![Australian Government — Department of Finance and Deregulation — Australian Government Information Management Office]
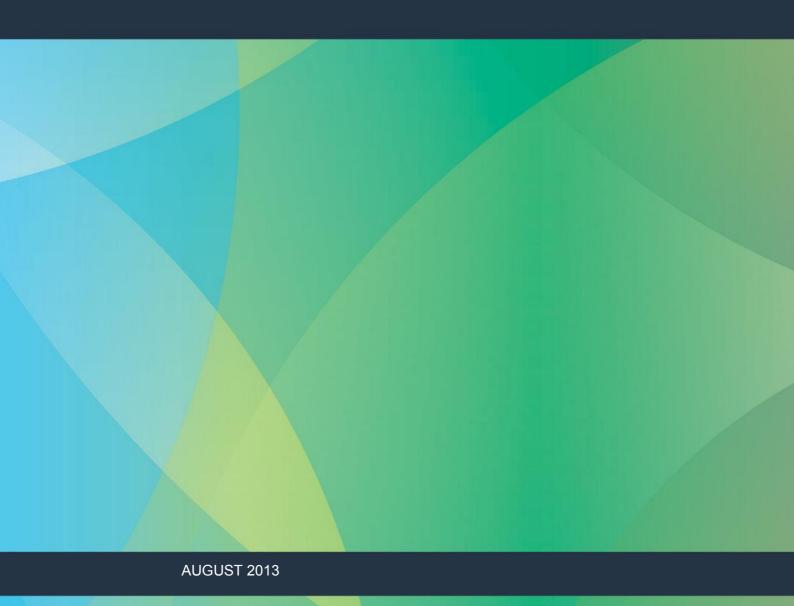
The Australian Public Service Big Data Strategy

Improved understanding through enhanced data-analytics capability

AUGUST 2013

Contents

This page is intentionally left blank

Preface

> *"The value of big data lies in our ability to extract insights and make better decisions"*[1]

The data held by Australian Government agencies has long been recognised as a government and national asset. The potential growth in this data due to the adoption of new technologies as well as the production of an increasing amount of both structured and unstructured data outside of government, suggest that big data analytics can increase the value of this asset to government and the Australian people.

Government policy development and service delivery will benefit from the effective and judicious use of big data analytics. 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 government operations - from the maintenance of our national infrastructure, through the enhanced delivery of health services, to reduced response times for emergency personnel.

The Big Data Strategy is intended for Australian Government agency senior executives with responsibility for delivering services and developing policy[2]. It outlines future work by the Government that will assist agencies to make better use of their data assets whilst ensuring that the Government continues to protect the privacy rights of individuals.

The development of a Big Data Strategy was identified in the *APS ICT Strategy 2012-2015*[3] which highlighted the need for a whole-of-government approach to big data to enhance data analytic capability of agencies in support of improved service delivery and the development of better policies. The Australian Government Information Management Office (AGIMO), within the Department of Finance and Deregulation, leads this work.

This Big Data Strategy was developed with the assistance of a multi-agency working group that was established in February 2013 (the Big Data Working Group).

---

[1] Dr. Michael Rappa Director of the Institute for Advanced Analytics and Distinguished University Professor North Carolina State University, http://analytics.ncsu.edu/?p=4770

[2] This Strategy does not aim to address the use of big data analytics by the intelligence and law enforcement communities.

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

On 15 March 2013, AGIMO released the *Big Data Strategy — Issues Paper*[4] on the AGIMO blog. The intention of the paper was to start the conversation about big data with the public, industry and academia. The input received from this public consultation process has been used to inform the development of the Big Data Strategy.

In parallel to the development of this Strategy, a whole-of-government Data Analytics Centre of Excellence (DACoE) has been launched and is led by the Australian Taxation Office.

The DACoE will build analytics capability across government by establishing a common capability framework for analytics, sharing technical knowledge, skills and tools, and building collaborative arrangements with tertiary institutions to shape the development of analytics professionals. The DACoE will work within guidelines established via the Big Data Strategy and other work of the Big Data Working Group.

---

[4] Department of Finance and Deregulation. Big Data Strategy — Issues Paper,
http://agimo.gov.au/2013/03/15/released-big-data-strategy-issues-paper/

Introduction

90% of the data in the world today has been created in the last two years alone. Some estimate that data production will be 44 times greater in 2020 than it was in 2009. Others estimate an additional 2.5 quintillion bytes of data is being generated every day.[5][6]

Big data and the accompanying tools and techniques that help to analyse and make sense of it, clearly present opportunities and challenges to government agencies. Indeed, big data, as an emerging concept, intersects with many data management issues that pre-date this concept. Big data again raises the issues of the value of data held by Commonwealth agencies and the responsibility to realise this value to benefit the Australian public, as well as the need to negotiate the privacy risks of linking, sharing and providing broader access to data.

This strategy aims to provide a pathway for Australian Government agencies to follow as they consider the use of big data to support their operations. This strategy aims to assist agencies in achieving productivity gains, through better service delivery and policy development while ensuring the privacy of individuals remains protected.

Before setting out the opportunities big data provides to government, and a vision, principles and actions aimed to assist agencies realise these opportunities, it is worth clarifying three issues:

- what we mean by big data;
- how this intersects with the concept of data as an asset and open government; and
- what the potential privacy implications are of this intersection.

---

[5] IBM, Big Data at the Speed of Business, http://www-01.ibm.com/software/data/bigdata/

[6] CSC, Big Data Universe Beginning to Explode, http://www.csc.com/insights/flxwd/78931-big_data_growth_just_beginning_to_explode

What is big data?

