Opening Up Government Data for Big Data Analysis and Public Benefit

Author

Hardy, Keiran, Maurushat, Alana

Published

2017

Journal Title

Computer Law & Security Review

DOI

10.1016/j.clsr.2016.11.003

Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Opening up government data for Big Data analysis and public benefit

Keiran Hardy [a,*], *Alana Maurushat* [b,c,d]

[a] *School of Criminology and Criminal Justice, Griffith University Gold Coast Campus, Southport, QLD 4222, Australia*

[b] *Faculty of Law, University of New South Wales, Sydney, NSW, Australia*

[c] *Law and Policy Program, Data to Decisions Cooperative Research Centre, Adelaide, SA, Australia*

[d] *Australian Centre for Cyber Security, UNSW Canberra, Canberra, ACT, Australia*

Keywords:
Open data
Big Data
Privacy law
Open Government Partnership
Open Government Declaration
Open Data Charter
Freedom of information

A B S T R A C T

Governments around the world are posting many thousands of their datasets on online portals. A major purpose of releasing this data is to drive innovation through Big Data analysis, as well as to promote government transparency and accountability. This article considers the benefits and risks of releasing government data as open data, and identifies the challenges the Australian government faces in releasing its data into the public domain. The Australian government has ambitious aims to release greater amounts of its data to the public. However, it is likely this task will prove difficult due to uncertainties surrounding the reliability of de-identification and the requirements of privacy law, as well as a public service culture which is yet to fully embrace the open data movement.

1. Introduction

There is a growing global trend for government data to be released to the public as 'open data'. Open data is data which is accessible for free or at minimal cost, and which can be accessed by anybody and re-used for any purpose.[1] Ideally, this means that the data should be made available online under a creative commons licence and in a machine-readable format (i.e. one that is capable of being analysed by computers and software programs).[2]

Governments committed to the open data movement are posting many thousands of datasets on online portals.[3] The purpose of releasing this data is not merely (or even primarily) to provide information to the public, but rather to drive innovation through 'Big Data' analysis. The idea is that by releasing large government datasets, businesses, academia and the general public will be able to draw new insights from that

* *Corresponding author.* School of Criminology and Criminal Justice, Griffith University Gold Coast Campus, Southport, QLD 4222 Australia.
E-mail address: k.hardy@griffith.edu.au (K. Hardy).

[1] Australian Government Information Management Office, *The Australian Public Service Big Data Strategy* (Australian Government, 2013) 27 ('*APS Big Data Strategy*'); Open Knowledge, *Open Data Handbook: What is Open Data?* <http://opendatahandbook.org/guide/en/what-is-open-data/> (last accessed 13 July 2016).

[2] Government 2.0 Taskforce, *Engage: Getting on with Government 2.0 – Report of the Government 2.0 Taskforce* (Australian Government Information Management Office, 2009), xiv–xv.

[3] See, eg, Australian Government, Data.gov.au <data.gov.au> (last accessed 13 July 2016).

http://dx.doi.org/10.1016/j.clsr.2016.11.003

data and contribute innovative solutions to complex policy problems.

The Australian government aims to catch up with some of the world's biggest producers of open data, including the United States.[4] Whilst this goal remains optimistic, open data and other areas of digital innovation are likely to play a central role in the policy agenda of the Turnbull government.

The aim of this paper is to consider the benefits and risks of opening up government data to the general public, and to identify the challenges the Australian government faces in releasing its data as open data. The paper argues that the Australian government will likely face ongoing challenges in meeting its open data targets due to barriers that are technical, legal and cultural: namely, ongoing uncertainty as to the reliability of de-identification, broad protections for personal information in national privacy legislation, and a public service culture which continues to favour the secrecy of information over public release.

Part 2 explains the open data movement and how this relates to the phenomenon of Big Data. Part 3 considers the benefits and risks of releasing government data as open data. Part 4 identifies the steps taken by the Australian government to open up its data to the general public, and Part 5 identifies the major barriers that the Australian government will continue to face in trying to release greater amounts of its data into the public domain. Part 6 considers whether these barriers are likely to be overcome.

