International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:10, No:4, 2016

Opening up Government Datasets for Big Data Analysis to Support Policy Decisions

K. Hardy, A. Maurushat

***Abstract*—**Policy makers are increasingly looking to make evidence-based decisions. Evidence-based decisions have historically used rigorous methodologies of empirical studies by research institutes, as well as less reliable immediate survey/polls often with limited sample sizes. As we move into the era of Big Data analytics, policy makers are looking to different methodologies to deliver reliable empirics in real-time. The question is not why did these people do *this* for the last 10 years, but why are these people doing *this* now, and if the *this* is undesirable, and how can we have an impact to promote change immediately. Big data analytics rely heavily on government data that has been released in to the public domain. The open data movement promises greater productivity and more efficient delivery of services; however, Australian government agencies remain reluctant to release their data to the general public. This paper considers the barriers to releasing government data as open data, and how these barriers might be overcome.

***Keywords*—**Big data, open data, productivity, transparency.

I. Introduction

THE capture and analysis of data is growing exponentially. Advancements in technology and its embedding in modern society mean that more and larger datasets are available to use and analyse. "Big Data" is a term which captures the proliferation of these datasets, as well as the extraction of information from large datasets through smart analytical tools. Often, this information is extracted in real-time, and correlations are drawn from disparate datasets to drive innovative approaches to policy and business. Big Data is, therefore, seen as key to improving the efficiency and effectiveness of services provided by both governments and the private sector. Examples range from smartphone applications which monitor public transport and traffic conditions, to software which matches lenders and investors by more accurately assessing credit risk, to targeted policies for addressing social disadvantage. In other words, big data analysis "enables businesses and governments to make informed, fact-based decisions about the complex world around us, create new products, reduce waste, and plan intelligently for the future" [1].

While much discussion has focused on the possible future applications of big data technologies, the capacity for data analysts to extract information from large datasets depends on a more immediate, practical issue: the opening up of datasets held by government agencies to the general public. In the day-to-day administration of government, agencies produce a huge volume and variety of data about individuals and society. This 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.

For good reason, much of the data produced by government is not released publically. Security and privacy concerns mean that personal, sensitive, and classified data is rightly protected. In other cases, such as information relating to weather, transport and government spending – there is a strong case for releasing data held by government to drive innovation and economic growth.

There is a growing trend around the world for government data to be released to the public through online portals. However, Australia lags behind other countries in embracing this open data movement. At October 2015, there were 7,400 datasets available on the Australian government's open data portal (data.gov.au). By contrast, the UK government by this time had posted 24,000 datasets on its national portal (data.gov.uk), and the United States an astonishing 187,000 (data.gov). This can partly be explained by the different sizes of these governments and population, and the amount of data they collect, although several government and independent reports have noted that Australian government agencies remain resistant to sharing their data with other agencies or releasing it as "open data" [1]-[5]. The reasons for this apparent reluctance will be explored below.

The aim of this paper is to identify the barriers to releasing government data as open data, and to consider how these barriers might be overcome. Part II explains the relationship between big data and open data. Part III outlines the barriers that prevent the release of government data for public benefit. The particular legislation and policies mentioned are specific to Australia, but otherwise the barriers discussed are more generally relevant to the open data movement, and may equally be present in other countries. Part IV considers how some of these barriers may be overcome. It sets out some steps that could help achieve the cultural change necessary in government agencies to facilitate the release of open data on a larger scale, whilst ensuring that individual privacy remains protected.

II. Big data and the Open Data Movement

The volume of data being generated by governments and businesses about individuals and the world we live in is increasing exponentially. According to some estimates, around 2.5 quintillion bytes of data are being generated each day, and

K. Hardy is with the Faculty of Law, University of New South Wales (phone: +61 2 9385 3445; fax: +61 2 9385 1175; e-mail: k.a.hardy@unsw.edu.au).

A. Maurushat is with the Faculty of Law, University of New South Wales, the Data to Decisions Cooperative Research Centre, and the Australian Centre for Cyber-Security (e-mail: a.maurushat@unsw.edu.au).

up to 90% of available data has been generated in the last few years [6]. In some cases, the collection of this data is explicit – such as when an individual provides personal details to a government agency or business in order to receive some benefit in return. More often – such as when our movements are tracked by networked sensors in smartphones and automobiles, or our Internet searches, online shopping habits and social media interactions are logged – the collection of data is less conspicuous. Reductions in the cost of data storage, improvements in computer processing speed, and the development of smarter analytic techniques mean that these ever-increasing amounts of data are being constantly stockpiled and analysed to drive policy and business innovation.