Big data refers to the vast amount of data that is now being generated and captured in a variety of formats and from a number of disparate sources.

Gartner's widely accepted definition describes big data 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."[7] These are known as the "three Vs". In addition to the "three Vs", there are two further "Vs" which are commonly discussed and which add further dimensions to the definition: the *veracity* or reliability of the data, and the *volatility* or sensitivity of data.

For the purposes of this strategy, big data analytics means that:

1. The data analysis being undertaken uses a high volume of data from a variety of sources including structured, semi-structured, unstructured or even incomplete data; and

2. The size (volume) of the data sets within the data analysis and velocity with which they need to be analysed has outpaced the current abilities of standard business intelligence tools and methods of analysis.

Big data exists in both structured and unstructured forms, including data generated by machines such as sensors, machine logs, mobile devices, GPS signals, as well as transactional records. According to IBM[8] some of the most common types of data being used in big data analytics include internal transactional, log and event data.

Access to advanced computing power for the analysis of large quantities of data is now more readily available. The emergence of powerful and cost-effective analytical tools, storage, and processing capacity (including cloud computing) removes the cost barriers to big data analytics and means that Australian Government agencies and other sectors of the economy can potentially realise benefits from the use of this technology more easily.[9]

Big data analytics offers new tools and techniques to address these data challenges. These tools and techniques allow us to delve deeper into data, potentially delivering important insights. In the case of government, important insights about the effectiveness and efficiency of services and policies may be realised through this analysis. Furthermore, the tools allow this analysis to occur more quickly and, generally, the value of the output from a data analytics project increases as the analysis approaches real-time.

---

[7]Gartner, The Importance of 'Big Data': A Definition,

 http://www.gartner.com/id=2057415

[8] IBM, Analytics: The real-world use of big data,

http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html

[9] McKinsey Global Institute, Big Data: The next frontier for innovation, competition, and productivity, May 2011

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Data as an asset

The government produces data as a result of its administrative and policy development activities and interactions with the Australian public. It is part of the government's responsibility to realise the value of this data and the information contained within it, and to recognise this data as a national asset to both the Government and the Australian public.[10]

Increasingly, it is being recognised that governments cannot realise this value without the assistance of industry and academia. As a result Australia, along with many other advanced economies, is increasingly seeking to provide data to third parties for analysis, or to support the provision of services, for example through the development of mobile apps.

The Declaration of Open Government[11], Australia's announcement in regards to joining the Open Government Partnership[12], the establishment of data.gov.au and the publication of the Principles on Open Public Sector Information[13] (PSI) were all important steps in opening up the data held by Commonwealth agencies data for reuse.

> "The Principles on open public sector information form part of a core vision for government information management in Australia. They rest on the democratic premise that public sector information is a national resource that should be available for community access and use.
>
> Transparency and public access to government information are important in their own right and can bolster democratic government. Information sharing better enables the community to contribute to policy formulation, assist government regulation, participate in program administration, provide evidence to support decision making and evaluate service delivery performance. A free flow of information between government, business and the community can also stimulate innovation to the economic and social advantage of the nation."
>
> **Office of the Australian Information Commissioner,**
> **Principles on open public sector information**

---

[10] Freedom of Information Act 1982, Part I – Preliminary, Clause 3 Objects – general, http://www.comlaw.gov.au/Details/C2013C00242/Html/Text#_Toc358042501

[11] Department of Finance and Deregulation, Declaration of Open Government, http://agimo.gov.au/2010/07/16/declaration-of-open-government/

[12] Attorney-General's Department, Australia joins Open Government Partnership, http://www.attorneygeneral.gov.au/Mediareleases/Pages/2013/Second%20quarter/22May2013-AustraliajoinsOpenGovernmentPartnership.aspx

[13] Office of the Australian Information Commissioner, Principles on open public sector information, http://www.oaic.gov.au/publications/agency_resources/principles_on_psi_short.html

Privacy

The Australian Government is committed to protecting the privacy rights of individuals. Big data raises new challenges in respect to the privacy and security of data.

The data management policies of government agencies will always be guided by the applicable legislative controls that already regulate government's use and release of data sets and information. Agencies will continue to need to comply with these controls for any big data related activities.

Maintaining the public's trust in the government's ability to ensure the privacy and security of the data stores that it has control over is paramount. This Strategy aims to ensure that this issue remains at the forefront of agency deliberations when they consider developing business cases for big data projects.

A number of specific issues around privacy need to be managed if agencies are to realise the benefits that big data can potentially provide, including:

- better practice in linking together cross-agency data sets;
- better practice use of third party data sets[14];
- de-identification[15] and the "mosaic effect"[16];
- the necessary considerations to make before releasing open data; and
- data retention and cross-border flows.

Forthcoming guidance will be produced to increase agencies understanding of these issues and provide advice and best practice for addressing these. The use of big data, like any other form of data or information, is subject to a number of legislative controls including the *Freedom of Information Act 1982*[17], the *Archives Act 1983*[18], the *Telecommunications Act 1997*[19] and the *Electronic Transactions Act 1999*[20]. Agencies also need to comply with the *Data-matching Program (Assistance and Tax) Act 1990*[21] wherever Tax File Numbers are used.

The use of big data is also regulated by the *Privacy Act 1988*[22] [23] which regulates the handling of personal information throughout the information lifecycle, including collection, storage and security, use, disclosure, and destruction.

---

[14] Non-government data includes open data created by external organisations or scientific/research data that is shared through specific arrangements.

