

Submission to The Productivity Commission Inquiry

Submitted by Sradha Kotwal, Sarah Coggan, Kellie Nallaiah and Martin Gallagher on behalf of The George Institute Australia

29 July 2016

Questions on high value public sector data

What public sector datasets should be considered high-value data to the: business sector; research sector; academics; or the broader community?

From the perspective of health research on the major burdens of disease in Australia, high value datasets should include:

1. Public sector datasets that offer longitudinal data on health outcomes, such as State-based admitted patient data collections and death registries.
2. Data from clinical registries, such as the Australia and New Zealand Dialysis and Transplant Registry (ANZDATA), offer invaluable and important information beyond capabilities of public sector data sets.
3. Australian Government datasets relating to outpatient services (Medical Benefits Schedule) and pharmaceutical usage (Pharmaceutical Benefits Schedule). These are of high value in tracking health behaviours, determinants of disease, and complications of treatments and patient safety. This is, especially important in regard to novel therapies that are incompletely understood at their time of introduction (of which there have been a number of examples in the last 5-10 years). Currently there are number of barriers making it difficult for researchers to access Australian Government datasets.

What characteristics define high-value datasets?

The characteristics that define high value datasets are:

- Broad population data, providing a highly representative sample of the Australian population and its vulnerable sub-groups.
- Routine and mandatory collections which include data that are otherwise expensive and logistically difficult to collect in a systematic manner
- Data collections that include otherwise hard to collect data in a systematic manner
- Data on health service utilisation at a population level, for example collected by hospitalisation datasets and MBS/PBS that can help answer questions about health services access and delivery.
- Data that can be adjusted (for research purposes) for social determinants of health, especially social, education, economic and residential information to provide an estimate of the impact of such variables on health
- Datasets that include routinely collected data from a significant period into the past which can be used to identify trends in particular populations. The WA Admitted Patient Data Collection and Mental Health datasets currently have data available for linkage from 1970 to 2014.
- Datasets that contain data that are difficult to collect accurately from individuals either because of the length of time that has passed or the level of detail that is required for the research being conducted.

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

The advantage of the public sector datasets for medical and health research is their broad-based population representation and insights, offering real potential for whole-of-system understanding and evidence-based reforms to the health system and clinical practice. There are numerous examples of how such data can be used to better understand the outcomes of health services and the safety of pharmaceutical and other treatments. With the increasing cost and complexity of medical care, the ageing population and high burden of multi-morbidity, such data will be essential to our understanding and affordability of health services into the future.

Large scale clinical trials are an expensive undertaking and using administrative datasets is a potential cost saving measure without compromising research output. This approach has been used in other countries to perform clinical trials that would otherwise have been impossible. An example of this is the TASTE trial in Sweden which used a clinical registry for randomisation and clinical outcome follow up. It is likely that the use of data tools in research will expand, thereby reducing the cost of studying medical therapies into the future. Reduced costs of research can equate to more research and more answers, benefitting the community both directly (if they suffer from the conditions or use the therapies studied) and indirectly (through the more efficient provision of medical therapies and services, effectively freeing up resources that can be used elsewhere).

The ready availability of data for medical research purposes that captures other social determinants of health is valuable as there is clustering of burden of disease with economic and social disadvantage. Understanding and defining pockets of higher disease burden may represent an efficient and effective means of tackling diseases such as diabetes and hypertension that have a high population burden. Such an approach may also offer the ability to tailor therapies and interventions more effectively thereby targeting the communities that have the most need.

Importantly, many of these patient groups, especially the elderly and those with chronic diseases such as diabetes, are not typically enrolled in clinical trials; and therefore our understanding of novel treatments in these population groups is limited. Public sector data collections offer a way of understanding the real world effects of novel and existing therapies and allows an insight into patient groups which would otherwise be neglected.

Questions on collection and release of public sector data

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

There are several complexities and barriers to data access at Federal and State levels:

There has been a general reluctance to release linked health data for research due to concerns that research is not the primary purpose of the data collection and the impact upon patient rights. Standardised rules and structures for the handling of such datasets by all parties would be a useful way ensuring data security and reducing the risks of misuse.