"Big Data" although difficult to reduce to a mutually agreed definition, is a popular term which captures these developments. Big Data is commonly defined according the "three Vs": Volume, Variety, and Velocity [7]. 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 data being collected [7]. These data may be structured (organised into formal databases), semi-structured (not organised into formal databases, but containing metadata or other identifying tags), or unstructured (having no identifiable structure, such as collections of images or text) [7]. Most definitions of Big Data include smart analytics and automation whereby there is a range of data analytical tools able to extract information, and glean patterns for a range of uses producing added value where innovative information technologies are combined with evolving mathematical approaches [19]. The data is processed in real-time with machine learning (artificial intelligence) algorithms [20]. In any of these cases, the data may also be incomplete. Other commentators use the term "Big Data" more broadly to describe the proliferation and analysis of large datasets as a social phenomenon [8].

However Big Data is defined, the implication is that large datasets are being analysed through a range of smart analytics which draw new insights to drive policy and business innovation. In practice, this usually means that advanced technologies, software, and algorithms are used to identify correlations across disparate datasets. For example, a retail chain might identify that most of its customers have a similar number of contacts on social media. Even though there is no apparent connection between the size of friendship circles and shopping in that particular store, the store might direct its marketing to others with a similar number of online contacts. While based on correlation rather than causation, predictions drawn from these kinds of insights can be alarmingly accurate: in one famous example, the retail chain Target sent marketing materials for baby products to a family before the family knew that their daughter was pregnant [8].

The analysis of large datasets can also drive innovative approaches to policy. For example, geographical and census data might be combined with information from other sources to identify new factors which indicate a community is more likely to suffer social disadvantage. Data on crops, livestock, and weather might improve current approaches to farming and industry, or data on hospital admissions and government spending might be used to improve the efficiency of health services in areas of need. The potential applications of big data analysis are wide, and could increase productivity across a range of sectors, including agriculture, property services, construction, health, transport, utilities, and mining [2]. As the Australian Public Service Big Data Strategy explains [7]:

> 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.

This is where 'open data' becomes important, as big data tools rely heavily on publically available data. Open data may be defined as data which is accessible for free or at minimal cost, without limitations as to user identity or intent [7]. Ideally, this means that the data should be made available online in a digital, machine readable format [7]. Discussions of open data often overlap with discussions about Public Sector Information (PSI), being information that is generated by or for government or other public institutions [2], [7]. The more PSI that is available as open data, the greater the capacity for industry, academia, government and the general public to draw new insights from that data and improve service delivery. Big Data and Open Data are therefore parallel and mutually-reinforcing trends.

Governments and organisations are increasingly opening up their datasets, making more of their data available so that businesses, academia and the general public can analyse that data and devise new services and technologies. As mentioned in the introduction, governments committed to releasing open data are doing so by posting many thousands of datasets on online portals. These datasets cover a range of topics including spatial, environmental 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, the release of data held by the Barcelona City Council has led Barcelona to be dubbed the first "smart city", as it relies on open data to drive more innovative approaches to transport, health, education and housing [9].

Australia lags behind comparable countries in embracing this trend. The Labour government committed to the open data movement by signing the Declaration of Open Government in July 2010, by joining the Open Government Partnership in 2013, and by launching the government's online data portal. However, it is unclear to what extent this policy will be supported by the Liberal Coalition government currently in office. This is despite the fact that "more vigorous open data policies" could be worth more than $64 billion per annum to the Australian economy [10]. The recent change in leadership may suggest that the Turnbull government will favour greater transparency compared to the Abbott government, which was known for its secretive approaches to immigration and national security, although this remains to be seen.

The Open Government Partnership is "an international platform for domestic reformers committed to making their governments more open, accountable, and responsive to citizens" [11]. It requires participating countries to develop an action plan in line with the Partnership's objectives, prepare yearly self-assessments on how that plan is being implemented, and submit to the Partnership's Independent Reporting Mechanism [11]. The Partnership is now joined by over 60 countries – including not only the major western countries but also some of Africa, much of South America and other countries with more chequered human rights records. The open data movement is therefore about more than the instrumental benefits that can be gained by improving big data analysis. It is also linked to ideals of transparency and accountability, as governments are increasingly exposing a greater proportion of their decisions and operations to the wider public.