2. The Open Data movement and Big Data analysis

Increasingly, governments around the world are releasing the data their agencies collect through online portals. These datasets cover a range of topics including spatial, transport and public health data. The use of these public datasets is wide and varied, and in some countries the open data trend has been embraced on a large scale. For example, Barcelona was dubbed the world's first 'smart city', as it relies heavily on data released by Barcelona city council to drive innovative approaches to transport, health, education, and housing.[5]

More than 60 countries have now joined the Open Government Partnership (OGP), an international initiative launched by eight countries including the United States, the United Kingdom, and Norway.[6] The OGP requires participating countries to endorse its Open Government Declaration, which has the aim of making governments 'more transparent, responsive, accountable, and effective'.[7] Countries who join the OGP must develop an action plan on open data, prepare yearly self-assessments on how that plan is being implemented, and submit to the OGP's Independent Reporting Mechanism.[8]

In late 2015 the OGP published an International Open Data Charter, which at the time of writing had been signed by nine national governments, several other municipal governments, the World Bank, IBM and other organisations.[9] The Charter builds on the G8 Open Data Charter, which was signed by all G8 leaders in 2013. It comprises six key principles, which state that data should be:

(1) Open by default
(2) Timely and comprehensive
(3) Accessible and usable
(4) Comparable and interoperable
(5) For improved governance and citizen engagement
(6) For inclusive development and innovation[10]

Importantly, these principles do not include protecting individual privacy.

A key aim of this open data movement is to drive policy and business innovation,[11] so open data must be understood alongside the phenomenon of Big Data. Big Data, although difficult to reduce to a mutually agreed definition, is a popular term which captures the proliferation of large datasets in our technology-driven society, as well as the extraction of new information from large datasets through smart analytical tools.

Big Data is commonly defined according to the 'three Vs': Volume, Variety, and Velocity.[12] That is, vast amounts of data are being collected from a wide range of sources, and this information is being processed at high speeds, often in real-time. According to some definitions, Big Data refers more narrowly to the datasets being collected.[13] Other commentators use the term more broadly to describe the proliferation and analysis of large datasets as a social phenomenon.[14]

Not all analysis of open data will involve Big Data analytics, but Big Data analytics rely heavily on open data. In other words, the more government data that is available as open data, the greater the capacity for industry, academia and the general public to contribute to policy innovation through Big Data analysis. Like open data, Big Data 'enables businesses and governments to make informed, fact-based decisions about the complex world around us, create new products, reduce waste,

[4] Malcolm Turnbull, 'Benefiting from Big Data: The Government's Approach' (Speech delivered to Australian Information Industry Association Navigating Analytics Summit, Canberra, 11 April 2014).

[5] Ajuntement de Barcelona, *BCN Smart City* <http://smartcity.bcn.cat/en> (last accessed 13 July 2016).

[6] Open Government Partnership, *Open Government Partnership* <http://www.opengovpartnership.org> (last accessed 13 July 2016).

[7] Open Government Partnership, *Open Government Declaration* <http://www.opengovpartnership.org/about/open-government-declaration> (last accessed 13 July 2016).

[8] Open Government Partnership, *How It Works* <http://www.opengovpartnership.org/how-it-works> (last accessed 13 July 2016).

[9] Open Data Charter, *Adopted By* <http://opendatacharter.net/adopted-by-countries-and-cities/> (last accessed 13 July 2016).

[10] Open Data Charter, *Principles* <http://opendatacharter.net> (last accessed 13 July 2016).

[11] Open Knowledge, *Open Data Handbook: Why Open Data?* <http://opendatahandbook.org/guide/en/why-open-data/> (last accessed 13 July 2016); PricewaterhouseCoopers, *Deciding with Data: How Data-Driven Innovation is Fuelling Australia's Economic Growth* (PricewaterhouseCoopers, 2014).

[12] *APS Big Data Strategy*, above n 1, 8.

[13] Ibid.

[14] Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray, 2013), 6–7.

and plan intelligently for the future'.[15] Big Data and open data are therefore parallel and mutually reinforcing trends.