[15] De-identification is a process by which a collection of data or information (for example, a dataset) is altered to remove or obscure personal identifiers and personal information (that is, information that would allow the identification of individuals who are the source or subject of the data or information).

[16] The concept whereby data elements that in isolation appear anonymous can amount to a privacy breach when combined.

[17] Freedom of Information Act 1982, http://www.comlaw.gov.au/Series/C2004A02562

[18] Archives Act 1983, http://www.comlaw.gov.au/Series/C2004A02796

[19] Telecommunications Act 1997, http://www.comlaw.gov.au/Series/C2004A05145

[20] Electronic Transactions Act 1999, http://www.comlaw.gov.au/Series/C2004A00553

[21] Data-matching Program (Assistance and Tax) Act 1990, http://www.comlaw.gov.au/Series/C2004A04095

[22] Privacy Act 1988, http://www.comlaw.gov.au/Series/C2004A03712

[23] Note: The Privacy Amendment (Enhancing National Privacy Protection) Act 2012 (http://www.comlaw.gov.au/Series/C2012A00197) was passed by Parliament on 29 November 2012, and includes significant changes to the Privacy Act 1988, including the replacement of the Information Privacy Principles and the

Security

The use of big data by government agencies does not change the existing security concerns that apply to data and information in general; rather it may add an additional layer of complexity in terms of managing these risks. Big data sources, aggregated data sets, the transport and delivery systems within and across agencies, and the end points for this data will become potential areas where data is exposed to being compromised, and will need to be protected through the use of appropriate security controls. This threat will need to be understood and carefully managed within the existing legislative and policy parameters, as outlined in the Protective Security Policy Framework's guidelines on the storage of aggregated information[24].

Data management

Commonwealth data, including data used by big data analytics projects, needs to be authentic, accurate and reliable if it is to be used to support accountability and decision making in agencies. The creation, collection, management, use and disposal of agency data is governed by a number of legislative and regulatory requirements.

Government data needs to be managed in a way that ensures it is discoverable, accessible and useable. An approach to achieving this is through the National Archives of Australia's *Digital Continuity Plan[25]*. Digital continuity ensures information is complete, available and usable by those with a need for it. It also ensures that information and data is not kept for longer than needed.

Australian Government data is considered a Commonwealth record and therefore subject to the *Archives Act 1983[26]*, which provides the legal authority for agencies to dispose of business information and data. Over-retention of data is common in agencies as it is often cheaper to store data rather than undertake the resource-intensive exercise to appropriately dispose of it. Over-retention of data can leave agencies open to a range of business risks including security and privacy breaches, reputational risks, and excess costs.

Consideration also needs to be given to any new data that may be created out of the use and re-use of existing data. This new data may have new or differing retention requirements and may also need to be managed as a Commonwealth record.

The management of data and information in line with the *Archives Act 1983* is fundamental for supporting other compliance requirements such as the *Freedom of Information Act 1983[27]* and the *Evidence Act 1995[28]*.

National Privacy Principles with the Australian Privacy Principles. These changes do not come into force until March 2014.

[24]Attorney-General's Department, Information Security management guidelines – Management of aggregated information, http://www.protectivesecurity.gov.au/informationsecurity/Documents/PSPF%20-%20ISMG%20-%20Management%20of%20aggregated%20information.pdf

[25] National Archives of Australia, Digital Continuity Plan http://www.naa.gov.au/records-management/agency/digital/digital-continuity/plan/index.aspx

[26] Archives Act 1983, http://www.comlaw.gov.au/Series/C2004A02796

[27] Freedom of Information Act 1982, http://www.comlaw.gov.au/Series/C2004A02562

[28] Evidence Act 1995, http://www.comlaw.gov.au/Series/C2004A04858

In addition to the legislation and regulation above and in previous sections, there are a number of other legislative and regulatory requirements government agencies must comply and align with when it comes to the management of Commonwealth data. This includes the *Electronic Transactions Act 1999*[29], the *Financial Management & Accountability Act 1997*[30], the *Intelligence Services Act 2001*[31], the *Crimes Act 1914*[32], the *Australian Government Protective Security Policy Framework*[33], the *Australian Government Information Security Manual*[34] and the *Australian Public Service Commissioner's Directions 2013*[35].

[29] Electronic Transactions Act 1999, http://www.comlaw.gov.au/Series/C2004A00553

[30] Financial Management and Accountability Act 1997, http://www.comlaw.gov.au/Details/C2013C00282/Html/Text#_Toc359413476

Note: The Public Governance, Performance and Accountability Act 2013, (http://www.comlaw.gov.au/Series/C2013A00123) was passed by Parliament on 28 June 2013 and will replace the Financial Management and Accountability Act 1997 and the Commonwealth Authorities and Companies Act 1997 from 1 July 2014.

[31] Intelligence Services Act 2001, http://www.comlaw.gov.au/Series/C2004A00928

[32] Crimes Act 1914, http://www.comlaw.gov.au/Series/C1914A00012

[33] The Australian Government Protective Security Policy Framework, http://www.protectivesecurity.gov.au

[34] The Australian Government Information Security Manual, http://www.dsd.gov.au/infosec/ism/

[35] Australian Public Service Commissioner's Directions 2013, http://www.comlaw.gov.au/Series/F2013L00448

Opportunities and benefits

Government agencies hold, or have access to, an increasing amount of data that is available in structured, semi-structured and unstructured formats. Australian Government agencies alone have installed an additional 93,000 terabytes of storage during the period 2008-2012[36] to cope with increasing data production. The analytical opportunities offered by this data have long been recognised and the growth of big data analytics technologies has expanded these opportunities.

