Health Data Governance

Re-Identification of Health Records

pwc

www.pwc.com.au

1. Introduction

Australian governments have long recognised the potential of big data, which can provide the depth of insight about the community to enable genuine evidence-based policy development, public infrastructure planning, and service delivery innovation.

The intensive analysis of data sets drawn from a range of different sources to uncover important trends and insights, commonly referred to as 'Big Data', has a particularly important role to play in Australia's health system, which comprises a complex web of medical, social and behavioral influences. Rich data sets exist at patient and system levels, reflecting the dominant role that Australian Governments play in our system, anchored in the Medicare Benefits Scheme (**MBS**), Pharmaceutical Benefits Scheme (**PBS**) and now-ubiquitous My Health Record (**MHR**).

Big data can improve population health, stimulate therapeutic innovation and enhance system efficiency and sustainability. The benefits can only be realised with the trust and confidence of citizens, bringing privacy issues into sharp relief. The use and management of data and health information is now regulated by the *Privacy Act 1988* (Cth) (**Privacy Act**) and health records legislation in most States and Territories.

In this article, we consider some of the regulatory issues relating to the de-identification and re-identification of health records data, particularly in light of a September 2016 incident in which de-identified PBS/MBS data was re-identified. We also highlight the challenge facing legislators in managing the tension between protecting personal information and allowing innovation to enable the promise of big data to be realised.

Increasingly sophisticated data management techniques have been developed to de-identify health records information, to enable data to be used, whilst complying with regulatory obligations. Those same techniques may, however, enable data sets to be re-identified, highlighting the need for a range of responses to adequately protect personal information.

There is an inherent tension between using big data sets to benefit the community and the personal privacy of the individuals to whom the data relates. Appropriate policy, contractual and technical controls are needed to maximise privacy protections without limiting the usability of the information (by, for example, over-use algorithmic de-identification which can reduce the integrity of the information). Finding the right mix of controls will maintain public confidence and spur further innovation in the health sector, realising the potential of big data for better health outcomes and system sustainability.

2. Government 2.0 and the creation of data.gov.au

In August 2016, the Commonwealth Department of Prime Minister and Cabinet (**PM&C**) launched the much anticipated data.gov.au website with the intention of providing an easy way to find, access and re-use public data sets. Data.gov.au publishes a range of datasets from various government entities, including data sets relating to health, science, civic infrastructure, community services, finance management, and communications.

It is intended that data.gov.au will become a comprehensive repository for data from Commonwealth, State and Territory governments, and City Councils, although at this point it is predominantly data sets from several Commonwealth agencies. PM&C has recognised that expanding the data stack will be an ongoing process.

The creation of data.gov.au followed on from the Government's 'Declaration of Open Government' (**Declaration**) and the declaration was a response to the 'Government 2.0 Taskforce Report'.[1] Through the Declaration, the Government committed to promoting greater participation in Australia's democracy by acting as an open government "built on better access to and use of government held information, and sustained by innovative use of technology". Data.gov.au facilitates the sharing of de-identified government information on a public platform.

De-identified PBS and MBS data sets were initially published, although following the re-identification incident, discussed below, that has now been removed. At this time the public disclosure of data relating to health and health care is relatively limited; there are 20 data sets available and a number of these are lists of sporting clubs and wellness providers, rather than therapeutic data sets.[2]

[1] Department of Finance (Cth), *Declaration of Open Government*, (http://www.finance.gov.au/archive/policy-guides-procurement/declaration-of-open-government/)

[2] Australian Government, data.gov.au, *Datasets* (https://data.gov.au/dataset?groups=health)

3. De-identification of data

A common method of safeguarding collected personal information (particularly in big data sets) is de-identification. Personal information is 'de-identified' where it is no longer 'about an identifiable individual or an individual who is reasonably identifiable' under section 6(1) of the Privacy Act.

The ability to successfully de-identify information is an important enabler of big data, as it allows for large quantities of personal information and, in the case of health, highly sensitive personal information, to feasibly be sourced and prepared for analysis.