One of the biggest advantages – but also one of the biggest dangers – of Big Data is the capacity to predict behaviour. Big Data analysis involves collecting vast amounts of unconnected and disparate data, and applying machine learning pattern matching and other techniques to discern trends and patterns in that data. Predictive analytics can then extrapolate from those trends and patterns to predict the behaviour of a given population or even of individuals.[16]

Predictive analytics promise great benefits for many stakeholders: town planners can extract detailed patterns in the way people and goods move about urban environments; maintenance engineers can predict wear and tear from how infrastructure is being used; retailers can analyse shoppers' behaviour and preferences; law enforcement officials can identify suspicious communications suggestive of criminal preparations; and epidemiologists can analyse patterns in health and lifestyle data to better manage a population's future well-being.[17]

This also represents one of the biggest dangers of Big Data, as predictive analytics may lead to controversial overreach by the state. Taken to its extreme, for example, the idea that somebody could be arrested and punished based on statistical analysis as to what they are likely to do at some future time would fundamentally conflict with the presumption of innocence.[18] A more realistic possibility is that predictive analytics may lead to discrimination through racial profiling if police target certain communities that are believed to pose a greater risk of criminal behaviour.[19]

The use of predictive analytics by private companies also raises significant concerns about privacy and overreach. In one famous example, the US retailer Target analysed purchasing patterns to determine when a customer was pregnant. An angry father who complained to the company subsequently found out that the analytics were indeed accurate, and the daughter had not yet told him about her pregnancy.[20]

Another significant posed by Big Data is the possibility that individuals may be re-identified from large datasets and their personal details exposed. The greater the amount of data collected and analysed by governments and businesses, the greater the risk that individuals may be re-identified from that data. This can be possible because unique combinations of identifying characteristics can be revealed by cross-referencing disparate datasets. In one high-profile study, for example, two Netflix users were re-identified from anonymised data which was offered publicly by that company in order to improve its movie recommendation service.[21] These risks to privacy are likely to increase over time as techniques for re-identification become more sophisticated and greater amounts of government data are released as open data.

De-identification reduces these risks, although debates continue as to whether current techniques are sufficiently reliable to protect individuals from being re-identified from anonymised data.[22] De-identification is the removal, stripping or obfuscation of directly identifying elements from a dataset such that the data is not immediately identifiable as associated or linked with a particular individual. In other words, all names, addresses and other obvious identifying information are removed so that only aggregate or statistical data remains. The data may be either anonymised, meaning that all identifying details are removed, or 'pseudonymised', meaning that key identifying information is replaced with a code or other identifier which can only be re-associated with an individual if it is cross-referenced with a secure lookup table.[23]

Government data released as open data will be de-identified, but this cannot completely remove the risk that multiple data sources might be cross-referenced to re-identify an individual. To date, it seems that new techniques for de-identifying data have been met with equally innovative attempts at re-identification.[24] As discussed below, de-identification also poses significant challenges for governments in releasing their data as it is frequently unclear whether anonymised data is subject to the requirements of national privacy legislation.

3. Benefits and risks of releasing government data

There are three major benefits associated with the releasing of government data to the general public as open data. First, open data is seen as key to improving the effectiveness and efficiency of government policy and services. For example, data on crops, weather and geography might be analysed to improve current approaches to farming and industry, or data on hospital admissions might be analysed alongside demographic and census data to improve the efficiency of health services in areas of need.

The idea is that releasing such data will allow governments to provide better quality services which intelligently target sectors of the population most in need of those services. This will lead to economic benefits if governments can provide the same quality of services at lower cost. It has been estimated that policy innovation based on open data could

[15] PricewaterhouseCoopers, above n 11, 2.
[16] See Mayer-Schönberger and Cukier, above n 14, 52–60.
[17] Ibid.
[18] Ibid., 159.
[19] Ibid., 160.
[20] See Mayer-Schönberger and Cukier, above n 14, 57–58.
[21] Robert Lamos, 'Researchers Reverse Netflix Anonymization', *Security Focus*, 4 December 2007 <http://www.securityfocus.com/news/11497>.
[22] Ann Cavoukian and Daniel Castro, *Big Data and Innovation, Setting the Record Straight: De-Identification* Does *Work* (Information and Privacy Commissioner, Ontario, 2014); Jules Polonetsky, Omar Tene and Kelsey Finch, 'Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification' 56 *Santa Clara Law Review* 593; Daniel C Barth-Jones, 'Does De-Identification Work Or Not?', *Fierce Big Data*, 23 June 2014.
[23] Simson L. Garfinkel, *De-Identification of Personally Identifiable Information* (National Institute of Standards and Technology, United States Department of Commerce, 2015), 5.
[24] Polonetsky, Tene and Finch, above n 22, 594.

