



**SCHOOL OF PUBLIC HEALTH AND PREVENTIVE  
MEDICINE**

**FACULTY OF INFORMATION TECHNOLOGY**

**Submission to Productivity Commission  
Inquiry into Data Availability and Use**

**July 2016**

***Inquiry into the benefits and costs of options for increasing availability of and improving the use of public and private sector data by individuals and organisations.***

**Executive Summary**

The health and social benefits of increased access to public sector data for research are not fully recognised. There would be economic benefits from increased availability to public data for research. There are substantial barriers to the use of public sector data. However there is evidence that the public is often less concerned about privacy issues than some governments perceive. Best practice data linkage, such as that conducted at the Farr Institute in the UK should inform future Australian Government investment. We have responded to selected questions on high value public sector data, the collection and release of public sector data, data linkage, resource costs of access, privacy protection and data security.

**General Considerations**

Australia has some of the best data systems globally, yet has fallen behind other developed countries in using data for public good research. The central issues appear to be:

1. The ***health and social benefits*** of increased access to public sector data for research are not fully recognised:
  - Timely access to big public data, especially when linked to other datasets, including surveys, allows important research questions to be answered, including about vulnerable and disadvantaged groups, with consequent health, social and economic savings.
  - If linked to health and welfare data, a wide range of public data from many sectors could contribute to answering important research and policy questions. These linked data not only identify public health priorities, but also disparities in health outcomes and health service use, and help develop universal and targeted health promotion efforts.
  - Such innovation in research is currently largely stifled in Australia compared with many other developed countries.
  
2. ***Economic benefits*** of increased availability to public data for research:
  - Potential efficiencies through evidence based policies, the availability of monitoring tools, the ability to evaluate practices and interventions, including cross-sectoral approaches which might have benefits for health, but are not initiated within the health sector (such as in social welfare, criminal and civil justice systems, etc).
  - Opportunity cost currently suffered by Australia in not using data for public good research.
  - Australian research in this area is non-competitive internationally and Australian researchers often leave to conduct research in countries with better data access. They may not return.
  - Feedback from data users on data shortcomings would have the potential to enhance current public sector data. For example, the School's Jean Hailes Research Unit has recently completed a technical review for DHHS about access to services for termination of pregnancy. The public sector data were very limited and of generally poor quality.
  - The enduring use of linked data for multiple studies (always subject to ethical and privacy considerations) would dramatically reduce start-up costs and provide time efficiencies in research and subsequent improvements in population health. The current requirement to delete datasets on the completion of studies is counter-productive (Issues Paper, p22).

3. **Barriers** to use of public sector data
  - Huge delays (e.g. 5 years) in access to Commonwealth health sector data for research result in waste of research dollars or failure to complete research.
  - Privacy concerns can be reduced by the use of de-identified data and secure data linkage processes. The public is probably unaware that public data are not available for important research, and there is evidence that they are less concerned about privacy issues than some governments perceive. There is little empirical research on what the public thinks about privacy protection in the context of health research. Unpublished research undertaken by the Monash Centre for Occupational & Environmental Health and the Michael Kirby Centre indicates that some groups, such as parents of disabled children, assume that data concerning their children is being linked for the purposes of health research and are disappointed to learn that it is not. Participants from a number of different population groups are happy for their data to be made available for research as long as the research is not-for-profit and the researchers are employed by reputable research institutions.
  - Some privacy-related legislation is out-of-date given the reduced privacy risk associated with modern, secure data linkage methods.
  - Much of the important national data are collected by states and territories and may not be routinely pooled nationally. Inappropriate barriers, such as use of the data as a leveraging tool, may delay the availability of national data. States and territories should sign up to the Australian Government's *Public Data Policy Statement* (Issues Paper, p13). We wonder what incentives would be necessary to achieve this.
  - The current requirement for state-by-state approval for use of pooled national data causes substantial delays and costs.
  - Inefficiencies in the actual data linkage process by authorised agencies result in queuing, uncertainty and reluctance to use linked data (e.g. for time constrained PhD studies).
  - Some data custodians are reluctant to release data for research. Positive data release policies are needed which are supported by appropriate KPIs and public reporting requirements.
  - The lack of unique identifiers for individuals in Australian data systems diminishes the utility of data.
  - The lack of consistency between data sets and incomplete data reduce the benefits of data linkage, though data linkage projects will highlight potential enhancements and drive improvements in data validity and quality.
  
