Australian Government

Department of Finance and Deregulation

Big Data Strategy – Issues Paper

MARCH 2013

Contents

1. Introduction

The data held by Australian Government agencies is both a national and government asset. It is also a potential source of opportunity. In this context, Australian Government agencies, like many other organisations, are aware of the challenges and opportunities that big data represents to the way they develop policy and deliver services to citizens.

The purpose of this issues paper is to provide an opportunity to consider the range of opportunities presented to agencies in relation to the use of big data, and the emerging tools that allow us to better appreciate what it tells us, in the context of the potential concerns that this might raise.

As an example, one of the major challenges facing agencies here is to leverage the value of big data sets while ensuring they continue to protect the privacy rights of the Australian public.

The Australian Government is committed to protecting citizen's rights to privacy, and as part of that commitment, has recently strengthened the provisions of the Privacy Act. The Australian Government Information Management Office (AGIMO) acknowledges that big data, and its associated analytical tools, can provide a challenge to these rights, but believe that, with proper considerations, agencies will be able to use big data to develop better policies and deliver better services without compromising the privacy rights of the public.

Our aim is to ensure that the use of the new technology and tools supporting big data will deliver benefits while maintaining compliance with privacy. To this end AGIMO will be working closely with the Office of the Australian Information Commissioner (OAIC), the Attorney General's Department (AGD) and experts across the public and private sectors as it develops a big data strategy.

1.1 Where are we now?

Data is being produced at an ever increasing rate. This growth in data production is being driven by:

- individuals and their increased use of media;
- organisations;
- the switch from analogue to digital technologies; and
- the proliferation of internet connected devices and systems.

There has also been an acceleration in the proportion of machine-generated and unstructured data (photos , videos, social media feeds and so on) compared to structured data such that 80% or more of all data holdings are now unstructured and new approaches and technologies are required to access, link, manage and gain insight from these data sets.

The commonly accepted definition of big data comes from Gartner who define it as high-**volume**, high-**velocity** and/or high-**variety** information assets that demand

cost-effective, innovative forms of information processing for enhanced insight, decision making, and process optimization. These are known as the "three Vs". Some analysts also discuss big data in terms of value (the economic or political worth of data) and veracity (uncertainty introduced through data quality issues).

Government agencies hold or have access to an ever increasing wealth of data including spatial and location data, as well as data collected from and by citizens. Experience suggests that such data can be utilised in ways that have the potential to transform service design and delivery so that personalised and streamlined services, that accurately and specifically meet individual's needs, can be delivered to them in a timely manner.

Improved service delivery could cover areas as diverse as remote medical diagnostics, major infrastructure management, personalised social security benefits delivery, improved first responder and emergency services, reduction of fraudulent or criminal activity across both government and private sectors, and the development of innovative new services as the growth and availability of Public Sector Information (PSI) becomes more prevalent.

The private sector holds huge amounts of data about its customers and in many cases leads the way in how this data is analysed and used to create new business models and services. Agencies have the opportunity to learn from the innovations occurring in the private sector to operate more efficiently and deliver services more effectively while ensuring that privacy and security matters are carefully considered.

Private sector organisations such as Google, Twitter and Facebook hold enormous data stores on Australian citizens and people across the world, and offer access to these on commercial terms. While needing to carefully consider the veracity of this data, it may be that agencies could consider using this data as part of big data analytics projects. The ethical, privacy and security implications of decisions such as these will need to be carefully considered.

1.2   Why a big data strategy?

The development of a big data strategy was initiated by the _APS ICT Strategy 2012 – 2015_[1] (ICT Strategy) which highlighted the need for a strategy to enhance cross-agency data analytic capability for improved policy and service delivery.

The big data strategy is intended for agency senior executives and business program managers. It is designed to highlight key opportunities and challenges that big data will bring to government agencies. The strategy will aim to assist agencies to take advantage of these opportunities and realise the potential benefits of these new technologies. This strategy will also consider a coordinated approach, across government, for the adoption of big data analytics technologies.

The strategy will deliver on elements of the ICT Strategy. Big data allows for more focused and evidence-based policy design and service implementation that in turn allows citizens to interact with the government in a personalised and seamless way.

As awareness of the benefits of big data increases there is likely to be an increase in public debate regarding the balance of benefits versus the challenges associated with the technology. Government agencies need to be in a position to consider external expert opinion and enunciate their own position on the use of the technology.

2. Opportunities for Australian Government agencies

The opportunity that big data presents to government agencies is in the potential to unlock the value and insight contained in the data agencies already hold via the transformation of information, facts, relationships and indicators. The value of big data for agencies is limited by the their ability to effectively manage the volume, velocity and variety of big data and the ability to derive useful information from this data. With every opportunity there come challenges or barriers and agencies must overcome these to enable the benefits of big data to be realised.