Big data offers organisations widespread potential opportunities and benefits. While the magnitude and nature of the value varies depending on industry sector, it is anticipated that government will be able to realise substantial productivity and innovation gains from the use of big data.[37]

It is expected that some big data projects will provide unanticipated insights into business problems. This is due, in part, to the process that big data analysis follows. While traditional scientific enquiry begins with a hypothesis that is tested through data collection, big data analysis is able to work in the opposite direction: beginning with a large amount of existing data which ,through analysis, reveals insights from which conclusions can be drawn. However, outcome oriented approaches to big data projects will remain critically important.

Industry experience suggests that these unanticipated correlations and discoveries may provide important insights that could lead to innovative solutions that might not otherwise have been reached. These insights may also provide opportunities to act and respond more rapidly to information and trends as they occur.

CSIRO — Vizie

The Commonwealth Scientific and Industrial Research Organisation (CSIRO), has developed a social media monitoring tool called Vizie which is a software tool that searches for and analyses keywords from publicly-available social-media data and metadata. It is designed to support individual engagement, not just provide summaries of social media activity.

Vizie uses two techniques to effectively analyse and provide insights about this data. Firstly, by building a baseline of normal discussions, Vizie is able to quickly alert staff to any changes that may require attention. Secondly by amalgamating online content together and grouping it according to discussion topics or issues, Vizie is able to present this content in a clear, structured manner.

A unique feature of Vizie is its visualisation component which is able to represent the retrieved posts in a simple overview that allows for the isolation of the original post in order to discover its context.

Vizie allows staff to prioritise posts, identifying which posts most need follow up comments to prevent the spread of misinformation. Prior to using Vizie, the Digital Media team at the Department of Human Services spent time conducting manual searches or using multiple off-the-shelf monitoring tools to track customer feedback on social media.

---

[36] Department of Finance and Deregulation. Australian Government ICT Expenditure Report 2008-09 to 2011-12, http://agimo.gov.au/files/2012/04/Australian-Government-ICT-Expenditure-Report-2008-09-to-2011-12.pdf

[37] McKinsey Global Institute, Big Data: The next frontier for innovation, competition, and productivity, May 2011 http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Service delivery

It is important to recognise the potential for agencies to utilise big data analysis in innovative ways that take advantage of patterns and correlations to improve service provision and outcomes. These insights can help to increase productivity and effectiveness by assisting agencies in better tailoring and targeting services, policies and programs. The improved targeting of services (in accordance with the Privacy Act and other relevant legislation) will help agencies to better manage and prevent over-servicing whilst ensuring that people are not missing out on any services to which they may be entitled and of which they might be unaware.

Improved service delivery could cover areas as diverse as the timely delivery of appropriate health and welfare services, major infrastructure management, personalised social security benefits delivery, improved emergency services and the reduction of fraudulent or criminal activity and errors across both government and private sectors. It may also result from the development of innovative new services as more PSI continues to be made available for reuse by third parties.

With big data analytics providing greater evidence to decision makers, better decisions can be made about how to tailor services reflecting specific individual and community needs and interests. This process of personalising services to the consumer's needs will allow for simpler and easier access to government services and may help government in reducing the costs of delivering these services by avoiding over-servicing or better matching of services to people and communities.

Personalising services will also improve the experience of individuals. For example the UK Cabinet Office Behavioural Insights team[38] applies insights from academic

DIAC — Border Risk Identification System

The Department of Immigration and Citizenship (DIAC) has developed a risk tiering system, the Border Risk Identification System (BRIS), for Australia's international airports to improve their ability to identify potential problem travellers in real time.

Australia's airports are receiving an average of 40 000 inbound travellers per day and this volume is increasing by some 5% annually. BRIS is able to assist DIAC in maintaining immigration and border integrity under this increasing pressure by providing a detailed and evidence based view of risk at the border so that resources can be allocated more effectively.

BRIS provides a smarter way to allocate border referral resources, by aligning resources to potential risk. Better targeted referrals means less inconvenience for genuine travellers, more efficient use of the departments time and resources, the avoidance of costs associated with onshore non-compliance and the ability to handle increasing caseloads without extra resources.

A successful prototype of the system was deployed in Sydney, Melbourne and Brisbane airports in early 2011. The system had halved the number of travellers undergoing additional checks at airport immigration points whilst detecting an increased number of suspicious travellers, many of which were eventually refused entry to Australia. The effectiveness of the system in turn saved tax payer dollars at an average of $60,000 saved per refusal.

In using advanced analytics, DIAC has substantially enhanced its ability to accurately identify risk while also reducing the need for delaying incoming travellers. The analytics-based system complements existing border risk identification and mitigation tools such as immigration intelligence, primary line referrals and Movement Alert List matches.

---

[38] Cabinet Office, Behavioural Insights Team,
 https://www.gov.uk/government/organisations/behavioural-insights-team

research in behavioural economics and psychology to public policy and services. The team uses a variety of data and statistical techniques to help find innovative ways of encouraging, enabling and supporting people to make better choices for themselves.

Policy development

It has been noted that *"the success of evidence-based policymaking depends on the quality of the evidence that underlies it"*[39]. There is inherent value in increasing the variety of data that is used and analysed in the evidence seeking process. By allowing government to access and perform analysis on rich layers of information from different sources, better government policy and better policy outcomes can be produced.

The final report from the APS200 Project[40] noted that knowledge management, integration and sharing within and across the APS and science agencies can facilitate access to and use of data and research services to support policy.