4. Assessment of **best practice data linkage models** (particularly enduring models) from elsewhere should inform government investment to achieve positive social and economic returns.

We have chosen to respond only to selected questions canvassed in the Commission's Issues Paper of April 2016.

## **QUESTIONS ON HIGH VALUE PUBLIC SECTOR DATA**

*What public sector datasets should be considered high-value data to the: research sector; academics?*

In our view, the high-value datasets include health data, census data, the electoral roll, welfare data, justice data, education data and environmental data. Some specific examples include national resources maintained by the Australian Institute of Health & Welfare, such as the National Death Index and Australian Cancer Database. Although not originally established for research purposes, the Medicare and Pharmaceutical Benefits Schedule databases have also proven to be valuable national resources. There are also many valuable datasets maintained at a state level, such as hospital admissions and emergency presentations, ambulance attendances, cardiac arrest registry, cancer registries, mental health, drug and alcohol treatment, disability data, welfare, justice and educational datasets, to name a few.

*What characteristics define high-value datasets?*

The defining characteristics include comprehensive population coverage, the nature of the data, high data quality with accessible data dictionaries, social determinants of health and disease. There also needs to be an appropriate governance structure.

*What benefits would the community derive from increasing the availability and use of public sector data?*

All of these data have been collected with public funds. They can be used to answer important research questions, improve the health of the community, and add value or enhanced use for the greater public good. A good example, and one which demonstrates the benefits of access to very large government datasets is a study by Mathews and colleagues at the University of Melbourne of Cancer risk in 680,000 people exposed to computed tomography (CT) scans in childhood or adolescence (1). The investigators conducted a data linkage study of 11 million Australians which linked Medicare data with cancer registries. They found a significantly increased incidence of cancer among those who had CT scans, which increased further with each additional scan. They concluded that future CT scans should be limited to situations where there is a definite clinical indication, and every scan optimised to provide a diagnostic image at the lowest possible radiation dose.

## **QUESTIONS ON COLLECTION AND RELEASE OF PUBLIC SECTOR DATA**

*What are the main factors currently stopping government agencies from making their data available?*

These factors include privacy laws and the barriers between jurisdictions. The practical issues that need to be overcome were examined in a proof of concept project by Mitchell et al (2). They concluded that the current complexity, duplication and lack of cohesion undermines any attempts to conduct research involving national record linkage in a timely manner. A table in this paper sets out the wide disparities in data linkage processes, with some States requiring up to 9 distinct stages of approval.

*How could governments use their own data collections more efficiently and effectively?*

While governments collect data, they do not necessarily analyse it optimally to inform public policy and importantly to evaluate practices and programs. They need to increase their in-house capacity to conduct these functions and/or engage with universities to conduct research to inform policy. There are also issues around transparency and public perception of “in-house” analyses.

*Should the collection, sharing and release of public sector data be standardised? What would be the benefits and costs of standardising? What would standards that are 'fit for purpose' look like?*

Standardisation would greatly improve the efficiency of data release and data linkage. Currently the Australian Institute of Health & Welfare (AIHW) requires separate approvals from each jurisdiction for each project. A better system would be if the data custodians released data to platforms, including large scale e-cohorts, such as those at the Farr Institute (see below) rather than project by project.

It has been a significant problem that the perinatal data are collected on a state-by-state basis, with some difference in fields and delays between collection at a state level and then integration at national level (can be up to five years). The Australian population is small enough for this to be done nationally with great improvements in efficiency and outputs.

Unfortunate practices have emerged in recent years where ad hoc non-standardised registries have been established in some, but not all, states and territories in such domains as suicide and family violence. Advice to the Commonwealth government on setting nationally agreed data sets for such registries at the earliest stages of these developments were ignored. Currently, there remains no clear national understanding of such issues to inform policy directions. Clearly, standardisation to international standards, where they exist, or otherwise nationally agreed standards for data collections would be optimal.

