

October 2017

Grattan Institute submission to the community consultation on 'Developing a framework for the secondary use of My Health Record data'

Stephen Duckett

Grattan Institute Support

Founding Members



Australian Government



THE UNIVERSITY OF
MELBOURNE



Affiliate Partners

Google
Medibank Private
Susan McKinnon Foundation

Senior Affiliates

EY
Maddocks
PwC
McKinsey & Company
The Scanlon Foundation
Wesfarmers

Affiliates

Ashurst
Corrs
Deloitte
GE ANZ
Jemena
The Myer Foundation
Urbis
Westpac

This submission was written by Stephen Duckett, Director, Health Program, Grattan Institute. Naveen Tenneti provided research assistance and contributed to the submission.

The opinions in this submission are those of the authors and do not necessarily represent the views of Grattan Institute's founding members, affiliates, individual board members, reference group members or reviewers. Any remaining errors or omissions are the responsibility of the authors.

Grattan Institute is an independent think tank focused on Australian public policy. Our work is independent, practical and rigorous. We aim to improve policy outcomes by engaging with both decision-makers and the community.

For further information on the Institute's programs, or to join our mailing list, please go to: <http://www.grattan.edu.au/>

This submission may be cited as: Duckett, S., 2017, *Grattan Institute submission to the community consultation on 'Developing a framework for the secondary use of My Health Record data'*, Grattan Institute

All material published or otherwise created by Grattan Institute is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License

Overview

- Data sets held by governments and businesses can be used for multiple purposes, not just the initial transactional record. They have the potential to transform policy development and evaluation and economic analysis
- The My Health Record, if widely adopted, will provide a significant opportunity to collect accurate, useful and relevant information for all stakeholders
- Secondary analysis of data sets collected for 'routine' or 'administrative' purposes is now a well-accepted type of research
- Secondary analysis of government data sets has a number of additional benefits:
 - Because the analysis is conducted on government data, the research is almost always policy-relevant. It can provide insight into patterns of spending or service use that would otherwise not come to the attention of policy makers; and
 - Secondary analysis leverages investments that have already been made in data collection and so is generally a less expensive form of research. Secondary analysis of Australian data sets has been used to improve the quality and efficiency of, and access to, health care
- Government data holdings should be seen as an important public resource to assist in policy-relevant research which will benefit the Australian community. Failure to harness fully the potential of these data sets represents a significant lost opportunity both for policy development and research
- Privacy risks can be mitigated by controlling data outputs and/or controlling the data released. The New Zealand approach to controlling data outputs should be explored as a potential additional data release strategy in Australia. Data custodians also limit release to approved data users. This is appropriate and should be continued
- Data release should be facilitated by
 - Releasing metadata
 - Developing streamlined and standardised release-approval processes; and
 - Developing common-use data sets.

Table of contents

Overview.....	3
1 Data for decision making	5
2 Controlling the risk to privacy.....	8
3 From rhetoric to action on data release	10
4 References	12

1 Data for decision making

Businesses, not-for-profit organisations, and governments are increasingly recognising the power of data to improve their decision making and service to their customers or clients¹. Data sets held by governments and businesses can now be used for multiple purposes, not just the initial transactional record. They have the potential to transform policy development and evaluation and economic analysis.²

The potential uses of data are now limited only by our imaginations:

Publishing, linking and sharing data can create opportunities that neither government nor business can currently envisage.³

Daily additions to world-wide data holdings are now measured in the quintillions.⁴ Increased computer power makes linking of data sets (e.g. social security and tax records) more feasible and increases the utility of data holdings.