Developmental Pathways Project

The Telethon Institute for Child Health Research Developmental Pathways Project is a landmark project taking a multidisciplinary and holistic approach to investigate the pathways to health and wellbeing, education, disability, child abuse and neglect, and juvenile delinquency outcomes among Western Australian children and youth.

Researchers from the Telethon Institute and the University of Western Australia have been working in collaboration with a number of state government departments. This collaboration has helped to establish a process for linking together de-identified longitudinal, population-based data.

Through the linking of these data collections the Telethon Institute for Child Health Research have been able to:

• Ascertain whether changes in factors at the child, family and community level increase or reduces vulnerability to adverse outcomes in mental and physical health, education, child maltreatment, juvenile offending, in all Western Australian children;

• Identify areas of prevention and intervention across multiple government sectors, particularly in regard to mental health, disabilities, child protection, juvenile justice, educational achievement and school attendance;

• Use this data to evaluate existing government initiatives and determine, at a population level, how initiatives have impacted on educational, social and health outcomes;

• Improve the collection, utilisation and reliability of Government department data in program evaluation and policy development; and

• Respond to the government departments' agendas and policy frameworks, while enhancing whole of government initiatives.

Through the effective communication of the research findings, future government agency policies, practice and planning initiatives will be more preventative, culturally appropriate and cost efficient. These findings have encouraged cross-agency collaboration to ensure improved health, well-being and development of children and youth, their families and their communities.

---

[39] Smith, Jeffrey and Sweetman, Arthur (2009) "Putting the evidence in evidence-based policy" Productivity Commission Roundtable *Strengthening Evidence-Based Policy in the Australian Federation* http://www.pc.gov.au/__data/assets/pdf_file/0018/96210/05-chapter4.pdf

[40] Innovation, APS200 Project: The Place of Science in Policy Development in the Public Service, http://www.innovation.gov.au/Science/Pages/APS200ProjectScienceinPolicy.aspx

By being able to better analyse big data, decision makers may be able to model different policy options and more accurately predict the outcomes of policies before they are implemented and use this information to inform and improve the policy development process.

Agencies could then use this granular information to make better informed and more responsive decisions, to achieve desired outcomes in a shorter amount of time, and at lower cost to the community.

A testament to the power of big data analytics in respect to decision making is its ability to provide near "real-time" insights from the in-stream analysis of data. This is often described as "nowcasting".

Hal Varian, chief economist at Google shares his thoughts on the future of big data and nowcasting:

> "I'm a big believer in nowcasting. Nearly every large company has a real-time data warehouse and has more timely data on the economy than our government agencies. In the next decade we will see a public/private partnership that allows the government to take advantage of some of these private-sector data stores. This is likely to lead to a better informed, more pro-active fiscal and monetary policy."[41]

The real time analysis of big data may also provide clues about other policy areas including public health, social services and environmental threats, especially where an urgent response is required.

Providing decision makers with the most up-to-date information on a subject may also prove to provide direction for future policy initiatives, enabling proactive policy responses to emerging issues.

These predictive capabilities may assist government in better managing threats such as natural disasters before they occur.

Statistics

Big data may also make an important contribution to the usage of statistics as a means of informing the Government and the public on economic, societal and environmental issues.

Traditionally, official statistics have been based almost exclusively on the administrative data collected by government programs and survey data collections.

While these data sources will continue to be important, when combined with properly integrated and analysed data from semi and unstructured sources, more relevant, insightful and timely statistics will become available to the government which should inform better policy and service delivery decisions.

Business and economic opportunities

Many industry proponents have identified the practical business opportunities that big data analysis presents including the optimisation of operations, the delivery of better, more informed decision making tools, the management and mitigation of

---

[41]Anderson J.Q, Rainie L, Big Data: *Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well*, Pew Research Center, July 2012, http://www.greenplum.com/sites/default/files/PIP_Future_of_Internet_2012_Big_Data.pdf

financial and other risks, and the development of new business models all of which will lead to an increase in productivity and innovation. Developments in big data analysis are also creating opportunities for entire new industries.

Industry experts[42] have highlighted that the commercialisation of the research and development into big data that is coming out of Australian scientific and academic institutions need to be recognised. So too will the early adopter or first mover advantages that Australian businesses may derive from innovating with big data.

The Department of Broadband, Communications and the Digital Economy (DBCDE) have recently released the Update to the National Digital Economy Strategy[43] which outlines a number of initiatives that aim to ensure Australia's place as a leading digital economy. The strategy recognises the importance of big data and open data as facilitators of innovation and increased productivity.

This Strategy also recognises and supports the work that is occurring in the open data space at the state government level. For example the recently released draft NSW Government Open Data Policy[44] that supports government transparency, accountability and efficiency.

Skills

Big data analytics has the potential to create new ICT jobs and even new professions. Many observers have noted that there is currently a major skills gap for data scientists with experience in big data analytics. According to Gartner, by 2015, big data demand will reach 4.4 million jobs globally, with two thirds of these positions remaining unfilled.[45]

The Government has recognised the ICT workforce challenge. The update to the National Digital Economy Strategy outlines initiatives for the completion of the development of a new curriculum for technologies, and the promotion of careers in ICT to school students.

Other initiatives such as GovHack[46] further promote and support the development of skills and interest in data mashups, apps and visualisations, all of which are central to big data analysis. GovHack is a 48 hour, competitive event that encourages teams to find new ways to produce innovative solutions with open data.

