECTIVE 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

SECTION I – PRINCIPLES RELATING TO DATA QUALITY

Article 6

1. Member States shall provide that personal data must be:

(a) processed fairly and lawfully;

(b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. **Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards**; [*University's emphasis*]

(c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;

(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use.

2. It shall be for the controller to ensure that paragraph 1 is complied with.

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Article 89

Directive 95/46/EC provided for a general obligation to notify the processing of personal data to the supervisory authorities. While that obligation produces administrative and financial burdens, it did not in all cases contribute to improving the protection of personal

data. Such indiscriminate general notification obligations should therefore be abolished, and replaced by effective procedures and mechanisms which focus instead on those types of processing operations which are likely to result in a high risk to the rights and freedoms of natural persons by virtue of their nature, scope, context and purposes. Such types of processing operations may be those which in, particular, involve using new technologies, or are of a new kind and where no data protection impact assessment has been carried out before by the controller, or where they become necessary in the light of the time that has elapsed since the initial processing.

Article 90

In such cases, a data protection impact assessment should be carried out by the controller prior to the processing in order to assess the particular likelihood and severity of the high risk, taking into account the nature, scope, context and purposes of the processing and the sources of the risk. That impact assessment should include, in particular, the measures, safeguards and mechanisms envisaged for mitigating that risk, ensuring the protection of personal data and demonstrating compliance with this Regulation.