benefit the Australian economy by up to $16 billion per year.[25]

The second major benefit is that open data is making gains in transparency and accountability, as a greater proportion of government decisions and operations are being shared with the public. These democratic values are made clear in the OGP's Open Government Declaration, which aims to make governments 'more open, accountable, and responsive to citizens'.[26]

The third benefit is that open data can improve democratic participation by allowing citizens to contribute to policy innovation. Events like GovHack,[27] an annual Australian competition in which government, industry and the general public collaborate to find new uses for open government data, epitomise a growing trend towards service delivery informed by 'user input'. For example, the winner of the 'Best Policy Insights Hack' at GovHack 2015 developed a software program for analysing which suburbs are best placed for rooftop solar investment.[28]

At the same time, the possibility of re-identification from government data means that the open data movement poses risks to individual privacy. Currently, much of the open data available is spatial (geographic or satellite) data,[29] which is relatively unproblematic to post online as it poses minimal privacy risks. However, for the full benefits of open data to be gained, this spatial data needs to be supplemented with information on welfare payments, hospital admission rates, income tax assessments and other potentially sensitive areas of government policy and administration which could drive innovation. As more and more datasets from these disparate policy areas are posted online, the greater the risk that individuals may be re-identified and their personal details exposed. Whilst debates continue over the reliability of de-identification,[30] even a small risk of exposing individuals' private details is a significant consideration for governments seeking to open up their data to the general public.

4. Australian government policy on open data

Both of the major political parties in Australia have taken important steps in line with the open data movement. In 2010, the Gillard Labor government issued a Declaration of Open Government similar to that endorsed by the OGP.[31] In May 2013, the Gillard announced that Australia would join the OGP and began developing a national action plan on open data.[32] The Gillard government also launched Australia's online data portal: data.gov.au.

After a change in government, the conservative Liberal-National Coalition government led by Tony Abbott built significantly on these efforts. Malcolm Turnbull, when he was Communications Minister, announced that the Abbott government had 'very ambitious targets' with regard to open data and was 'committed to turning around our slow start'.[33] The Abbott government significantly increased the number of government datasets available online, and it launched the Australian Open Data 500, a register of private sector companies that rely on open government data.[34] However, work on Australia's national action plan stalled during this period, and it was not clear if Australia would meet the requirements of the OGP.

Some of Malcolm Turnbull's earliest changes in the Prime Ministership signalled that he would place digital policy and innovation at the centre of the Coalition government's policy agenda. The new Communications Minister, replacing Turnbull, was bestowed with the additional title of 'Minister Assisting the Prime Minister for Digital Government'. The Digital Transformation Office, which was established in mid-2015 to update government services for the digital era, was moved from the Communications portfolio into the Department of Prime Minister and Cabinet (PM&C). Three separate policy units relating to government data (from the Departments of Finance and Communications) were also merged into a Public Data Branch within PM&C.[35]

In November 2015, Turnbull renewed Australia's commitment to the OGP and opened a consultation process with the goal of submitting a national action plan to the OGP by mid-2016.[36] The consultation is led by Pia Waugh, head of the Data Infrastructure team within the PM&C's new Public Data Branch. A workshop was held in April 2016 to draft a national action plan to submit to the OGP.[37]

Individual Australian States have also taken a proactive approach. In late 2015, the New South Wales Parliament passed the *Data Sharing (Government Sector) Act 2015* (NSW). That Act established a new Data Analytics Centre, which will aggregate and analyse data from New South Wales government agencies, state-owned corporations and local councils. In addition, each of the Australian States and the Australian Capital

[25] Lateral Economics, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target* (Lateral Economics and Omidyar Network, 2014), ii.

[26] Open Government Partnership, *Open Government Declaration* <http://www.opengovpartnership.org/about/open-government-declaration> (last accessed 13 July 2016).