De-identification can be thought of as removing or modifying a person's name, address or date of birth, although more sophisticated techniques exist; the Office of the Australian Information Commissioner (**OAIC**) has highlighted a number of these, including:

- removing or modifying quasi-identifiers (for example, significant dates, profession, income) that are unique to an individual, or in combination with other information are reasonably likely to identify an individual;
- combining information or data that is likely to enable identification of an individual into categories. For example, age may be combined and expressed in ranges (e.g. 25 – 35 years) rather than in single years (e.g. 27, 28 years of age);
- altering identifiable information in a small way such that the aggregate information or data is not significantly affected — a tolerable error — but the original values cannot be known with certainty;
- swapping identifying information for one person with the information for another person with similar characteristics to hide the uniqueness of some information;
- using algorithms to generate 'synthetic' data from original data sources and substituted for it, while preserving some of the patterns contained in the original data. This allows systems to be tested with data that is realistic but poses less risk of re-identification; and
- suppressing data, which involves not releasing particular information that may enable re-identification, or deleting that information from the dataset. Data suppression may impair the utility of an information asset.[3]

The challenge that the OAIC has raised in this respect is understanding how organisations ensure de-identification is "correctly done" (if, indeed, there is a "correct" way to do so).The appropriate de-identification method would depend upon the sensitivity of the data, relevant organisational controls and the analytical requirements, highlighting the need for an effective risk assessment, often called a privacy impact assessment, to be undertaken. The objective in undertaking the assessment should be to ensure that personal data is appropriately protected whilst retaining sufficient detail in the relevant data set to enable it be used for its intended purpose.

4. Overview of regulatory regime

The Privacy Act and the *Australian Privacy Principles* contained in Schedule 1 of the Privacy Act (**APPs**), set out a regime for the collection, holding, use and disclosure of data applicable to Commonwealth agencies and organisations. The regime is based on guiding principles rather than prescriptive requirements. Importantly:

- when an agency or organisation collects personal data from an individual, it must inform the individual of the purpose for which it is collected so that the individual can provide consent to its use and disclosure for that purpose. If the organisation wishes to obtain consent for a secondary purpose at the time of collection, it must state a specific secondary purpose – a broad, 'catch-all' purpose will be insufficient;
- the agency or organisation must not use or disclose the personal data other than for the consented purposes unless an exception under the Privacy Act applies. One exception is the existence of a 'permitted health situation', which includes where the use or disclosure is necessary for research relevant to public health and safety and occurs within relevant guidelines, and obtaining consent is impractical; and
- where the agency or organisation no long requires the personal data for the purposes that it was collected, it must destroy the data or take reasonable steps to de-identify the data.

The opportunities presented by big data sets, including health data sets, has led to a preference for de-identifying rather than destroying data. If de-identified sufficiently, this means that organisations may still be able to use, share, and publish such information whilst preserving the privacy of individuals. As noted, the method of de-identification which is 'reasonable' in the circumstances depends on factors such as the type and sensitivity of the data, and whether the data is qualitative or quantitative. The matters for consideration of what is 'reasonable' in the circumstances include:

- the amount and sensitivity of the information;
- the nature of the organization (including size, resources and business model);

[3] Office of the Australian Information Commissioner, *Privacy Business Resource 4: De-identification of data and information* (https://www.oaic.gov.au/agencies-and-organisations/business-resources/privacy-business-resource-4-de-identification-of-data-and-information)

- possible adverse consequences to the individual if personal information is not destroyed or de-identified;
- the organisation's information handling practices (including whether it outsources or transfers information to third parties or overseas); and
- the practicability (such as time and cost).[4]

Under the Privacy Act, the OAIC can also approve for the purposes of the APPs, guidelines that are issued by the CEO of the National Health and Medical Research Council which relate to the use and disclosure of health information for the purposes of research, or the compilation or analysis of statistics, relevant to public health or safety. These guidelines sit side-by-side to (and not in place of) the Privacy Act and APPs. The most recent guidelines were published in 2014.[5]

In addition, the Productivity Commission identified in its recent Report on *Data Availability and Use* that there is a need for better guidance on robust de-identification.[6] The final report of the Productivity Commission was issued to the Commonwealth Government on 31 March 2017, and was released publicly on 8 May 2017.

Many Australian States have a health records statute, or a corresponding privacy statute which incorporates obligations about collecting, handling and providing access to health records. Generally, these statutes impose similar obligations to the Privacy Act, in that notification or consent of the individual is required for collection of health information; and destruction or de-identification of health information on end-of-useful-life of the information.