Open data may also strengthen democratic participation, as greater availability of government data means greater capacity for the general public to contribute to policy and business innovation. Events like GovHack, 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' [2].

III. Barriers to Releasing Government Data as Open Data

The release of government data to the general public is beneficial, then, not only because it facilitates big data analysis, but also because it contributes to greater accountability and transparency of government. Despite this, there are several reasons – many of them valid – why governments may not want or be able to release data they have generated into the public domain. This section identifies these barriers to releasing government data as open data, and Part IV suggests ways that some of these barriers might be overcome.

The most obvious barrier preventing the disclosure of government data is privacy legislation. In Australia, government agencies and large businesses are subject to the requirements of the *Privacy Act 1988* (Cth) ('Privacy Act'), which regulates the collection, use, disclosure, and storage of 'personal information'. Personal information is defined as 'information or an opinion about an identified individual, or an individual who is reasonably identifiable' (Privacy Act, Section 6). The most obvious examples of personal information include names, telephone numbers, date of births, and residential addresses. Because this definition encompasses information from which a person is *reasonably identifiable*, protections under the Privacy Act may also apply to data which has been de-identified (i.e. from which names, addresses and other identifying information have been removed). This may be because the information has not been reliably de-identified, so it is still possible to extract a person's identity from the remaining data; or because the de-identified data can be combined with *other* sources of information to reveal a person's identity. Some of the issues with this approach are discussed in Part IV.

Personal information may only be disclosed by government agencies if the disclosure satisfies certain criteria. The individual the subject of the information must either consent to the information being disclosed (Privacy Act, schedule 1, Australian Privacy Principle 6.1) or the disclosure must fall within one of several specified exemptions. The broadest of these exemptions is that the individual would reasonably expect the entity to use or disclose the information for a secondary purpose *and* the secondary purpose is related to the primary purpose for which the information was collected (Privacy Act, schedule 1, Australian Privacy Principle 6.2). This exemption does not facilitate the public release of government data for big data analysis, as individuals would not reasonably expect their personal information to be disclosed for that purpose, and big data analysis would not likely be related to the original purpose for collecting the information. Other exemptions for disclosing personal information under the Privacy Act relate to situations which involve suspicion of unlawful activity or serious misconduct, or threats to life, health, or safety (Privacy Act, Section 16A).

Stronger statutory protections apply to 'sensitive information', being personal information that also relates to the individual's racial or ethnic origin, political opinions, religious beliefs, sexual orientation, or similar categories (Privacy Act, Section 6). Sensitive information also includes health, genetic and biometric information (Privacy Act, Section 6). In addition, the Privacy Act sets out more detailed schemes for protecting credit reporting information and tax file numbers. Other data which contain private details about individuals, including health records, financial transactions and telecommunications data (metadata), are subject to further privacy protections in other legislation. For example, metadata retained by service providers can only be disclosed to specified investigative agencies if doing so would be reasonably necessary to investigate a criminal offence, find a missing person, or enforce a law which imposes a pecuniary penalty or protects the public revenue (Telecommunications (Interception and Access) Act 1979 (Cth), Sections 178-179).

Government agencies are also prohibited from disclosing information which would harm Australia's security interests. The Australian Government Protective Security Policy Framework sets out guidelines for government agencies which handle classified material [12] and the employees of the Australian intelligence agencies are subject to severe penalties for disclosing (or even unlawfully copying) information obtained in the course of their employment (Intelligence Services Act 2001 (Cth), Part 6).

There are many security concerns when using and storing large volumes of data. Current data storage systems often classify types of data in order to comply with appropriate legal frameworks. It has been argued that information classification becomes even more critical when you are drawing from disparate data sources in large volumes as organisations are no longer sticking to their own datasets, but are incorporating third party datasets. Who owns the data? Who is responsible for the data? What are the security standards required for the data? Attributed based encryption may be required for use of

some types of data, or when sharing data between agencies. While all of these practices are known today, the logistics of encrypting such large volumes of data may have implications.