[27] GovHack, *GovHack 2016* <https://www.govhack.org> (last accessed 13 July 2016).

[28] GovHack Hackerspace 2015, *Synergising Synergies for Sitizens* <http://2015.hackerspace.govhack.org/content/synergising-synergies-sitizens> (last accessed 13 July 2016).

[29] See Australian Government, *Datasets* < https://data.gov.au/dataset> (last accessed 13 July 2016).

[30] See Cavoukian and Castro, above n 22; Polonetsky, Tene and Finch, above n 22; Barth-Jones, above n 22.

[31] Australian Government Department of Finance, *Declaration of Open Government* (16 July 2010) <http://www.finance.gov.au/blog/2010/07/16/declaration-open-government/>.

[32] Australian Government Department of Finance, *Australian Joins the Australian Government Partnership* (22 May 2013) <http://www.finance.gov.au/blog/2013/05/22/australia-joins-open-government-partnership/>.

[33] Ibid.

[34] Open Data 500 Global Network, *Open Data 500 Australia* <http://www.opendata500.com> (last accessed 24 March 2016).

[35] James Riley, 'Data policy unit created in PM&C', *Innovation Australia*, 2 November 2015 <http://www.innovationaus.com/2015/11/Data-policy-unit-created-in-PM-C>.

[36] 'Turnbull Signs Australia Up to Open Government Partnership', *The Mandarin*, 18 November 2011 <http://www.themandarin.com.au/57015-turnbull-signs-australia-open-government-partnership/>.

[37] Australian Government, *Open Government Partnership – Australia* (2016) <https://ogpau.govspace.gov.au/>.

Territory (ACT) maintains an Open Data Portal similar to that maintained by the federal government.[38]

Australia's commitment to open data will likely be at its strongest under the Turnbull government, although the extent to which greater amounts of government data will be released to the public remains unclear. Currently, Australia lags behind other countries in posting government data online. As of July 2016, there were 8500 datasets available on data.gov.au. By contrast, the UK government by this time had posted nearly 34,000 datasets on its national portal (data.gov.uk), and the United States an astonishing 184,000 (data.gov).

This can partly be explained by the size of different administrations. It is very doubtful, despite Turnbull's aspirations, that Australia will ever 'catch up' with the United States in releasing such enormous quantities of data.[39] Nonetheless, several government and independent reports have observed that Australian government agencies remain reluctant to release their data to the general public.[40] The major reasons for this reluctance are identified below.

5. Barriers to releasing open data

As in other countries, there are good reasons why Australian government agencies do not release the sensitive data they collect. It would clearly be irresponsible for government agencies to release intelligence assessments, health records, credit records, tax file numbers or other information, the release of which would endanger national security, invade privacy or facilitate identity theft. For this reason, these categories of data are subject to stringent privacy protections, including in some cases the possibility of criminal penalty for disclosure.[41]

At the same time, there is a strong case for releasing much of the statistical and policy data produced by government agencies in order to drive innovation and improve transparency and accountability. Such data could include, for example, detailed statistical information on income tax assessments, welfare payments and hospital admissions in different local government areas across Australia. Releasing such data could allow researchers and private organisations to develop new, innovative approaches to public health and welfare policy.

At the same time, such benefits come with additional risks to privacy. It is these risks which drive the three main barriers to opening up Australian government data. These barriers are technical, legal and cultural. These barriers are not unique to Australia, although given that Australia has released less datasets than other countries like the UK and United States, it is useful to consider these barriers in the Australian context and whether they are likely to be overcome.

The technical barriers relate largely to the problems, discussed above, of adequately de-identifying government data. Ongoing uncertainty about the reliability of de-identification is compounded because Australian government agencies decide internally whether to open their datasets, and often this is done without the relevant expertise (such as in data scrubbing and data mining). This not only places unrealistic expectations on agencies to appropriately de-identify and release their own data, but also poses significant risks to privacy if government data is de-identified without sufficient expertise.