The industry, research and academic sectors have been working on big data analytics projects for some time and continue to invest heavily in the skills, technologies and techniques involved with big data analysis. These sectors are also identified as being key custodians of valuable data collections, and potential

---

[42] Australian Information Industry Association, AGIMO Big Data Strategy Issues paper – AIIA Response, http://c.ymcdn.com/sites/www.aiia.com.au/resource/collection/F22C309D-816A-467D-9A09-473D98C7F3A2/AGIMO_Big_Data_Strategy_Paper.pdf

[43] Department of Broadband, Communications and the Digital Economy, Advancing Australia as a Digital Economy: An Update to the National Digital Economy Strategy, http://www.nbn.gov.au/nbn-benefits/national-digital-economy-strategy/

[44] New South Wales Government, Draft NSW Government Open Data Policy, http://engage.haveyoursay.nsw.gov.au/opendata

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

[46] http://www.govhack.org/

partners with the Government, for the delivery of insights from big data analytics which promote the public good.

Government will work with these sectors by leveraging and sharing expertise in big data analytics and related fields, and will also work with these sectors to promote the continued development of skills in this area of increasing demand.

The government is also strengthening the skills across agencies through initiatives such as the Data Analytics Centre of Excellence. This purpose of this initiative is to bring together representatives from across government and from a multitude of disciplines to share technical knowledge, skills and tools whilst building analytics capability.

The government is also looking to increase learning through the identification and initiation of a number of big data pilot projects. These pilot projects will help to showcase the potential for big data analytics to improve the way government operates and deliver tangible value to individuals.

Productivity benefits

The efficient use of big data analytics in the public sector has been identified as a potential driver of productivity gains in international jurisdictions.

For example, the McKinsey Global Institute[47] has estimated that Europe's public sector could potentially reduce the costs of administrative activities by 15 to 20 per cent.

According to these estimates, savings would equate to approximately €150 billion to €300 billion. McKinsey has also identified annual productivity growth by up to 0.5 percentage points over the next 10 years as a result of improved services and government efficiency.

The UK Policy Exchange organisation estimate[48] that greater productivity can be achieved through the use of big data tools in reducing fraud and error and closing the `Tax Gap' (the difference between actual tax collected and theoretical liabilities). These reductions would lead to savings of £16 to £33 billion per year.

---

[47] McKinsey Global Institute, Big Data: The next frontier for innovation, competition, and productivity, May 2011
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
[48] Policy Exchange, The Big Data Opportunity,
http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf

The vision —
what the future will look like

> **Vision**
>
> *The Australian Government will use big data analytics to enhance services, deliver new services and provide better policy advice, while incorporating best practice privacy protections and leveraging existing ICT investments.*
>
> *The Australian Government will be a world leader in the use of big data analytics to drive efficiency, collaboration and innovation in the public sector.*

The vision supports the following capabilities:

Enhanced services

- provide better information about service delivery outcomes and inform future models for the provision of these services as well as identifying where gaps exist under current service delivery arrangements;
- allow government agencies to better target services to those that need them, thus allowing more efficient and effective delivery of services; and
- enable agencies to improve services by tailoring service delivery based on the individual needs of businesses and communities.

New services and business partnership opportunities

- the analysis of big data is expected to lead to the development of new services based on insights derived from the analytics process; and
- industry developments and the maturity of tools and services that utilise big data analytics will create entirely new business opportunities and industries based on using open government data.

Improved policy development

- support better policy development by strengthening evidence-based decision making and provide more immediate information about policy settings and their impacts.

Protection of privacy

- incorporate "privacy by design" into big data analytics projects, and proactively ensure the privacy of the individual's data and information; and
- adopt better practice methodologies that address the potential risk to privacy posed by big data analytics and "the mosaic effect".

Leveraging the Government's investments in ICT technologies

- leverage the Government's investments in technology such as the National Broadband Network; and
- help to lower the cost of entry for big data analytics projects through use of the National Broadband Network in conjunction with cloud technologies.[49]

---

[49] Department of Finance and Deregulation, Australian Government Cloud Computing Policy, http://agict.gov.au/policy-guides-procurement/cloud

Big data principles

The following principles are intended to guide agencies in their approach to big data.

Principle 1: *Data is a National asset*

Data sets that government holds are a national asset and should be used for public good.

✓ Sharing this data, in accordance with the Declaration of Open Government, and other legislative requirements, will enhance the culture of engagement.

Principle 2: *Privacy by design*

Big data projects will incorporate 'privacy by design'.

✓ This means that privacy and data protection is considered throughout the entire life cycle of a big data project.

✓ All data sharing will conform to the relevant legislative and business requirements.

Principle 3: *Data integrity and the transparency of processes*

✓ Agencies embarking on the use of big data analytics to deliver improved service delivery are encouraged to seek peer review and conduct public consultation on these projects where appropriate.

✓ Agencies are encouraged to conduct Privacy Impact Assessments (PIA) for any new big data projects and publish these PIAs (or modified versions if necessary)[50].

✓ Each party to a big data analytics project must be aware of, and abide by their responsibilities regarding: the provision of source data and the obligation to establish and maintain adequate controls over the use of personal or other sensitive data they are entrusted with; and/or the management of the project from start to finish and the obligation for its sound conduct, in line with the agreed requirements of the responsible agencies.

✓ These processes will help to strengthen the integrity of big data analytics projects and help to maintain the public's confidence in the Government's stewardship of data.

---

[50] Office of the Australian Information Commissioner, Privacy Impact Assessment Guide,
http://www.oaic.gov.au/privacy/privacy-resources/privacy-guides/privacy-impact-assessment-guide

Principle 4: *Skills, resources and capabilities will be shared*

Skills and expertise in data analytics will be shared amongst government agencies and industry where appropriate.

- ✓ Resources such as data sets and the analytical models used to interrogate them, as well as the infrastructure necessary to perform these computations, will be shared amongst agencies where appropriate and possible to do so.