The release of data into the public domain is further complicated by copyright law. Data generated by government agencies is not necessarily protected by copyright, as copyright does not extend to the protection of mere facts; it will be triggered only by a minimum degree of intellectual input (usually the original expression of an idea). However, where government data is covered by Crown copyright, restrictions on the re-use of that information may prevent access to that information at a cost which is reasonable for members of the general public. This may be driven by economic incentives, as agencies may want (or need) to charge for access to that information in order to maintain or increase revenue [3].

Overarching these more formal legal restrictions is an issue which is more difficult to quantify and address. Several government and independent reports have noted that Australian government agencies remain resistant to sharing their data or releasing it into the public domain [1]-[5]. This resistant culture may arise from a number of factors [2]-[4], including:

- A belief that secrecy is the default position regarding the disclosure of information, and that disclosure will breach the Privacy Act or other legislation;
- Fear of what might be revealed if the information is published, such as mistakes or misconduct on behalf of government employees;
- Concerns about the quality or accuracy of the information being released;
- Limited understanding of the benefits that can be gained from open data; and
- A lack of leadership to help drive a shift towards the greater release of data.

On its face, then, it appears a simple proposition that government agencies should release greater amounts of the data they generate to aid big data analysis and contribute to greater accountability and democratic participation. However, the barriers to releasing this information are varied and significant. Strong legal protections prevent the disclosure of many categories of data, and public service culture appears to favour secrecy as the default position.

In many cases, a cultural reluctance to releasing government data will be useful and appropriate, as it will mean that agencies in doubt about releasing their data will err on the side of protecting personal and other sensitive information. However, this culture may become problematic if information which could otherwise be released appropriately is withheld due to a lack of understanding of the law or of the public benefit that might be gained from making that data more widely available. Section IV considers how some of the barriers could be overcome whilst ensuring that existing protections for individual privacy are retained.

IV. Overcoming Barriers to Releasing Government Data

Many of the barriers identified above are appropriate to protect well-established categories of private and sensitive information. It would be irresponsible for governments to release intelligence information, health records, credit records, tax file numbers or other data, the release of which would endanger security, invade privacy or facilitate identity theft. Beyond this, important questions remain as to whether the Privacy Act strikes an appropriate balance between facilitating the release of open data and protecting personal information. The issue is not that the prohibitions on disclosing personal information under the Privacy Act are too restrictive – on the contrary, they are appropriately designed to prevent the release of information held by government agencies that would allow individuals to be identified. Rather, the difficulties lie in understanding which categories of data fall within these protections. Beyond obvious categories like names, addresses and telephone numbers, it is difficult for agencies to know whether a particular dataset qualifies as 'personal information' for the purposes of the Privacy Act. This is largely because the definition of personal information extends to information from which a person is 'reasonably identifiable' (Privacy Act, Section 6). As mentioned in Part III, this means that de-identified information could constitute personal data. There would need only to be a reasonable possibility that a person could take that data and identify an individual from it – either by reversing the de-identification techniques, or by combining it with other sources of information. This poses a major challenge to government agencies seeking to release their data in to the public domain. If a particular dataset could contribute to big data analysis, but there is a reasonable possibility that this data could somehow be used to identify an individual, then the agency will not be permitted to disclose that information absent express consent. This may be so even if the agency has removed any obvious identifying information from the dataset. Understandably, government agencies will be reluctant to release their data publically in these conditions. The key problem, then, lies in understanding and applying the Privacy Act to specific categories of data, beyond the obvious cases of names, addresses and other basic identifying information. Would de-identified statistical data on welfare payments, for example, constitute personal information for the purposes of the Privacy Act? On face value, it would not, as statistical data such as this would not include any obvious identifying information. However, if there was a reasonable possibility that somebody with sufficient skills to identify an individual – either by reinstating the identifying information into the dataset, or by combining that dataset with other datasets containing geographical and demographic information – then that statistical data *would* constitute personal information for the purposes of the Privacy Act. Certainly, there is significant uncertainty in such a case as to whether that information could be validly released. Government agencies are likely to err on the side of withholding that information before allowing the possibility that information about a person's welfare payments or other private details

might be exposed. This uncertainty appears to be the major reason behind the reluctance of Australian government agencies to release their data publically. In its final report, the Government 2.0 Taskforce concluded [2] that a 'risk-averse' attitude towards breaching privacy law – as opposed to the law itself – had prevented the release of government data as open data:

> Often something will not be released, not because it is clear that is in breach of some stipulation – for instance the *Privacy Act 1988* (Privacy Act) – but because someone thinks it just could be, and of course privacy regulation, like so many areas of regulation can be complex. So rules of thumb are needed for practitioners. They may not precisely reflect the details of that act, or of any other or other possible obstacles, but they may nevertheless lead to suppression of information, even if the technical details of the Privacy Act actually permit release.