New Zealand similarly has a *Health Information Privacy Code* which falls under the ambit of the New Zealand Privacy Commissioner's regulation.[7] Generally the rules are similar to those in Australia: an organisation must notify an individual prior to collection of health information. When health information is no longer required, it must be transferred or destroyed in a manner that ensures its confidentiality.

5. Re-identification – what happened?

The difficulty of successfully de-identifying information was demonstrated in September 2016. Researchers from the Department of Computing and Information Systems at the University of Melbourne were able to re-identify certain de-identified PBS and MBS data published on data.gov.au (**Sample Health Data**).[8] The Sample Health Data constituted 10% of the MBS data collected between 1984 and 2014 and separate PBS data collected between 2003 and 2014. The data included in the set was selected randomly.

The identity of individuals associated with Sample Health Data was guarded in several ways:

- not including any names or addresses of doctors or patients; and
- encrypting the identification numbers of doctors and patients.

Details of the services provided were not encrypted.

Partial details of the encryption algorithm used to de-identify the identification numbers of doctors and patients was also published.

The researchers set about analysing the Sample Health Data with the aim of:

> ***'understanding mathematical facts about encryption and anonymization, in order to ensure that the security of government data is preserved in the face of inevitable efforts of external parties who may be prepared to break the law and attempt to re-identify the data' [9].***

[4] Office of the Australian Information Commissioner, APP guidelines, *Chapter 11: APP 11 – Security of personal information* (https://www.oaic.gov.au/agencies-and-organisations/app-guidelines/chapter-11-app-11-security-of-personal-information)

[5] National Health and Medical Research Council, *Guidelines approved under Section 95A of the Privacy Act 1988* (2014) (https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/pr2_guidelines_under_s95a_of_the_privacy_act_140311.pdf)

[6] Productivity Commission, *Data Availability and Use* (Report, March 2017) (http://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf)

[7] New Zealand Privacy Commissioner, *Health Information Privacy Code* (1994) (https://www.privacy.org.nz/assets/Files/Codes-of-Practice-materials/HIPC-1994-incl.-amendments-revised-commentary-edit.pdf)

[8] Dr Chris Culnane, Dr Benjamin Rubinstein and Dr Vanessa Teague, *Understanding the maths is crucial for protecting privacy* (September 2016) (https://pursuit.unimelb.edu.au/articles/understanding-the-maths-is-crucial-for-protecting-privacy)

[9] Dr Chris Culnane, Dr Benjamin Rubinstein and Dr Vanessa Teague, *Understanding the maths is crucial for protecting privacy* (September 2016) (https://pursuit.unimelb.edu.au/articles/understanding-the-maths-is-crucial-for-protecting-privacy)

The researchers used cryptographic methods to reverse the encryption algorithm used to de-identify doctor identification numbers, and were able to re-identify the doctor identification numbers (**Re-identification Event**). The researchers are reported not to have sought to re-identify the patient identification numbers.

The researchers notified the Commonwealth of the vulnerability on 12 September 2016; the data set was immediately removed from data.gov.au and the OAIC was notified. The OAIC has commenced an investigation to assess whether any personal information is or was at risk, and assess the adequacy of the process to de-identify the personal information it published. The outcome of this investigation will be made public upon its conclusion.[10]

6. A swift legislative response

The Attorney-General announced proposed amendments to the Privacy Act on 28 September 2016 and on 12 October 2016 the *Privacy Amendment (Re-identification Offence) Bill 2016* (Cth) (**Bill**) was introduced into the Senate. The Bill creates the following new offences:

- the intentional re-identification of de-identified personal information made available by a Commonwealth agency; and
- the intentional disclosure of re-identified personal information.

Data.gov.au publishes government information for both Federal and State governments, however the Bill relates only to information made available by Commonwealth agencies. In most States there is a separate privacy law applicable to State agencies. At present, no State has proposed an equivalent amendment to the Federal Bill.

Significant sanctions will apply to contraventions, both criminal (up to 2 years imprisonment and/or a fine of up to $21,600) and civil (up to $108,000).

The Bill also introduces offences of counselling, procuring, facilitating or encouraging another to re-identify de-identified personal information.

The Bill creates an obligation on an entity whose de-identified personal data has been re-identified to:

- notify the responsible Government agency of the re-identification; and
- comply with any directions from the agency about handling of the information.