- ✓ Big data analytics capability will be strengthened by a Whole-of-Government approach through efforts such as the DACoE.

Principle 5: *Collaboration with industry and academia*

The industry, research and academic sectors have been working on big data analytics projects for some time and continue to invest heavily in the skills, technologies and techniques involved with big data analysis.

- ✓ Government agencies recognise the research sector as a key partner in delivering insight from big data analytics as well as a key producer and custodian of valuable data collections.

- ✓ Government agencies will collaborate with industry, academia, non-government organisations and other relevant parties locally and internationally on big data analytics. This engagement will be encouraged by the Big Data Working Group, the Australian Government Chief Technology Officer and the whole-of-government DACoE.

- ✓ These engagements will leverage private and public sector experience and expertise in big data analytics and increase government agency knowledge and skills in this area.

Principle 6: *Enhancing open data*

Open data will continue to be enhanced as agencies are encouraged to release information with the objective of outsourcing and encouraging innovation.

Government agencies will approach big data analytics projects under the principles on open PSI. These principles rest on the Gov 2.0 premise that PSI is a national resource that should be available and discoverable for community access and use. Government agencies will need to balance these considerations with any conflicting legal obligations (such as those relating to security and privacy).

- ✓ Where appropriate, the results of any government big data projects will be made public and the data sets used or created in the analytical process will be released onto data.gov.au, under open licences, for public consideration and consumption.[51]

- ✓ Big data analytics will build on the implementation of Gov 2.0. The data.gov.au portal will be enhanced and will continue to serve as the central data repository for discoverable and useable government information.

---

[51] Although making data open should be considered as the default, there will be some instances where a cost recovery model still brings the most public benefit, particularly if the quality of the data relies on the beneficiaries of the data to pay to support it.

Actions

Action 1: *Develop big data better practice guidance [by March 2014]*

The Big Data Working Group will work in conjunction with the DACoE to develop better practice guidance that will aim to improve government agencies' competence in big data analytics. This guidance will:

- include advice to assist agencies identify where big data analytics might support improved service delivery and the development of better policy;
- identify necessary governance arrangements for big data analytics initiatives;
- assist agencies in identifying high value datasets;
- advise on the government use of third party datasets, and the use of government data by third parties;
- promote privacy by design;
- promote Privacy Impact Assessments (PIA) and articulate peer review and quality assurance processes; and
- include reference to policy and guidance in regards to the use of cloud computing[52].

The guidance will also incorporate existing advice from agencies where there is an opportunity to do so.

For example, the guidance will reference the Principles for Data Integration Involving Commonwealth Data for Statistical Research Purposes[53] which were created by the National Statistical Service (NSS).

The guidance will reference the Statistical Spatial Framework (SSF)[54] developed by the NSS, which provides a common approach to the integration of socio-economic and location data, with a view to improving the accessibility and usability of spatially-enabled information.

The guidance will also reference documents produced by the OAIC including resources to assist agencies in de-identifying data and information.[55]

Input from industry and academia will be sought in the preparation of this guidance. This guidance will also provide advice around assessing risks and managing security when undertaking a big data analytics project.

---

[52] For further information and guidance regarding cloud computing see: http://agimo.gov.au/policy-guides-procurement/cloud/

[53] National Statistical Service. High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes.
http://www.nss.gov.au/nss/home.nsf/NSS/7AFDD165E21F34FDCA2577E400195826?opendocument

[54] National Statistical Service. Statistical Spatial Framework.
http://www.nss.gov.au/nss/home.NSF/pages/Statistical%20Spatial%20Framework

[55] Office of the Australian Information Commissioner. De-identification resources May 2013,
http://www.oaic.gov.au/privacy/privacy-engaging-with-you/previous-privacy-consultations/de-identification-resources-may-2013/

Action 2: *Identify and report on barriers to big data analytics [by July 2014]*

The Big Data Working Group will work in conjunction with the DACoE to identify barriers to the effective use of big data across government. These barriers include technical, policy, legislative skill, resource, organisational and cultural barriers.

Whilst not all barriers can be resolved, a report will be produced that details these barriers and considers possible mitigation and remedial strategies and actions.

Action 3: *Enhance skills and experience in big data analysis [by July 2014]*

The Big Data Working Group will work in conjunction with the DACoE to identify and support a number of big data pilot projects, including existing projects that take advantage of big data analytics as well as the initiation of new big data projects to be led by selected Government agencies. These pilot projects will enhance the development of big data related skills by promoting learning, innovation and collaboration.

Additionally, the Big Data Working Group will work in conjunction with the DACoE to advocate for the wide variety of specific skills for big data analytics to be considered alongside broader skills in ICT in any initiatives that aim to enhance educational curriculums. For example, these skills may include information and communication technology, informatics and statistics, mathematics, socio-economics, business, linguistics and impact evaluation skills.

Action 4: *Develop a guide to responsible data analytics [by July 2014]*

The Big Data Working Group will work in conjunction with the DACoE to develop a guide to responsible data analytics. This guide will focus on the governance of big data projects and will incorporate the recommendations and guidance of the OAIC in regards to privacy.

The guide will also include information for agencies on the role of the National Statistical Service (NSS) and the Cross Portfolio Data Integration Oversight Board and its secretariat.[56]

The guide will incorporate the NSS produced *High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes*[57], this includes how and when agencies should interact with the secretariat as they develop big data projects that involve the integration of data held by Commonwealth agencies. The guide will also investigate the potential for a transparent review process to support these projects.