The requirement for individual patient consent to access MBS and PBS data is another barrier. The rationale for the processes currently in place should be made more transparent so solutions for gaining access to these invaluable data can be explored. Processes akin to the current Federal Government MBS Review might be a good opportunity to explore ways to do this.

State governments have differing understanding, rules and regulations around the access and use of their datasets. A national standardised approach to data security would provide robust guidelines for appropriate data access and research use and would benefit state governments. A national approach also can help deliver considerable efficiencies for researchers, government and others accessing the data.

An additional barrier to accessing government data is the widely varying conditions applied by research ethics committees across the country and the lack of recognition of approvals from other states/jurisdictions. Many of these entities seem to have limited experience of data linkage and this may also be a factor in their reticence. Although the National Mutual Acceptance (NMA) of single ethical review of multi-centre clinical trials was extended to include all human research (including data linkage studies) in December 2015, there are currently only four participating jurisdictions (Queensland, South Australia, Victoria and NSW). The participation of the remaining jurisdictions would undoubtedly help to streamline the process of obtaining linked data.

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

Governments could be more open and transparent about the type of information it collects and what can be made available to researchers within current legislation. Knowledge of what is available can increase its use (internally and externally) to improve patient outcomes.

Governments could actively promote and support the use of data collected with appropriate oversight to ensure adherence to data management and use principles. In some circumstances this requires a shift of focus with data custodians facilitating rather than restricting data use. Governments should be able to release data that complies with privacy principles and legislation to appropriate persons with adequate security clearances who are working for organisations that aim to improve health outcomes for patients.

Efficient and effective use of data could be achieved by reviewing the purpose for which data is collected. For data that is collected to improve the delivery of care to patients (should arguably be the majority of health care data), systems and processes should be put in place to ensure the maximal use of this data to achieve its' intended aim. This aim extends beyond treatment of the individual to treatment of populations of people with similar conditions for widespread impact on the health of communities.

Governments across the country have highly variable rules and understanding of the potential use of their datasets. Streamlining and standardising these processes offers the promise of considerable efficiencies and enhancements in data security.

The development of collaborations between government bodies and independent academic institutions for research purposes also have potential to offer real benefit to research translation. Researchers who understand the real value of data should work closely with policy makers to enable the incorporation of findings into their deliberations and where appropriate, the rapid translation of research into policy.

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?

Standardising the collection, sharing and release of public sector data nationally would be a major efficiency gain for the research and health sector, allowing for an increased amount of resources

dedicated to improving health outcomes for patients. Standardisation would enable linkage across multiple datasets and jurisdictions much more readily, as it is currently an expensive, protracted and debilitating process. In the healthcare setting, many of the important definitions for datasets are already standardised through ICD 10 and other state and national data coding standards. The costs of further standardisation across different jurisdictions would be overshadowed by the gains in healthcare efficiencies and effectiveness.

The most significant standardisation would be around the means of accessing data and the rules around storage and analysis of data. This could represent a significant cost saving for State governments, as they have varied processes for such work that they must develop themselves and maintain. In addition, it would provide a further layer of reassurance for data custodians that any data that is released is handled and analysed appropriately. The standardisation of the cost charged by the data linkage units for identical data would also be useful in planning and budgeting for studies.

Standardisation of the collection of public sector data would allow access to identical datasets across jurisdictions and therefore a more complete answer to research questions. Currently, there are variations in the detail of the data collected and made available for example some jurisdictions collect full date of birth while some only collect month and year of birth.

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?

As applicants for and users of the data, we are not aware of the criteria and tools that governments currently use in this decision-making. However, we have seen significant variation between states in their readiness, rules and application processes around data sharing such that the decision-making tools do not appear standard across the nation.

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?

- At the NSW state level, developing the capabilities of the Centre for Health Record Linkage has led to big advances in the use and understanding of health data in NSW.
- The establishment of the population health research network to develop nationwide infrastructure has also helped although advances have been very slow.
- The inclusion of data linkage studies in the National Mutual Acceptance (NMA) scheme.

Questions on data linkage

Which datasets, if linked or coordinated across public sector agencies, would be of high value to the community, and how would they be used?