Civil penalties have been proposed for failures to notify breaches and in this case the Australian Information Commissioner has jurisdiction to investigate the matter.

The Bill has a different focus to most other provisions of the Privacy Act, which apply only to 'agencies' and 'organisations', as defined by the Privacy Act. The Bill applies to individuals and small businesses, but does not apply to 'agencies', Commonwealth contracted service providers, entities that enter into agreements with agencies, and entities exempted by the Minister.

The amendments to the Privacy Act recognise that de-identification techniques may become susceptible to re-identification in the future and so there is a need to develop a network of non-technical data protections which support technical de-identification.

If passed, the provisions in the Bill may provide a deterrent against attempted re-identification, however there are some obvious limitations; an attempt must first be identified before an offence can be alleged and it is possible that re-identification may occur without ever having come to light (recall that the researchers at the University of Melbourne voluntarily informed the OAIC and the Department). There will also be practical difficulties in enforcing the proposed legislative provisions on persons situated outside of Australia.

The Bill was the subject of a review and report by the Legal and Constitutional Affairs Legislation Committee (**Committee**),[11] which received submissions raising the following concerns:

- that the risk of re-identification of government data sets may be too great to warrant releasing them;
- that criminalising re-identification activities may not be equitable and/or proportionate to other offences contained in the Privacy Act;
- that the offences proposed by the Bill are framed too broadly;
- that the Minister's power to exempt entities from the offences contained in the Bill are too broad;
- that retrospective operation of the Bill may not be appropriate;
- that the burden of proof to demonstrate that an exemption applies may not be appropriate.

The OAIC has also suggested that it is the responsibility of Government agencies to strengthen their management of privacy risks.[12]

[10] Privacy Commissioner, *Australian Privacy Commissioner's investigation into published MBS and PBS data sets* (29 September 2016) (https://www.oaic.gov.au/media-and-speeches/statements/australian-privacy-commissioner-s-investigation-into-published-mbs-and-pbs-data-sets)

[11] Senate Standing Committees on Legal and Constitutional, *Privacy Amendment (Re-identification Offence) Bill 2016, Report* (February 2017) (http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/PrivacyReidentification/Report)

[12] Allie Coyne, *OAIC tells govt to fix its privacy before criminalising data re-identification*, https://www.itnews.com.au/news/oaic-tells-govt-to-fix-its-privacy-before-criminalising-data-re-identification-445132

Nevertheless, the majority of the Committee recommend the Bill be passed. In dissent, the Labor and Greens senators argued that the Bill does not provide an appropriate balance between the need for privacy and the need to encourage research into areas including information security, cryptology and data analysis. Rather, they argue, it shifts the responsibility for protecting the privacy of individuals that are the subject of such information away from government agencies. The Labor and Greens senators also expressed concern about the retrospective application of the Bill. On balance, they recommend that the Bill not be passed.[13]

The Bill is presently before the Senate, and has not yet been presented to the House of Representatives.

The New Zealand Privacy Commissioner recommended amendments to the New Zealand Privacy Act (**NZ Privacy Act**) in a report released on 3 February 2017, including protections against the risk of re-identification of personal information.

In contrast to Australia's approach of criminalizing re-identification, the New Zealand Privacy Commissioner instead recommended:

- the addition of a privacy principle setting out limited circumstances in which re-identification of de-identified information can occur, to give individuals a right of action where harm is caused by the re-identification;
- the addition of provisions in the NZ Privacy Act which clarify the obligations of agencies in undertaking de-identification activities.

7. Implications for healthcare

The trust and confidence of citizens is key to unlocking the potential of big data in delivering enhanced health outcomes and system sustainability.

The Re-identification Event highlighted potential weakness in the method of de-identification and the security with which de-identification processes are guarded. The legislative response in criminalising de-identification of certain data may have some deterrent effect, however that alone will be insufficient. Robust and secure de-identification techniques, in conjunction with legal, policy and contractual controls, deployed across government and continually refined in response to changes in technology and approaches, are also needed.

In determining the appropriate portfolio of technical de-identification methods and non-technical governance controls, it is important to assess the risk of re-identification having regard to matters such as technology and the amount of information, along with the need to retain detail in the data for it to be useful for research purposes.