Government agencies also face challenges in formatting data according to common metadata standards. This standardisation is important so that government data can be easily searched and analysed. Metadata in this context means a 'structured description of the content, quality, condition or other characteristics of data'.[42] In other words, metadata is important identifying information about a dataset, such as its author, title, date and time last modified, and a summary of its content. Such information also provides important information about the provenance of government data, so that users can be assured of its quality. Currently, Australian government agencies lack such consistency in how their data is formatted.[43]

There are also significant legal barriers to opening up Australian government data. The *Privacy Act 1988* (Cth) (Privacy Act), Australia's national privacy legislation, regulates the collection, use, disclosure and storage of 'personal information'. The requirements of the Privacy Act apply to federal government departments and large businesses.[44]

Personal information is defined in the Privacy Act as 'information or an opinion about an identified individual, or an individual who is reasonably identifiable'.[45] Obvious examples of personal information include names, addresses, and dates of birth. Such identifying information should clearly be removed from any government data before it is released into the public domain. Personal information can only be disclosed under a specified exemption,[46] such as where an agency reasonably believes that the disclosure of information is necessary to lessen or prevent a serious threat to life, health or safety.[47]

The requirement for government agencies to protect personal information becomes complicated due to uncertainty surrounding the reliability of de-identification. The words 'reasonably identifiable' in the definition of personal information mean that the Privacy Act requirements can also apply to de-identified, aggregate data. If it would be reasonably possible to re-identify an individual from de-identified data, either by

[38] See, eg, NSW Government, *Data NSW: Providing Access to NSW Government Data* <data.nsw.gov.au> (last accessed 13 July 2016); Queensland Government, *Queensland Government Data* <data.qld.gov.au> (last accessed 13 July 2016); Victoria State Government, *Discover and Access Victorian Government Open Data* <data.vic.gov.au> (last accessed 13 July 2016).

[39] Turnbull, above n 4.

[40] PricewaterhouseCoopers, above n 11, 21; Government 2.0 Taskforce, above n 2, 51; Australian Government Information Management Office, *National Government Information Sharing Strategy: Unlocking Government Information Assets to Benefit the Broader Community* (2009), 7 ('*Information Sharing Strategy*'); Australian Government Information Management Office, *Australian Government Information Interoperability Framework: Sharing Information Across Boundaries* (2006) 20 ('*Information Interoperability Framework*').

[41] See, eg, *Intelligence Services Act 2001* (Cth), pt 6.

[42] Australian Institute of Health and Welfare. 2016. *METeOR: Metadata Online Registry – About Metadata* <http://meteor.aihw.gov.au/content/index.phtml/itemId/181414> (last accessed 13 July 2016).

[43] Productivity Commission, *Annual Report 2012-13* (Australian Government, 2013), 12.

[44] *Privacy Act 1988* (Cth) ss 6, 6C.

[45] *Privacy Act 1988* (Cth) s 6.

[46] *Privacy Act 1988* (Cth), ss 16A, 16B.

[47] *Privacy Act 1988* (Cth), s 16A, Item 1.

reversing the de-identification techniques or by combining that data with *other* sources of data, it would not be lawful under the Privacy Act for government departments to disclose that information.[48] This broad approach to defining personal information is similar to that adopted by the UK and European Union,[49] but differs from that in the United States, where personal information is typically defined narrowly according to specified categories such as names, addresses and social security numbers.[50]

Even an agency's best efforts at de-identifying its data may not prevent that data from being combined with other sources of information to re-identify an individual. Because of this risk, it is difficult for Australian government agencies to be confident that they have adequately de-identified their data for the purposes of privacy law. Reluctance to release data under the Privacy Act has become so common that the acronym 'BOTPA' (Because Of The Privacy Act) has become shorthand for agencies refusing to share their data, even with other government agencies and where disclosure would not actually breach the requirements of the Privacy Act.[51]

The third major barrier to opening up government data is a public service culture which appears to favour secrecy of information as the default position. This has parallels in ongoing debates about Freedom of Information (FOI) legislation,[52] and the different approaches to FOI taken by the Labor and Liberal governments. The Labor governments implemented reforms to improve public service culture around FOI, such as by creating a scheme for the proactive release of government information and establishing the Office of the Australian Information Commissioner (OIAC). By contrast, the Abbott government introduced a Bill to disband the OIAC,[53] and was more generally recognised for its secretive approaches to implementing government policy (particularly with regard to asylum seekers and national security).