In the past, government data bowerbirds have collected, but obstructed the release of, data. Now, the Australian Government promises

to optimise the use and reuse of public data; to release non-sensitive data as open by default; and to collaborate

with the private and research sectors to extend the value of public data for the benefit of the Australian public.⁵

1.1 My Health Record data

The My Health Record's primary purpose is to provide a

central, lifelong depository for each patient's medical records, regardless of provider, in an electronic format accessible by providers and patients.⁶

But an important secondary purpose is to use the data to understand patterns and trends. Secondary analysis of data sets collected for 'routine' or 'administrative' purposes is now a well-accepted type of research.⁷

Despite this treasure trove of data, researchers typically rely on survey data to gain insight. Yet surveys are a poor source of information. The reasons for this are well-documented but worth repeating. Non-response rates are high and increasing. Longitudinal surveys are rare – and suffer from high attrition. Although the largest surveys conducted in Australia are sufficient for calculating basic statistics with confidence, they are still too small to be useful on other important research questions. The closure of a major factory would likely not even be detectable in any survey currently available to Australian researchers.

¹ McAfee and Brynjolfsson (2012)

² Einav and Levin (2013)

³ Turnbull (2015)

⁴ Kim, *et al.* (2014); a quintillion has 18 zeroes

⁵ Turnbull (2015)

⁶ Productivity Commission (2017)

⁷ This and other sections of this Submission drawn on a previous paper on data access. See Duckett, *et al.* (2017) and Academy of the Social Sciences in Australia (2013)

In an era where some Australians are feeling increasingly alienated and isolated from policy decision-making, the ability to conduct detailed public policy research is needed urgently.

Although survey data is typically less sensitive than administrative data, the barriers to access are nonetheless onerous. Linking and matching is often forbidden. And confidentiality requirements often make good-practice research methods (such as obtaining review from external colleagues) practically impossible.

Secondary analysis of government data sets has additional benefits:

- Because the analysis is conducted on government data, the research is almost always policy-relevant. It can provide insight into patterns of spending or service use that would otherwise not come to the attention of policy makers; and
- Secondary analysis leverages investments that have already been made in data collection and so are generally a less expensive form of research. Secondary analysis of Australian data sets has been used to improve the quality and efficiency of, and access to, health care.

Secondary analysis is increasingly common. Many data agencies have established procedures to ensure confidentiality of data, so that individual privacy is not infringed by allowing further (secondary) use of the routinely collected data. Data released is not able to be identified to any individual (to protect privacy), and small cells are also suppressed to prevent any possible identification. Data release agreements generally require

researchers to certify that they will not attempt to identify individuals and/or specific services.

Government data holdings should be seen as an important public resource to assist in policy-relevant research which will benefit the Australian community. Failure to harness fully the potential of these data sets represents a significant lost opportunity both for policy development and research.

The Grattan Institute has been fortunate to use such resources to enhance its research. A Grattan report released in August 2015 used routinely collected data to analyse the use of ineffective treatments in Australian hospitals.⁸ International research shows that ineffective and inappropriate care is a major source of waste in hospitals. Further, it poses a risk to patients who could otherwise have avoided the stress, cost, inconvenience and risk of a hospital stay.

Australia currently has no system for identifying and addressing provision of inappropriate care in hospitals. Duckett, *et al.* (2015) developed a model for doing so; however, because our data sets were not linked, and what is considered 'inappropriate' depends on the patients' history of treatment, we could only measure use of about one eighth of the treatments listed as ineffective in the United Kingdom. If we were able to access this data, we would be able to observe four times as much ineffective care.

1.2 Making data useful for policy makers

Data should be collected not just for storage, but to be used by policy makers. The patterns, trends and insights generated should

⁸ Duckett, *et al.* (2015)

be used to improve service delivery. The Productivity Commission's 'Shifting the Dial' report recognises the possibilities that the My Health Record presents. It notes:

There are compelling grounds to use My Health Record as a platform for providing clinically-proven advice to patients, rather than just as a method for collecting data for clinicians.⁹

Understanding how a platform can be augmented requires the access and use and secondary use of routine data.

⁹ Productivity Commission (2017)

2 Controlling the risk to privacy

Big data comes with risks, including risks to privacy.¹⁰ Privacy risks can be to individual consumers and to providers, but risks to the two different types of stakeholders are quite different.