The potential benefits of big data in healthcare are growing as sources of information expand. The change to an 'opt out' model for the My Health Record is designed to ensure this becomes a ubiquitous part of the health administration system. The uptake in wearable devices has opened the door to even richer, real-time, insights to assist people in managing their own health and fueling innovation in the delivery of services. The continued development in this area is dependent upon public trust in the privacy of personal information.

A 2016 survey by Research Australia found that 90% of respondents were in favour of sharing de-identified health data to advance medical research and patient care.[14] This insight shows that Australians are aware of the benefits that greater use of health data can deliver, but not at the expense of putting their personal information at risk.

8. What to do?

Protecting personal health data requires a holistic approach, encompassing technical, legislative and process elements, which is difficult, but not insurmountable. The deterrent impact of the Bill has a role to play, but is not a complete solution.

New Zealand provides an example of a different model to resolve the tension between privacy and access to data. The Integrated Data Infrastructure (**IDI**) is a large research database containing microdata about people and households[15], and housing data from a range of agencies, including health information for the NZ population. Access to IDI requires an application and will only occur after the application has been assessed and the 'five safes' have been met:[16]

- Safe people – referee checks, signed secrecy declarations and strict protocols;
- Safe projects – must have statistical purpose and be in the public interest;
- Safe settings – access only permitted in a secure environment with no external connections;

[13] Senate Standing Committees on Legal and Constitutional Affairs, *Privacy Amendment (Re-identification Offence) Bill 2016, Dissenting Report of the Australian Labor Party and Australian Greens* (February 2017) (http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/PrivacyReidentification/Report/d01)

[14] Cited in Productivity Commission, *Data Availability and Use* (Report, March 2017) (http://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf)

[15] Stats NZ, *Integrated Data Infrastructure* (http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx)

[16] Stats NZ, *How we keep IDI data safe* (http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx)

- Safe data – de-identification and encryption is used rigorously; and
- Safe output – output must be made confidential before release.

Elements of this model could be adopted by Australian Governments to establish a more comprehensive approach, which includes:

- consistent application of robust de-identification methods, supported by advice and guidance from the OAIC and relevant statistical agencies, to promote better practice methodology;
- appropriate security of de-identification techniques and process of continual refinement to ensure that techniques continue to keep pace with new approaches;
- consideration of applying restrictions to access to data, gov.au along the lines of the New Zealand model;
- greater oversight of use of information and contractual mechanisms which limit the use and distribution of data and re-identification;
- providing researchers with analysis of data, rather than providing access to it, for example, running an analysis of the data and providing the result rather than the raw data;
- greater consumer/community education of and engagement with the potential benefits of big data; and
- strong legislative deterrence.

The selection of appropriate model elements is a manifestation of the tension between freedom of access and innovation and privacy, which is a societal choice that will need to be made by political leaders and supported by informed consumers (who are the data subjects). It would be useful for there to be a consistent approach across the country, which may mean that this is best resolved by the Council of Australian Governments.

The Re-identification Event highlights the challenges of de-identification of data to enable the promise of big data to be realised. It provides a timely call to action to develop a more comprehensive and risk-free approach to de-identification which extends beyond legislative deterrence, and which will ultimately help secure vital trust and confidence of citizens.

For a deeper discussion of how these issues might affect your business, please contact:

Tony O'Malley

Partner, Legal
+61 (2) 8266 3015
tony.omalley@pwc.com

Emily Prior

Partner, Big Data
+61 (2) 8266 0698
emily.prior@pwc.com

Simon Lewis

Director, Legal
+61 (2) 8266 2161
simon.lewis@pwc.com

Roland Fan

Director, Big Data
+61 (2) 8266 2033
roland.fan@pwc.com

Sylvia Ng

Director, Legal
+61 (2) 8266 0338
sylvia.ng@pwc.com

Steph Baker

Solicitor, Legal
+61 (2) 8266 5054
steph.baker@pwc.com

www.pwc.com.au

PwC refers to the Australia member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.

This content is for general information purposes only, and should not be used as a substitute for consultation with professional advisors.

Liability limited by a scheme approved under Professional Standards Legislation.

At PwC Australia our purpose is to build trust in society and solve important problems. We're a network of firms in 157 countries with more than 223,000 people who are committed to delivering quality in assurance, advisory and tax services. Find out more and tell us what matters to you by visiting us at www.pwc.com.au.

127049355