It seems, then, that government data which *could* be validly released under the Privacy Act is being withheld from the general public out of a mistaken belief that releasing the information would breach privacy protections. What ends up happening then, is low value datasets become openly available while higher value datasets remained closed. This affects the underlying utility of the value to be derived from the use of open data for big data analytics.

After interviewing representatives from Australian government agencies, members of the Data2Decisions Cooperative Research Centre concluded [5] that it was these concerns about protecting individual privacy – and not a cultural preference for secrecy – which was preventing government data from being released:

> Concerns expressed by agencies were mostly linked to concerns about proper levels of protection for sensitive, confidential, privacy-related or other information types. Most concerns were thus linked to key risks and hence understandable, rather than being based on a purely cultural reluctance to share data or investigate such options, or consider new approaches.

The encouraging insight from these reports is that the major reasons for agencies withholding their data appear to be based in concerns for individual privacy (which should be maintained) and uncertainty surrounding legal protections for privacy (which can be improved). If the major obstacle to releasing government data was a public service culture, which favoured secrecy over transparency and innovation, this would be significantly more difficult to overcome. Instead, what appears to be a cultural preference for secrecy is actually a more practical problem grounded in uncertainty as to when Privacy Act protections apply. How, then, might this uncertainty surrounding the Privacy Act be addressed, to facilitate the release of government data in appropriate circumstances and on a larger scale? The most direct solution would be to amend the definition of personal information in the Privacy Act so that it does not extend to information from which an individual might reasonably be identified. This would provide government agencies with significantly greater scope to release their data into the public domain. Provided that an agency had removed any obvious identifying information (such as names and addresses) from a dataset, that data would not qualify as personal information and could be released. The problem with this solution is that it would be achieving the ideals of the open data movement by significantly curtailing existing privacy protections. Debates continue over the reliability of de-identification techniques [13]-[15] which suggests that the technology is not yet sufficiently advanced for governments to be fully confident that de-identified information will protect individual privacy [20]-[23]. The words 'reasonably identifiable' will serve a useful purpose in the definition of personal information until such time as de-identification techniques can provide this level of confidence.

A preferable approach would instead be to maintain the current definition of personal information in the Privacy Act, while generating greater clarity around how that legislation operates and greater recognition of how and why government data should be released as open data. Below we set out some suggestions for how this might be achieved.

First, significant research should be undertaken which itemises the specific datasets held by government agencies and whether or not these fall within the Privacy Act definition of personal information. This is a large and complex task, and will require cooperation between government departments and academia. Currently, only basic examples of personal information (such as names and addresses) are set out in guidance on the Privacy Act from the Australian Information Commissioner [16]. This provides little guidance to agencies as to whether the Privacy Act applies in more difficult cases. Agencies will receive legal advice on these questions, but this advice will be ad hoc and may differ between agencies (or even between the same agency at different times). These advices should be shared with researchers and consolidated to construct a clear and comprehensive database setting out which government datasets are subject to privacy protections.

Second, ongoing research is needed into which de-identification techniques are the most reliable when applied to government datasets, as well as clear guidance on how these techniques can be applied by government agencies. This will help government agencies to release de-identified data with the confidence that individual privacy will be protected.

Third, there needs to be greater recognition of the benefits that open data can provide. As mentioned in Part II, a greater commitment to releasing government data as open data could benefit the Australian economy by up to $64 billion each year [10]. Government data is regularly updated, highly accurate and collected over a long period of time, which means that it provides a fruitful source for research and innovation, including valuable longitudinal studies [17].

The growing number of reports and guidance on open data will help to generate this recognition amongst government agencies [1]-[5], but this should be supplemented with greater public discussion of the issues surrounding open data. A greater understanding of the open data movement among the

general public could help to influence government policy further in the direction of transparency and accountability.