Health datasets such as admitted patient data collections, Medicare Benefits Schedule, Pharmaceutical Benefits Schedule, emergency department collections, and education datasets would be valuable in understanding inpatient and outpatient use of healthcare.

The Pharmaceutical Benefits Schedule could also be used to conduct post marketing surveillance of new medication, including their clinical outcomes. An example of this would be evaluating the long-term impact of novel biological agents for immunological diseases and oncology that is difficult to measure because of the rare nature of many of these conditions.

We note that the Census 2016 is collecting identifiable data and this could represent an important tool for data linkage going forward. We understand the concerns about linking such data but the potential benefits would likely outweigh the potential risks identified so far.

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

When performing multi-jurisdictional data linkage, the process of needing separate applications for each data linkage unit and, separate ethics applications is burdensome, time-consuming and creates barriers to linking datasets and research inefficiencies, for example, expanded time-frames and increasing costs. This erodes one of the putative benefits of linked data - cost-effectiveness.

Furthermore, access to the MBS and PBS datasets require MBS/PBS specific consent forms in addition to study consent forms. Good arguments have been made for not seeking consent when using mandatory data collections as consent costs money, limits the extent of the patient population included and perhaps reduces patient confidentiality and privacy (as patients have to be identified to be consented).

These processes also add to the cost of such work (which in itself is quite costly) and overall, discourage researchers and others to undertake more complex multijurisdictional linkage projects.

How can Australia's government agencies improve their sharing and linking of public sector data? What lessons or examples from overseas should be considered?

As detailed above, improvements in sharing and linking public sector data can be done by streamlining national processes and reducing bureaucracy in accessing data, which will reduce costs and expedite the valuable results of such research.

It is also important that consultation from researchers and navigators of the approval process is part of re-designing these systems.

Countries such as Canada, Sweden and New Zealand are great examples of countries with efficient and cost effective processes around data linkage.

Questions on high value private sector data

What private sector datasets should be considered high-value data to: public policy; researchers and academics; other private sector entities; or the broader community? In each case cited, what characteristics define such datasets?

What would be the public policy rationale for any associated government intervention?

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

Private sector datasets such as private pathology laboratory data would be considered high-value. Population level analyses of private pathology data has only been conducted on a very small scale in Australia and is currently very difficult to do due to the lack of standardised processes. At The George Institute we have experienced this limitation first hand specifically because of external concerns of about data safety and confidentiality.

Providing easy access to this data would not only improve patient care in hospitals but likely reduce health costs due to the reduced likelihood of repeating tests. Real time linkage of private pathology

datasets to hospital laboratory data would also provide continuity in patient care and improved safety.

Access to health insurance data would be valuable as has been demonstrated by the results from the Kaiser Permanente databases in the United States.

Questions on access to private sector data

Are there any legislative or other impediments that may be unnecessarily restricting the availability and use of private sector data? Should these impediments be reduced or removed?

No comments in this section.

What are the reasonable concerns that businesses have about increasing the availability of their data?

From our observations and experience most concerns around increasing availability is the increase in risk to patient confidentiality. Despite reassurance that this risk is low, not all parties the Institute deals with have agreed to take it on. Perhaps there is a role for government, as part of a standardised national approach, to reduce this risk (or fear of) if all the relevant and appropriate processes (including adherence to legislation and oversight) for such linkages are performed.

What principles, protocols or legislative requirements could manage the concerns of private sector data owners about increasing the availability of their data?

Should the collection, sharing and release of private sector data be standardised in some way? How could this be done and what would be the benefits and costs? What would standards that are 'fit for purpose' look like?

To what extent can voluntary data sharing arrangements — between businesses / between businesses and consumers / involving third party intermediaries — improve outcomes for the availability and use of private data? How could participation levels be increased?

Would such voluntary arrangements raise competition issues? How might this change if private sector information sharing were mandated? Is authorisation (under the Competition and Consumer Act 2010 (Cth)) relevant?

What role can governments usefully play in promoting the wider availability of private datasets that have the potential to deliver substantial spillover benefits?

How can the sharing and linking of private sector data be improved in Australia? What lessons or examples from overseas should be considered?