*What criteria and decision-making tools do government agencies use to decide which public sector data to make publicly available and how much processing to undertake before it is released?*

The decision making process often seems capricious to researchers, the criteria are opaque and rely too heavily on the individual discretion of public servants. A good example would be the chief legal officer of the Australian Electoral Commission, who has repeatedly questioned the ability of Human Research Ethics Committees constituted following NHMRC guidelines and including legal practitioners, to correctly apply Australian Privacy Principles (3).

*What specific government initiatives (whether Australian Government, state, territory or local government, or overseas jurisdictions) have been particularly effective in improving data access and use?*

A particularly effective initiative in the United Kingdom has been the Farr Institute of Health Informatics Research (<http://www.farrinstitute.org/>). The Farr Institute works to advance informatics, statistics and data science and to develop governance and infrastructure frameworks that underpin the safe and trusted use of patient data. The Farr Institute works with the owners and custodians of data to support the safe and trusted sharing and linkage of datasets and the establishment of research data platforms. For example, the Secure Anonymised Information Linkage (SAIL) system (<http://www.saildatabank.com>) is a fundamental component of the Centre for the Improvement of Population Health through eRecords Research (CIPHER), one of the four coordinating centres of the Farr Institute. Population health, social and other datasets are linked to form research platforms with coordinated and streamlined governance processes, established custodian approvals, and full compliance with data protection legislation. The research platforms include population dynamic electronic cohort studies, disease registries and randomised controlled trials. There is a 2-stage approval process with researchers contacting SAIL in the first instance to establish the scope and feasibility of the project, and then submission of a project application to the

Information Governance Review Panel (IGRP) who provide independent advice on Information Governance and review all proposals to use SAIL data to ensure that they are appropriate and in the public interest.

#### **QUESTIONS ON DATA LINKAGE**

*Which rules, regulations or policies create unnecessary or excessive barriers to linking datasets?*

The barriers include a lack of national identifiers, idiosyncratic interpretations of privacy by public servants, the burden of making multiple ethics committee applications and inconsistent rulings by these committees.

*How can Australia's government agencies improve their sharing and linking of public sector data?*

We understand that the Australian Digital Health Agency has recently provided MBS and PBS data to the States. They are clearly working towards making more health data available for the public good and to improve the use of the e-health record. They are also well aware of the AMA's expressed concern about patients restricting access to e-health record data. At least this Agency recognises the Government's efforts to permit health and other data to be used to improve the health of the community. We would like the Productivity Commission to consider this example and help solve the sharing and linking of public sector data, provided patient privacy and confidentiality are not compromised.

*What lessons or examples from overseas should be considered?*

The British perinatal data collection system, now known as 'MBRRACE-UK: Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK' has been incredibly successful in addressing adverse perinatal outcomes, risks associated with social and economic disadvantage and identifying previously unascertained causes of maternal and newborn deaths. It is a UK wide scheme and is able to elucidate rare but catastrophic outcomes, because of the size of the dataset and the consistency, including standardised protocols with which it is collected.

The Farr Institute involves 21 academic institutions and health partners in England, Scotland and Wales, and is publically funded by a consortium of ten organisations led by the Medical Research Council, to deliver high-quality, cutting-edge research using 'big data' to advance the health and care of patients and the public. The Farr Institute is not only producing world-leading, high quality health care research but has invested heavily in data innovations in methods for privacy protecting data linkage, storage and access, new statistical approaches and algorithms for big data, and improving data quality. Examples showing the value of this approach are too numerous to list individually, but some recent publications from the Farr Institute have examined the prognosis of acute myocardial infarction(4) and anti-platelet therapy in coronary artery disease(5). Such studies would be very difficult if not impossible to conduct in Australia.

We have limited experience of working with private sector data and offer no specific comments.

## **QUESTIONS ON RESOURCE COSTS OF ACCESS**

*How should the costs associated with making more public sector data widely available be funded?*

*To what extent are data-related resources in agencies being directed towards dealing with data management and access issues versus data analysis and use?*

*What pricing principles should be applied to different datasets? What role should price signals play in the provision of public sector data?*

The way of the future is likely to be user pays, but this will require greater transparency of the costs and understanding by funding bodies of legitimate research costs – currently there is no consistency around this. With respect to making the data available to users, privacy protecting data safe havens will need to be developed. At the moment there is only SURE which is too expensive for most researchers. Another option is on-site analysis which limits the capacity for collaboration. One of the best things about SAIL and its GATEWAY (safe haven) is that researchers can work on projects from anywhere in the world.