Finally, clear guidance should be provided to government agencies on how they should collect and organise their datasets before releasing them into the public domain. Some useful guidance on this is already appearing, but this should be supplemented with more technical guidance that is tailored to specific datasets. For example, Australia's Gov 2.0 taskforce has recommended [2] that PSI should, by default, be: free, based on open standards, easily discoverable, understandable, machine-readable, freely reusable, and transformable. This means that government data should be released under a Creative Commons Licence where possible or otherwise with clear guidance on re-use rights [2], [18]. It also means that government agencies should consider the re-use of their data for big data analysis when collecting that data in the first instance.

The shift from a government culture which errs on the side of withholding information from the public, to one which generates new data with big data analysis and the open data movement in mind, could take a decade or more to fully achieve. The suggestions above set out some practical steps by which this cultural shift could begin to take effect. This task should begin now, lest Australia lag further behind other countries in reaping the full benefits that open data can provide.

REFERENCES

[1] PricewaterhouseCoopers, *Deciding with Data: How Data-Driven Innovation is Fuelling Australia's Economic Growth*. PwC, 2014.
[2] Government 2.0 Taskforce, *Engage: Getting on with Government 2.0 – Report of the Government 2.0 Taskforce*. Australian Government, 2009.
[3] Australian Government Information Management Office, *Australian Government Information Interoperability Framework: Sharing Information Across Boundaries*. AGIMO, 2006.
[4] Australian Government Information Management Office, *National Government Information Sharing Strategy: Unlocking Government Information Assets to Benefit the Broader Community*. Australian Government, 2009.
[5] 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.
[6] IBM, *Bringing Big Data to the Enterprise: What Is Big Data?* <http://www-01.ibm.com/software/au/data/bigdata/> (last accessed 8 October 2015).
[7] Department of Finance and Deregulation, *The Australian Public Service Big Data Strategy: Improved Understanding Through Enhanced Data-Analytics Capability*. Australian Government, 2013.
[8] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2013.
[9] Barcelona Inspires, *BCN Smart City* <http://smartcity.bcn.cat/en> (last accessed 8 October 2015).
[10] Lateral Economics, *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*. Lateral Economics and Omidyar Network, June 2014.
[11] Open Government Partnership, *What is the Open Government Partnership?* <http://www.opengovpartnership.org> (last accessed 8 October 2015).
[12] Australian Government, Information Security Management Guidelines. Australian Government, 2013.
[13] Jane Yakowitz, "Tragedy of the data commons" *Harv Journal Law & Tech*, vol 25, no. 1, pp. 1-67.
[14] Stephen Wilson, "You can de-identify but you can't hide', *Lockstep*, 25 April 2015 < http://lockstep.com.au/blog/2015/04/25/de-identification-risks> (last accessed 8 October 2015).
[15] Dan Barth-Jones, 'Does de-identification work or not', *FierceBigData*, 23 June 2014 < http://www.fiercebigdata.com/story/does-de-identification-work-or-not/2014-06-23> (last accessed 8 October 2015).
[16] Australian Information Commissioner, *What is covered by privacy?* <http://www.oaic.gov.au/privacy/what-is-covered-by-privacy> (last accessed 8 October 2015).
[17] Australian Government Productivity Commission, *Annual Report 2012-13*. Australian Government, 2013.
[18] Australian Information Commissioner, *Principles on Open Public Sector Information*. Australian Information Commissioner, 2011.
[19] Peter Mell, "NIST Presentation: Overview of Big Data and Security Implications" National Institute of Standards and Technology (2015) http://breakinggov.sites.breakingmedia.com/wp-content/uploads/sites/4/2012/11/bigdata.pdf.
[20] Alana Maurushat, Lyria Bennett-Moses and David Vaile, "Using 'Big' Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards," (June, 2015) Proceedings of the 15th International Conference on Artificial Intelligence and Law 196 http://dl.acm.org/ft_gateway.cfm?id=2746110&ftid=1596343&dwn=1&CFID=542654113&CFTOKEN=24738647 (last accessed October 10, 2015).
[21] Jane Bambauer, Krish Muralidhar, and Rathindra Sarathy, "Fool's Gold: an Illustrated Critique of Differential Privacy" (2014) 16 Vanderbilt Journal of Entertainment & Technology Law.
[22] Ed Felten, and Narayanan, "No Silver Bullet: De-identification Still doesn't Work" (July 9, 2014) http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf (last accessed October 9, 2015).
[23] Ann Cavoukian and Jonas, "Privacy by Design in the Age of Big Data" (June 8, 2012) http://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf (last accessed October 9, 2015).