Who should have the ownership rights to data that is generated by individuals but collected by businesses? For which data does unclear ownership inhibit its availability and use?

Questions on consumer access to, and control over, data

What impediments currently restrict consumers' access to and use of public and private sector data about themselves? Is there scope to streamline individuals' access to such data and, if there is, how should this be achieved?

Are regulatory solutions of value in giving consumers more access to and control over their own data?

Are there other ways to encourage greater cultural acceptance amongst businesses of consumer access to data about them?

What role do third party intermediaries currently play in assisting consumers to access and use data about themselves? What barriers impede the availability (and take-up) of services offered by third party intermediaries?

What datasets, including datasets of aggregated data on consumer outcomes at the product or provider level, would provide high value to consumers in helping them make informed decisions? What criteria should be used to identify such datasets? What, if any, barriers are impeding consumers' access to, and use of, such data?

We believe that consumers should have easy access to their data whether it is directly or indirectly collected. Design of processes for access to data should have consumer input.

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?

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?

The infrastructure costs of setting up the processes for access to datasets should be funded by the government while users of the data (Government departments, research institutions, health institutions) should pay for access to the data. The cost benefits from the data sets would be an indirect 'reinvestment' back into the health system through better research and thus health outcomes.

There needs to be increased investment in developing a skilled labour force, whether it be internal or external to government, but needs to be supported and facilitated by government. There is a wealth of untapped health data out there that could be used to improve health services and health outcomes for patients. Having a skilled workforce, consisting of people who are appropriately trained, provided regular professional development opportunities and valued for what they contribute, is key to achieving this.

Questions on privacy protection

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

Any data that is considered to be sensitive or any data that makes an individual potentially identifiable.

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

Wherever possible, analyses should be performed using re-identifiable or unidentifiable data only. If the data is re-identifiable, the strongest measures should be taken to ensure that identification does not occur without the appropriate approvals.

Also, access to any data should be restricted solely to those listed on the application.

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?

The weight should be on protecting the privacy of individuals first and foremost. So long as all the correct measures are put into place to ensure this occurs then there is very little, if any, opportunity for misuse. Once data is aggregated with strict adherence to the privacy principles (that should be reviewed regularly along with the changes in the capabilities of technology) there should be no reason why data should not be made available and used when appropriate.

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?

Standardisation on processes and agreement between jurisdictions on this standardisation.

How effective are existing approaches to confidentialisation and data security in facilitating data sharing while protecting privacy?

What lessons from overseas jurisdictions can Australia learn from regarding the use of individuals' and businesses' data, particularly in regard to protecting privacy and commercially sensitive or commercially valuable information?

What are the benefits and costs of allowing an individual to request deletion of personal information about themselves? In what circumstances and for what types of information should this apply?

What competing interests (such as the public interest) or practical requirements would indicate that the ability to request deletion should not apply?

We have no comments on this section.

Questions on other restrictions

Having regard to current legislation and practice, are further protocols or other measures required to facilitate the disclosure and use of data about individuals while protecting privacy interests? What form should any such protocols or other measures take?

Is there need for a more uniform treatment of commercial-in-confidence data held by the Australian Government and state and territory governments?

Are there merits in codifying the treatment and classification of business data for privacy or security purposes? What would this mean in practice?

No comments on this section

Questions on data security

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

How do data security measures interact with the Privacy Act?

The Information Privacy Principles (IPPs) govern agencies of the Commonwealth in their collection, management, and use of data containing personal information as set out in the Privacy Act 1988. Data security measures for data linkage studies comply with IPPs 1-9 and are in breach of IPP 10 (limits on use of personal information) and/or 11 (limits on disclosure of personal information). A justification for these breaches are the high level of security measures that include the separation of identifiable data (required for accurate linkage) from the linked data before the researchers have access to it for analysis, the requirement to use SURE (Secure Unified Research Environment) for high risk projects involving Commonwealth data.

How should the risks and consequences of public sector and private sector data breaches be assessed and managed? Is data breach notification an appropriate and sufficient response?

Provision of training and guidelines on how to deal with data breaches would help in the management and mitigation of data breaches.