The Abbott government's preference for maintaining secrecy of government information seemed to pervade the culture of its public service. One FOI and privacy law specialist described the term of the Abbott government as 'dark days' for FOI, remarking that 'no prizes or honours' were being awarded for public servants who efficiently complied with FOI requests.[54] This was in part driven by fears of what might be exposed by releasing government information, with public servants reluctant to put decision-making processes down on paper for fear of these being released into the public domain.[55]

With regard to open data, a culture resistant to releasing government information appears to be driven by several similar factors, including:

- a generational preference amongst public service management for maintaining secrecy of information, whereas younger generations expect that data should be made freely available;
- concerns about the quality or accuracy of information being released;
- fear that mistakes or misconduct on behalf of government employees might be exposed;
- limited understanding of the benefits that can be gained from open data; and
- a lack of leadership to help drive the open data movement[56]

These factors may be more relevant to the protection of sensitive data by law enforcement and intelligence agencies than in preventing other government agencies from releasing data on the implementation of their policies.[57] Reluctance to release such policy data appears to be driven to a large extent by concerns about protecting individual privacy.[58] This is a positive sign, as it suggests that the reluctance of Australian government agencies to open up their data is not wholly driven by self-interest or a desire for secrecy, and could be overcome if sufficient confidence in de-identification can be developed.

6. Future challenges for open data in Australia

These technical, legal and cultural hurdles will pose ongoing challenges for the Australian government in seeking to release greater amounts of its data as open data.

As for other countries seeking to open up greater amounts of their data, achieving sufficiently reliable de-identification procedures in Australia will take time – particularly for the cost of these techniques to reduce sufficiently so that all government agencies can apply them on a larger scale. One way to overcome these difficulties would be to set up a group of government experts (or for the government to employ a third party contractor) which is responsible for aggregating and de-identifying government data and releasing that data to the public. This would work most appropriately as a core branch within the Digital Transformation Office, with the goal of

[48] Unless a specified exemption applies: *Privacy Act 1988* (Cth), ss 16A, 16B.

[49] Council of Europe, Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, 28 January 1981 (Strasbourg), ch 1 art 2 < http://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680078b37>.

[50] United States Government Accountability Office, *Privacy: Alternatives Exist for Enhancing Protection of Personally Identifiable Information* (GAO, 2008) 1; United States Senate Bill No 1386, cl 1(e), available at: <http://www.leginfo.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.pdf>.

[51] Government 2.0 Taskforce, above n 2, 51; *Information Sharing Strategy*, above n 40, 15.

[52] *Freedom of Information Act 1982* (Cth).

[53] Freedom of Information Amendment (New Arrangements) Bill 2014 (Cth).

[54] Stephen Easton, 'FOI Laws: Fixing the Chilling Effect on Frank Advice', *The Mandarin*, 18 June 2015 <http://www.themandarin.com.au/40043-abbott-takes-secrecy-new-heights-public-servants-care/?pgnc=1>.

[55] Ibid.

[56] *Information Interoperability Framework*, above n 40, 20–21; *Information Sharing Strategy*, above n 40, 7; Government 2.0 Taskforce, above n 2, 49–51.

[57] Data to Decisions Cooperative Research Centre, *Review of Barriers to Open Data and Related Re-Use of Information in Five Exemplar Federal Data Sets* (Data to Decisions CRC, November 2014, Draft Report on file with authors).

[58] Ibid.

meeting Australia's open data targets. That group should focus significant resources on developing reliable de-identification techniques that can compete with and overcome innovations in re-identification.

A standardised technical approach along these lines would also ensure that Australian government data is formatted and released according to common metadata standards.[59] This is the purpose of the Australian Government Technical Interoperability Framework – which sets out national standards for consistently describing and encoding government data – but this framework is now more than a decade old and should be updated.[60]

In terms of the legal barriers, uncertainty over the Privacy Act requirements could be resolved quickly and directly by removing the words 'reasonably identifiable' from the definition of personal information. However, this would improve access to open data by significantly reducing existing privacy protections. It would allow the release of any de-identified government data, regardless of whether that data could be combined with other sources to identify private details about individuals. This would significantly increase the risks that Australian citizens might be re-identified from open government data. A preferable approach, as outlined above, is to focus on achieving more reliable de-identification of government data than to relax existing privacy protections.