*Is availability of skilled labour an issue in areas such as data science or other data-specific occupations? Is there a role for government in improving the skills base in this area?*

Availability of high level skills in data management and extraction is an issue. There has been serious under-investment in many jurisdictions over a long time in training and retaining appropriately qualified staff. The Universities are well placed to provide training in biostatistics, data management, informatics etc.

## **QUESTIONS ON PRIVACY PROTECTION**

*What types of data and data applications (public sector and private sector) pose the greatest concerns for privacy protection?*

Research undertaken Monash Centre for Occupational & Environmental Health and the Michael Kirby Centre indicates, not surprisingly, that there is heightened sensitivity around data concerning sexual health and mental health.

*How can individuals' and businesses' confidence and trust in the way data is used be maintained and enhanced?*

We have found that the public form their beliefs about data use for health research purposes from their experience with health service provision. Confidence in medical research could be improved by promoting knowledge about the steps that are taken to protect personal data in research. This includes formal research governance requirements, ethical guidance and scrutiny of research by ethics committees.

*What weight should be given to privacy protection relative to the benefits of greater data availability and use, particularly given the rate of change in the capabilities of technology?*

People will express a variety of views in response to the question. Instead of relying on individual opinions, it would be preferable to form a view that is the product of research and is evidence-based.

*Are further changes to the privacy-related policy framework needed? What are these specific changes and how would they improve outcomes? Have such approaches been tried in other jurisdictions?*

*How could coordination across the different jurisdictions in regard to privacy protection and legislation be improved?*

Rather than attempting to improve coordination, which is bound to fail, the same standards should be adopted in all jurisdictions.

#### **QUESTIONS ON DATA SECURITY**

The power of Data Science to generate novel and useful medical hypotheses, from a wide range of disparate data sources, has been recognised more and more in recent years. Relevant data may include information at the molecular level, such as genetic information and chemical markers, as well as macro information such as diet and lifestyle, and levels in between (such as heart-rate and body temperature). Pulling such data together from different sources within a single hospital, or from across the nation, leads to novel insights that drive medical discoveries. Australian Researchers are world leaders in Data Science, Data Wrangling, Knowledge Discovery and also in the CyberSecurity techniques necessary to ensure the required quantity and breadth of data can be provided to researchers without breach of privacy.

*Are security measures for public sector data too prescriptive? Do they need to be more flexible to adapt to changing circumstances and technologies?*

From a data security point of view, it is essential to always have adequate security measures in place to ensure that privacy, authenticity and integrity requirements are satisfied. However, rather than regulating data security by prescribing the use of particular technologies, one should define security requirements that need to be satisfied. This would enable the use for new or future security mechanisms. Advanced cryptographic mechanisms, for example, can substantially improve data security as well as enhance privacy. Approaches like searchable encryption or homomorphic encryption enable computations over encrypted data. Thus, data can be encrypted at the sources and only the results of statistical evaluation or data aggregation is decrypted. The original data stays encrypted. The use of such technologies can remove the need of explicit anonymisation of data.

The process of providing researchers access to data sources could be improved and accelerated with clearly defined technical processes using adequate security measures. Changing technologies in collection of data (Internet of Things, mobile sensors, etc.) as well as the distribution of resources require to frequently adapt the mapping of the administrative and formal processes to actual technical realisations. This includes a mapping of security requirements to adequate security mechanisms. By integrating mechanisms to enforce limitations of access to what is actually needed for particular research tasks, effects of breaches can be minimised. Furthermore, by using advanced security technology, the risk that comes with providing access to sensitive data becomes containable. Thus, decision processes for granting access to data collections could be accelerated.

Security technology should be seen as an enabler rather than as restricting use of data sources for research. However, advanced security technology and secure interfaces to data collections need to be designed into the systems and also be reflected in regulations and administrative processes.