Custodians should not necessarily regard all data as private. For instance, currently institutions with the best survey data impose strict confidentiality requirements on researchers. These largely self-imposed requirements are onerous.

Surveys by institutions such as the ABS are clearly intended to be used for research. Survey respondents volunteering information do so with full knowledge of this purpose. It is therefore unlikely that the assurance of confidentiality has any effect on the information offered. Yet once this assurance is given, it must be honoured.

Before placing restrictions on data release, or issuing assurances about confidentiality, custodians should think critically about whether release would in fact breach someone's privacy or substantially damage the integrity of the data. Restrictions should be withdrawn or not imposed unless there is clear evidence of their benefit.

In the past, data have only been released in broad, aggregated, tabular form, either in standard published form or through user-specified tabulations. A more recent advance has been the provision of front-ends to data sets, to allow users to generate tables on-line from a data extract. Typically the data access tools

are structured so as not to allow personally identifiable data to be generated (e.g. by using age-ranges in drop-down menus).

Statistics agencies and other data custodians have released unit record data sets for about 30 years, and data custodians have adopted two broad approaches to mitigating privacy risks.

The first is to *control the outputs*. Data custodians can allow data analysis directly on data sets in controlled premises, and vet all outputs (such as computer-generated cross-tabulations taken off the premises) to ensure none allow personal identification.

The second is to *control the data released*. This approach includes data modification strategies (e.g. introducing perturbations in small cells¹¹), as well as limiting the variables released in one data set (e.g. not releasing detailed age and detailed geography as it may enable identification).

In both cases, data custodians may *control the users to whom data are released*. This may include police checks on data users, or only releasing data to reputable organisations or organisations with the ability to guarantee security of access, and so on.

2.1 Whose privacy?

Controls over data release are primarily designed to protect privacy. The public's focus thus far has been centred on ensuring the privacy of personal medical records. The providers of information, however, also have a stake. Data custodians often treat privacy of the provider on the same footing as privacy of the

¹⁰ Richards and King (2013)

¹¹ Mehmood, *et al.* (2016); Polonetsky, *et al.* (2016)

consumer. They should not, especially where providers get public funding.

There are very different issues involved in release of information about providers, where the risks are generally associated with commercial issues, compared to the privacy risks about individuals, which are about sensitive personal information. These should not be treated as equivalent.

Medicare data which may identify medical practitioners in terms of their volume and type of activity should also be able to be released into the public domain. But again, data users should not be allowed to publish anything which identifies individual medical practitioners.¹²

¹² Governments should publish identifiable information about bulk-billing rates and average out-of-pocket costs for general practices and specialists

3 From rhetoric to action on data release

To move from rhetoric to action on data release, Australia should:

3.1 Release metadata

Routine or 'administrative' data holdings are complex, with vast numbers of data elements collected. Some of these will be obvious and known to researchers, e.g. Medicare Benefits Schedule (MBS) item numbers, but others may be system created, e.g. information could be extracted about medical services based on the year the service was delivered or the year a claim was processed.

Information about the variables held and their definitions ('metadata') could be standardised and made publicly available. The necessary resources must be provided to maintain and update the metadata.

Publication of information on data elements should be accompanied by standardisation of de-identification methods and publication of information about the rules used to protect privacy (e.g. suppression or randomisation of low frequency events).

Both publication of metadata and of information about de-identification methods will help to make researchers aware of the data available. This will stimulate use of the data.

Better information about data characteristics will also improve the quality of requests for data extraction, because researchers would be better placed to specify requests precisely in terms of the data elements held.

3.2 Develop standard approval mechanisms

Current administrative arrangements provide that claims-processing for major government functions (e.g. claims against the MBS) must be done by the Department of Human Services, while policy responsibility for these functions is vested in the relevant policy line agency (in the case of the MBS, the Department of Health and Ageing). Similar arrangements apply for social services data.