Consideration of advances in big data technology has shown that it has potential to enhance the government's analysis capability in areas such citizen-centric service delivery. It is evident that big data also provides insights into social networks and relationships as well as allowing for the development of predictive models for a number of applications.

Of interest more broadly to agencies, big data analysis may provide profound insights into a number of key areas of society including health care, medical and other sciences, transport and infrastructure, education, communication, meteorology and social sciences.

Key areas that big data analytics may influence are detailed below:

Data management — there are potential savings in time and money if agencies implemented smarter data management practices that were conscious of the needs of big data analysis. Data sources from differing organisations and operational areas would be of greater benefit to multiple agencies and for multiple purposes if there were greater transparency. For example, through better business process management, redundant data collection processes can be reduced by reusing data collected from separate processes.

Personalisation of services — big data analytics may produce value by revealing a clear picture of an individual customer or customer group. Big data is able to achieve this due to its characteristic granularity (its ability to combine multiple and differing small pieces of data to generate revealing insights). This granularity may assist in unlocking the possibility of personalised services tailored to the individual and delivered by government.

Problem solving and predictive analytics — the unification of multiple datasets from disparate sources in combination with advanced analytics techniques and technologies will advance problem solving capabilities, and in turn will improve the ability of predictive analytics to reveal insights that can effectively support decision-making.

Productivity and efficiency — the analysis of big data sources can be used to identify cost savings and opportunities to increase efficiency, which will directly

contribute to an improvement in productivity. This can in turn help to encourage further innovation.

2.1   What the future looks like

A successful big data strategy is expected to assist in realising each of the priority areas observed in the ICT Strategy.

The delivery of better services — big data analytics will allow government agencies to deliver more personalised services that are tailored to meet citizen's needs and preferences. For example, the identification of individuals or groups who are eligible for certain entitlements without the need for them to be aware of or explicitly apply for that benefit.

Improved efficiency of government operations — more effective use of big data for predictive analysis will allow government agencies to better assess risk and feasibility, and detect fraud and error. This will in turn deliver improved productivity as resources can be directed towards projects with greater confidence of the outcome.

Open engagement — through the Big Data Working Group and other groups such as the Australian Tax Office's Data Analytics Centre of Excellence, agency stakeholders in big data and its related technologies will be able to engage with industry, academia, non-government organisations and other interested parties locally and internationally. These engagements will help to build knowledge, spark ideas, generate growth and better inform decisions and solutions that meet the needs of the government, both on a national and local level.

Further to this, the pursuit of big data technologies by government agencies will see a number of potential collaboration opportunities between agencies that will strengthen existing networks and help develop new partnerships.

3. Challenges

Meeting the challenges presented by big data will be difficult. The volume of data is already enormous and increasing every day. The velocity of its generation and growth is increasing, driven in part by the proliferation of internet connected devices. Furthermore, the variety of data being generated is also expanding, and organisation's capability to capture and process this data is limited.

Current technology, architecture, management and analysis approaches are unable to cope with the flood of data, and organisations will need to change the way they think about, plan, govern, manage, process and report on data to realise the potential of big data.

3.1 Privacy, security and trust

The Australian Government is committed to protecting the privacy rights of its citizens and has recently strengthened the *Privacy Act* (through the passing of the Privacy Amendment (Enhancing Privacy Protection) Bill 2012) to enhance the protection of and set clearer boundaries for usage of personal information.

Government agencies, when collecting or managing citizens data, are subject to a range of legislative controls, and must comply with the a number of acts and regulations such as the *Freedom of Information Act* (1982), the *Archives Act* (1983), the *Telecommunications Act* (1997) ,the *Electronic Transactions Act* (1999), and the *Intelligence Services Act* (2001). These legislative instruments are designed to maintain public confidence in the government as an effective and secure repository and steward of citizen information.

The use of big data by government agencies will not change this; rather it may add an additional layer of complexity in terms of managing information security risks. Big data sources, the transport and delivery systems within and across agencies, and the end points for this data will all become targets of interest for hackers, both local and international and will need to be protected.

The public release of large machine-readable data sets, as part of the open government policy, could potentially provide an opportunity for unfriendly state and non-state actors to glean sensitive information, or create a mosaic of exploitable information from apparently innocuous data. This threat will need to be understood and carefully managed.

The potential value of big data is a function of the number of relevant, disparate datasets that can be linked and analysed to reveal new patterns, trends and insights. Public trust in government agencies is required before citizens will be able to understand that such linking and analysis can take place while preserving the privacy rights of individuals.