In terms of the cultural challenges, new initiatives like the Digital Transformation Office will help to drive a change in public service culture towards a more transparent, innovative government. The Turnbull government is tasking a younger, tech-savvy generation of public service employees with devising digital solutions to difficult policy problems. Events like GovHack suggest that such policy innovation will also depend on contributions from members of the general public, who can apply their knowledge of coding, smart-phone applications and other technology to open government data.

These are very positive signs for the open data movement in Australia, although it will likely take some years for this kind of approach to be adopted across the public service as a whole, and it may in any case be resisted by sectors of the public service which are more resistant to change. Given the sensitive nature of much data held by government, there are also inherent limits to how much government data can eventually be released.

Despite the Australian government's strong commitments to the open data movement, then, it will likely take a significant time – certainly several terms of government and perhaps up to a decade or more – to achieve something resembling the open data ideal. Developing a public service culture which freely and reliably publishes a large proportion of its data, so that the general public can contribute innovative solutions to complex policy problems, is a long-term project in democracy and accountability.

7. Conclusion

In their day-to-day operations, government agencies produce a huge volume and variety of data about individuals and society. The open data movement aims to release this information to the general public in order drive policy innovation and to improve transparency and accountability.

The open data movement is closely related to the phenomenon of Big Data, which focuses on analysing large datasets to develop innovative approaches to policy and business. The greater release of open data to facilitate Big Data analysis could increase productivity across a range of sectors including agriculture, property services, construction, health, transport, utilities, and mining.[61] As the Australian Public Service Big Data Strategy explains:

> *Big data analytics can be used to streamline service delivery, create opportunities for innovation, and identify new service and policy approaches as well as supporting the effective delivery of existing programs across a broad range of operations.*[62]

In other words, government data must be 'out there' in the public domain for it to be fed into Big Data systems and the full benefits of Big Data technologies to be gained. In addition to economic benefits, greater access to government data will improve the transparency and accountability of governments by exposing a greater proportion of their operations to public scrutiny. Open data can also improve democratic participation, as members of the general public can analyse that data and contribute innovative policy solutions.

At the same time, these benefits need to be weighed against the privacy risks in releasing opening up greater amounts of government data to the general public. Because of uncertainty over the reliability of de-identification, agencies remain reluctant to release the data they collect because it might be combined with other data sources to reveal private details about Australian citizens. This reluctance stems from the definition of personal information in the Privacy Act, which prevents the disclosure of information from which an individual is 'reasonably identifiable'.[63]

The definition of personal information in the Privacy Act is a major hurdle to releasing government data in Australia. However, until such time as de-identification techniques are considered sufficient to prevent re-identification, the words 'reasonably identifiable' in that definition will play an important role in protecting Australian citizens from possible large-scale invasions of privacy. The Australian government should focus on developing more reliable techniques for de-identification, and not on amending the Privacy Act simply to permit the release of government data from which obvious identifying information has been removed.

[59] Productivity Commission, above n 43, 12.

[60] Interoperability Framework Working Group, *Australian Government Technical Interoperability Framework* (Australian Government, 2005).

[61] Government 2.0 Taskforce, above n 2, 43.

[62] *APS Big Data Strategy*, above n 1, 5.

[63] *Privacy Act 1988* (Cth) s 6.

The other major hurdle to releasing government data is a public service culture which is yet to fully embrace the open data movement. It is likely that this culture will change under the Turnbull government, which sees digital innovation as a central pillar of its policy agenda. However, it will likely take some years to achieve such cultural change across the public service as a whole. Despite the Coalition government's commitment to turning around Australia's 'slow start',[64] then, equalling other countries' commitment to the open data movement is likely to be a slow process.

It is encouraging that much of the Australian government's concerns about releasing data to the public relate to protecting individual privacy. The slower release of datasets in Australia compared to other countries may therefore not be a failure, but rather a virtue from which other countries can learn. The important and difficult task for the future will be learning how to strike the right balance between releasing government data for public benefit whilst ensuring that this same level of concern for individual privacy is maintained.

[64] Turnbull, above n 4.