This separation means that although claims data is held by the Department of Human Services, decisions about data release are made by the Department of Health.

Data release decisions require co-ordination between the two Departments. At present this process tends to be slow and cumbersome. This problem is exacerbated when data from more than one Department is requested. A more streamlined management process should be implemented for cross-agency data release decisions.

A list of precedents of previous data approvals – and non-approvals – should be published, to further streamline data requests and approvals.

3.3 Develop common use datasets

Common or public use data sets are standardised extracts of data or reports from surveys which are made available for external use. They can exist at various levels: detailed published tables, data cubes (interactive tables which allow for personalised extraction of aggregate data), confidentialised unit record files based on a

sample of records, and researcher access to unit record data under secure conditions.

Development and release of common use data sets will help to streamline access and provide an important resource for research and policy analysis. In addition to facilitating research which would answer defined questions, a common use data set would also facilitate exploratory research, allowing researchers to undertake preliminary analyses to test whether specific research questions may be worth pursuing and whether special data extraction requests are warranted.

Conditions for release of common data sets should be modelled on existing ABS procedures, including:

- Approval of the body to which data is released;
- Conditions on secure storage;
- Conditions on not identifying individuals in any publications;
- Prohibitions on on-passing data to third parties; and
- The payment of a nominal charge. Each application for access to the common use data set should be required to provide a short outline of proposed use of the data, to allow a determination that the proposed use meets a 'public interest' test as required by existing legislation.

3.4 Learn from the United States

Many overseas jurisdictions release common use data sets. The United States Department of Health and Human Services has a very open policy about data availability. Research using important data sets, such as the Medicare claims data and hospital utilisation data, has provided the basis for thousands of policy evaluations and research papers.

The U.S. Department has taken great strides in recent years to facilitate data access. It has established a website (healthdata.gov), which is

dedicated to making high-value health data more accessible to entrepreneurs, researchers, and policy makers in the hopes of better health outcomes for all.

This Department says it has established the website to 'liberate' data.

Appropriate release of My Health Record data would help create a similar data access revolution in Australia.

4 References

- Academy of the Social Sciences in Australia (2013) *Facilitating access to routine data for research benefiting the Australian people*, The Academy
- Duckett, S., Breadon, P., Romanes, D., Fennessy, P. and Nolan, J. (2015) *Questionable care: avoiding ineffective treatment*, Grattan Institute
- Duckett, S., Goss, P. and Terrill, M. (2017) *Better access to data will lead to better decisions: Submission to the Productivity Commission's Inquiry into Data Availability and Use*, Grattan Institute
- Einav, L. and Levin, J. D. (2013) 'The Data Revolution and Economic Analysis', *National Bureau of Economic Research Working Paper Series*, No. 19035,
- Kim, G.-H., Trimi, S. and Chung, J.-H. (2014) 'Big-data applications in the government sector', *Communications of the ACM*, 57(3), p 78-85
- McAfee, A. and Brynjolfsson, E. (2012) 'Big data: the management revolution', *Harvard business review*, 90(10), p 60-66
- Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G. and Guo, S. (2016) 'Protection of Big Data Privacy', *IEEE Access*, 4, p 1821-1834
- Polonetsky, J., Tene, O. and Finch, K. (2016) 'Shades of gray: Seeing the full spectrum of practical data de-identification', *Santa Clara Law Review*, 56(3), p 593-629
- Productivity Commission (2017) *Shifting the Dial: 5 Year Productivity Review*, Productivity Commission
- Richards, N. M. and King, J. H. (2013) 'Three Paradoxes of Big Data', *Stanford Law Review Online*, 66, p 41-46
- Turnbull, M. (2015) *Australian Government Public Data Policy Statement (released 7 December 2015)*, from https://www.dpmc.gov.au/sites/default/files/publications/au_st_govt_public_data_policy_statement_1.pdf