The public trust in government agencies and systems needs to be maintained. As the volume of government data holdings increase, the trust that Australians have in these agencies and their ability to securely hold information of a personal nature can easily be affected by leakage of data or information into the public domain. Agencies need to be able to maintain the public's trust and will need to consider this issue at the forefront when developing secure systems for managing big data stores. Liaison with industry experts is an important first step in this process.

3.2 Data management and sharing

Accessible information is the lifeblood of a robust democracy and a productive economy.[2] Government agencies realise that for data to have any value it needs to be discoverable, accessible and usable, and the significance of these requirements only increases as the discussion turns towards big data.

Government agencies must achieve these requirements whilst still adhering to privacy laws. The processes surrounding the way data is collected, handled, utilised and managed by agencies will need to be aligned with all relevant legislative and regulatory instruments with a focus on making the data available for analysis in a lawful, controlled and meaningful way.

Data also needs to be accurate, complete and timely if it is to be used to support complex analysis and decision making. For these reasons, management and governance focus needs to be on making data open and available across government via standardised APIs, formats and metadata. Improved quality of data will produce tangible benefits in terms of business intelligence, decision making, sustainable cost-savings and productivity improvements.

The current trend towards open data and open government has seen a focus on making data sets available to the public, however these 'open' initiatives need to also put focus on making data open, available and standardised within and between agencies in such a way that allows inter-governmental agency use and collaboration to the extent made possible by the privacy laws.

There are opportunities for agencies to create incentives that promote competition and innovation within the marketplace. Various releases of public sector information (PSI) into public 'mash-up' sessions have already seen new and innovative uses for public data be developed.

In order to increase the use of government data across agencies, individual agencies need to be encouraged to establish and maintain publically available information asset registers. Information asset registers are lists of an agencies information assets intended to increase the discoverability, access and reusability of agency information assets by both internal and external users.

The OAIC, in the paper _Open Public Sector Information – from Principles to Practice_ [v], has highlighted that agencies have identified establishing and maintaining an information asset register as the largest challenge in ensuring more robust information management, and this has been highlighted as a priority action area.

The release of PSI also needs to be considered in the light of new technologies that can provide an opportunity for unfriendly state and non-state actors to glean sensitive information, or create exploitable information from seemingly innocuous data sets.

There are concerns over the ability for personally identifiable information to be extracted through the correlation of separate data-sets that have individually been anonymised, otherwise known as the 'mosaic effect'.[3]

AGIMO is seeking the input and advice of the OAIC and other key agencies such as the AGD and the Defence Signals Directorate (DSD) in regards to big data and privacy for inclusion in the strategy.

3.3 Technology and analytical systems

The emergence of big data and the potential to undertake complex analysis of very large data sets is, essentially, a consequence of recent advances in the technology that allow this. If big data analytics is to be adopted by agencies, a large amount of stress may be placed upon current ICT systems and solutions which presently carry the burden of processing, analysing and archiving data. Government agencies will need to manage these new requirements efficiently in order to deliver net benefits through the adoption of new technologies.

In particular these technologies include low cost storage arrays, in-memory processing, and cloud based storage and processing together with a range of new software developments underpinned by high performance processing platforms. The emergence of Cloud Computing over the last few years represents the single most important contributor here – with Cloud based storage, compute and analytical tools and apps now widely available. Cloud computing offers agencies the flexibility to store, and perform computational analysis on, increasingly large data sets in a manner previously not possible.

Big data and the cloud

Big data initiatives impose large technology overheads in order to meet processing, storage and transportation requirements. The size and scale of these resources mean they are likely to be expensive to procure and manage, and impractical for many agencies to operate in-house. The cloud environment is suited to this kind of initiative as it offers capacity, scalability, resilience, efficiency and availability. **Figure 1** demonstrates the efficiency that the use of cloud infrastructure can have on big data analysis, because capacity is available on demand.