## Authors & Affiliations

Michael Abramson<sup>1</sup>, Jane Fisher<sup>1</sup>, Belinda Gabbe<sup>1</sup>, Danny Liew<sup>1</sup>, Bebe Loff<sup>1</sup>, John McNeil<sup>1</sup>, Joan Ozanne-Smith<sup>1,2</sup>, Carsten Rudolph<sup>3</sup>, Mark Wallace<sup>3</sup>, John Zalcborg<sup>1</sup>

1. School of Public Health and Preventive Medicine, Monash University
2. Victorian Institute of Forensic Medicine, Melbourne
3. Faculty of Information Technology, Monash University

With diverse leadership across 4 campuses, the **School of Public Health and Preventive Medicine** is Monash University's principal source of skills in epidemiology (including clinical epidemiology), biostatistics and large scale clinical data-management. The School emphasises expertise in large epidemiological studies, multi-centre clinical trials, clinical registries, evidence synthesis, health social sciences and forensic medical investigations. Continued collaborative work with the major Monash affiliated hospitals, research institutes and public health units within Victoria ensures the School provides a key state, national and international resource underpinning translational research.

The School is the home of a substantial number of important clinical registries. These include clinical quality, condition / disease and drug / device / product registries. Further details can be found at <http://www.med.monash.edu.au/sphpm/depts-centres-units/registries/index.html>. The experience of running these registries and conducting large scale epidemiological studies is the particular perspective that we bring to this submission.

The School's Department of Forensic Medicine is collocated with the **Victorian Institute of Forensic Medicine**, to which it provides the academic teaching and research arm. It is also collocated with the Coroners Court of Victoria and the National Coronial Information System. These agencies combined are both custodians of rich sources of data and users of data for research. While such death data are sensitive, they are available for research after appropriate scientific and ethical / privacy scrutiny. Linkage with data from other sectors (to provide de-identified data sets) would add enormously to the value of the public good research that could be conducted e.g. unexpected deaths data linked with mental health, aged care, or corrections data.

The School was instrumental in establishing the National Coronial Information System and managing it for more than a decade. This nationally significant statistical asset would be invaluable to Australia's public good research if linked to other important data sets to form an enduring data platform.

The mission of the **Faculty of Information Technology** at Monash is to lead global IT research and education. Our goal is to innovate tomorrow's information technology for the benefit of society and to educate our students to become global technology leaders. Society today is in the midst of an information revolution. The social and economic changes we are witnessing are as basic and wide-ranging as the Industrial Revolution. The scope of these changes is reflected by the vast range of activities that now constitute IT research. Much of the richness of modern IT research arises from its intersection with other disciplines. The Faculty's flagship programmes reflect this intersection. One of the flagships focuses on Data systems and cybersecurity (see [www.infotech.monash.edu.au/research/about/flagships/data-systems-cybersecurity/](http://www.infotech.monash.edu.au/research/about/flagships/data-systems-cybersecurity/) )

## References

1. Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ*. 2013;346:f2360. PubMed PMID: 23694687. Pubmed Central PMCID: PMC3660619. Epub 2013/05/23. eng.
2. Mitchell RJ, Cameron CM, McClure RJ, Williamson AM. Data linkage capabilities in Australia: practical issues identified by a Population Health Research Network 'Proof of Concept project'. *Aust N Z J Public Health*. 2015;39(4):319-25.
3. Loff B, Campbell EA, Glass DC, Kelsall HL, Slegers C, Zion DR, et al. Access to the Commonwealth electoral roll for medical research. *Med J Aust*. 2013 Jul 22;199(2):128-30. PubMed PMID: 23879514. Epub 2013/07/25. eng.
4. Chung S-C, Gedeberg R, Nicholas O, James S, Jeppsson A, Wolfe C, et al. Acute myocardial infarction: a comparison of short-term survival in national outcome registries in Sweden and the UK. *The Lancet*.383(9925):1305-12.
5. Timmis A, Rapsomaniki E, Chung SC, Pujades-Rodriguez M, Moayyeri A, Stogiannis D, et al. Prolonged dual antiplatelet therapy in stable coronary disease: comparative observational study of benefits and harms in unselected versus trial populations. *BMJ*. 2016 2016-06-22 12:36:24;353.