---

[56] National Statistical Service. Statistical Data Integration involving Commonwealth Data.
www.nss.gov.au/dataintegration

[57] National Statistical Service, High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes,
http://www.nss.gov.au/nss/home.NSF/pages/High+Level+Principles+for+Data+Integration+-+Content?OpenDocument

Action 5: *Develop information asset registers [ongoing]*

The Big Data Working Group will work in conjunction with the DACoE to produce guidance for agencies to assist in the development of agency specific information asset registers.

These information asset registers will support visibility between agencies about what data-sets they have available for re-use.

This action builds on the implementation of Gov 2.0 across agencies and will help to better manage data held by Commonwealth agencies and increase the number of data sets released onto data.gov.au.

This guidance will leverage existing documentation including the guide to publishing PSI[58] and the work surrounding the data.gov.au initiative.

Action 6: *Actively monitor technical advances in big data analytics. [ongoing]*

Members of the Big Data Working Group, supported by AGIMO, will actively monitor technical advances in big data analytics, and call upon industry, research and academic experts to provide updates to the working group.

---

[58] Australian Government Web Guide, Publishing Public Sector Information,
http://webguide.gov.au/web-2-0/publishing-public-sector-information/

Glossary

Cloud computing

Cloud computing is an ICT sourcing and delivery model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

This cloud model promotes availability and is composed of five essential characteristics: on demand self service, broad network access, resource pooling, rapid elasticity and measured service.

Data exhaust

Data exhaust (or digital exhaust) refers to the by-products of human usage of the internet, including structured and unstructured data, especially in relation to past interactions.[59]

Data scientists

A data scientist has strong business acumen, coupled with the ability to communicate findings to both business and IT leaders in a way that can influence how an organization approaches a business challenge. Good data scientists will not just address business problems; they will pick the right problems that have the most value to the organization.

Whereas a traditional data analyst may look only at data from a single a data scientist will most likely explore and examine data from multiple disparate sources. The data scientist will sift through incoming data with the goal of discovering a previously hidden insight, which in turn can provide a competitive advantage or address a pressing business problem. A data scientist does not simply collect and report on data, but also looks at it from many angles, determines what it means, then recommends ways to apply the data.[60]

De-identification

De-identification is a process by which a collection of data or information (for example, a dataset) is altered to remove or obscure personal identifiers and personal information (that is, information that would allow the identification of individuals who are the source or subject of the data or information).[61]

Information assets

Information in the form of a core strategic asset required to meet organisational outcomes and relevant legislative and administrative requirements.

Information assets register

In accordance with Principle 5 of the Open PSI principles, an information asset register is a central, publicly available list of an agency's information assets intended to increase the discoverability and reusability of agency information assets by both internal and external users.

---

[59] McKinsey Quarterly, Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, http://www.mckinseyquarterly.com/Clouds_big_data_and_smart_assets_Ten_tech-enabled_business_trends_to_watch_2647

[60] IBM, What is data scientist, http://www-01.ibm.com/software/data/infosphere/data-scientist/

[61] Office of the Australian Information Commissioner, Information Policy Agency Resource 1, http://www.oaic.gov.au/privacy/privacy-engaging-with-you/previous-privacy-consultations/de-identification-resources/information-policy-agency-resource-1-de-identification-of-data-and-information-consultation-draft-april-201

Mosaic effect

The concept whereby data elements that in isolation appear anonymous can lead to a privacy breach when combined.[62]

Open data

Data which meets the following criteria:

Accessible (ideally via the internet) at no more than the cost of reproduction, without limitations based on user identity or intent.

In a digital, machine readable format for interoperation with other data; and

Free of restriction on use or redistribution in its licensing conditions.[63]

Privacy by design

Privacy by design refers to privacy protections being built into everyday agency/business practices. Privacy and data protection are considered throughout the entire life cycle of a big data project. Privacy by design helps ensure the effective implementation of privacy protections.[64]

Privacy impact assessment (PIA)

A privacy impact assessment (PIA) is a tool used to describe how personal information flows in a project. PIAs are also used to help analyse the possible privacy impacts on individuals and identify recommended options for managing, minimising or eradicating these impacts.[65]

Public sector information (PSI)

Data, information or content that is generated, created, collected, processed, preserved, maintained, disseminated or funded by (or for) the government or public institutions.[66]

Semi-structured data

Semi-structured data is data that does not conform to a formal structure based on standardised data models. However semi-structured data may contain tags or other meta-data to organise it.

Structured data

The term structured data refers to data that is identifiable and organized in a structured way. The most common form of structured data is a database where specific information is stored based on a methodology of columns and rows.

Structured data is machine readable and also efficiently organised for human readers.

Unstructured data

The term unstructured data refers to any data that has little identifiable structure. Images, videos, email, documents and text fall into the category of unstructured data.

[62] Office of the Australian Information Commissioner, FOI guidelines - archive,
http://www.oaic.gov.au/freedom-of-information/freedom-of-information-archive/foi-guidelines-archive/part-5-exemptions-version-1-1-oct-2011#_Toc286409227

[63] Open Data White Paper (UK), Unleashing the Potential,
http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf

[64] http://www.privacybydesign.ca/

[65] Office of the Australian Information Commissioner, Privacy Impact Assessment Guide,
http://www.oaic.gov.au/privacy/privacy-resources/privacy-guides/privacy-impact-assessment-guide

[66] Office of the Australian Information Commissioner, Open public sector information: from principles to practice,
http://www.oaic.gov.au/information-policy/information-policy-resources/information-policy-reports/open-public-sector-information-from-principles-to-practice#_Toc348534327