Figure 1: Big data and cloud computing value graph
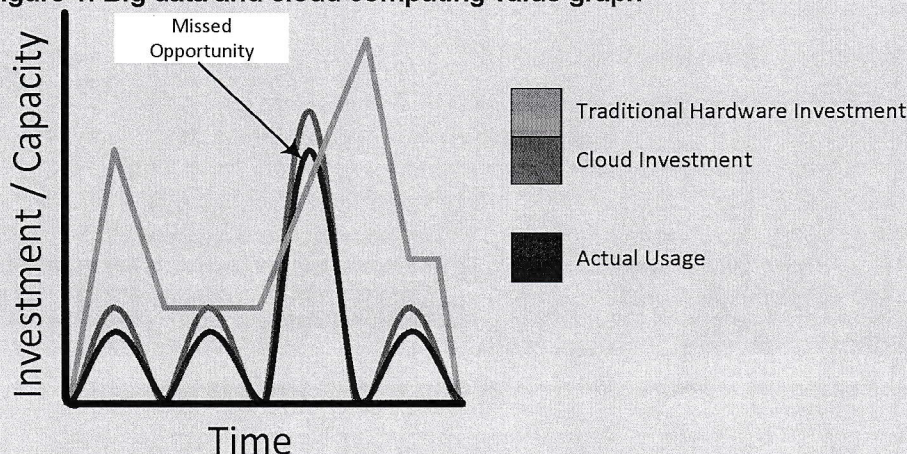
The National Broadband Network (NBN) will assist in providing the necessary bandwidth to transport the data and may help to enable data to be analysed in a cloud environment and in near real-time.

Agencies will need to determine the feasibility of any big data analytics projects based upon their ability to meet the business needs of the agency rather than any preconceived technological preference.

Vendors and developers of big data systems and solutions, including open source software, are developing more capable tools to simplify the challenges of big data analysis.

3.4 Skills

Due to its relative youth and complexity, big data will require agencies to attract employees with diverse new skill sets. These skills include science, technological, research, statistical, analytical and interpretive skills, business acumen and creativity — as well as an understanding of the underlying nature of the business process or policy intent. These skill sets are unlikely to be found in any one person, and this means that collaborative teams of specialists will need to be assembled to allow agencies to achieve optimal results from their data analysis efforts.

Many observers have noted that there is currently a major skills gap for data scientists with experience in big data analytics. According to Gartner[4], by 2015, big data demand will reach 4.4 million jobs globally, with two thirds of these positions remaining unfilled.

There is currently a shortage of university degrees that have a curriculum focused on big data analytics. Industry is looking for more qualified professionals with skills related to big data analytics. There is a push for education providers to develop courses that provide suitable education and training in this area of expertise.

Government agencies will look for opportunities to leverage support from the expertise and experience in big data that is found inside and outside of government. Agencies will also consider opportunities for partnerships with industry (including vendors and solutions providers) as well as independent research bodies and academia that will allow agencies to attract, retain and maintain expertise in this area much as it does with senior staff across other functional areas of government.

The Advanced Analytics Institute (AAI)

In Australia, the University of Technology Sydney has opened the Advanced Analytics Institute (AAI), a research institute that focuses on data and analytics science as well as evidence-driven decision making research. The AAI brings together researchers with a variety of backgrounds and aims to support and mentor generations of high-calibre analytics graduates. It has worked with a number of organisations and Australian Government departments including AMP, IBM, SAS, Microsoft Research, Nokia, Westpac, the ATO and DHS (Centrelink).

4. The Way Forward

To realise the outcomes of the *ICT Strategy* in relation to big data, Australian Government agencies need to achieve the following objectives:

- leveraging the big data experience and skills that exist within a small number of government agencies, and across the ICT industry and the research community to address the skills deficit that exists in this area. There are great opportunities to use this skill and experience to:
  - improve the management and analysis of the data sets that the government holds and
  - improve government operations, policy development and service delivery;
- ensuring that privacy issues are addressed up front in order to:
  - protect the privacy rights of the Australian public and
  - give agencies clarity around how they can use big data to develop better policies and services; and
- encouraging the release of public sector information (PSI) consistent with all privacy and security legislative instruments and guidance.

To succeed in these endeavours and finalise a big data strategy, agencies, facilitated by AGIMO, will engage in the following supporting activities:

- support the work of the cross-agency big data working group through the development of a detailed work plan for the creation of relevant policy, guidance and advice;
- work in collaboration with the Whole-of-Government Data Analytics Centre of Excellence to provide assistance in overcoming skills, analysis and technological challenges;
- develop specific guidance in response to:
  - the OAIC's recommendations for agencies in creating information asset registers; and
  - privacy, security and legislative implications of big data across government.

5. Endnotes

[1] Department of Finance and Deregulation, Australian Public Service Information and Communications Technology Strategy 2012-2015, http://agimo.gov.au/ict_strategy_2012_2015/

[2] Office of the Australian Information Commissioner, Open public sector information: from principles to practice, http://www.oaic.gov.au/publications/reports/Open_public_sector_information_from_principles_to_practice_February2013.pdf

[3] Jaikumar Vijayan, Sidebar: The Mosaic Effect, http://www.computerworld.com/s/article/91109/Sidebar_The_Mosaic_Effect

[4] Gartner, Gartner Reveals Top Predictions for IT Organisations and Users for 2013 and Beyond, http://www.gartner.com/it/page.jsp?id=2211115