Australian Government Statistical Forum (AGSF) Agenda
Thursday 6 October 2016
9:30am-12:30pm

***Meeting theme: Privacy and secrecy – methods, tools and approaches to
protecting an individual's personal information***

Archer Board Room, Ground Floor, ABS House, 45 Benjamin Way, Belconnen
Sydney ABS Office, Level 10, 44 Market Street, Sydney

Attendees

Chair	Dr Phillip Gould (ABS)
Members	Ms Milly Lubulwa (ABARES) Mr Geoff Neideck (AIHW) Mr Matt Power (ATO, proxy for Ms Julia Neville (ATO) Mr Glen Malam (BITRE, proxy for Dr Gary Dolman) Ms Robyn Shannon (Employment) Ms Karen McGuigan (DIBP, proxy for Mr Jason Russo) Mr Andrew Lalor (Industry) Ms Annette Neuendorf (Social Services, proxy for Mr David Dennis) Ms Toni Cavallaro (NCVER, proxy for Dr Craig Fowler) Ms Jeanette Corbitt (Environment, proxy for Ms Kerry Olsson)
Apologies	Ms Gemma Van Halderen (ABS), Ms Julia Neville (ATO), Dr Gary Dolman (BITRE), Mr Jason Russo (DIBP), Mr David Dennis (Social Services), Dr Craig Fowler (NCVER), Dr Louise Minty (BoM), Ms Kerry Olsson (Environment), Ms Helen Owens (PM&C)
Presenters	Ms Angelene Falk (OAIC) Dr Christine O'Keefe (CSIRO) Mr Andrew Webster (ABS) Mr Robert Wells (Sax Institute) Mr Geoff Neideck (AIHW)
Observers	Mr Adam Harris (ABS) Mr John Brewer (ABS) Ms Sybille McKeown (ABS) Mr Jeremy Michell (ABS) Ms Joanna Khoo (Sax Institute) Mr Anthony Ashton (Sax Institute)
Minutes	Ms Cassy Short (ABS), Ms Frances Eberbach (ABS)

Agenda Item 1 – Welcome and Chair's Update

1.1   Dr Phillip Gould (ABS) chaired the meeting as a proxy for Ms Gemma Van
Halderen who sent her apologies. Dr Gould acknowledged the traditional owners of
the land and welcomed attendees, inviting them to introduce themselves.

1.2   Dr Gould provided an update on key issues and priorities for the ABS:

- Census – enumeration has now concluded with a 96% response rate and majority of population reporting online; review headed by Alastair McGibbon (the government's Cyber Security Advisor); inquiry by the Senate Economics Reference Committee will report on 24 November – Alistair McGibbon provided submission to the Senate inquiry; Post Enumeration Survey currently underway.
- Statistical business transformation process – a gateway review has been undertaken by the Department of Finance; ABS data lab is up and running; work is underway to develop statistical training.
- PC Inquiry into Data Availability and Use – draft report expected in November (relevant for government data initiatives and privacy issues, sharing data and making it more available across the Australian economy and the private sector).
- Stakeholder relationship health assessment – results of independent assessment indicate the ABS has improved in this area but from a very low base; there is recognition within the ABS that there is still more to do, such as inviting key stakeholders on board with our strategic directions and targeted secondments.

Agenda Item 2 - The OAIC's Draft Guide to Big Data and the Australian Privacy Principles: an Update (presentation circulated)

2.1   Ms Angelene Falk (Assistant Commissioner, Office of the Australian Information Commissioner) delivered a presentation that provided an update on the OAIC's draft guide to big data and the Australian principles. Ms Falk outlined the role of OAIC – providing guidance in the use of information whilst protecting privacy of the individual. A key issue for the OAIC is currently de-identification, with a workshop occurring on 16 November.

2.2   The context for the OAIC Guide to Big Data and the Australian Privacy Principles:

- Defining big data in terms of high volume/velocity/variety assets that require processing to optimise to enhance policy insights;
- Determining whether the privacy principles are fit for purpose to protect information in a big data context;
- Agency/organisation accountability for how personal information is handled.

2.3   The Guide to big data proposes that we start by using data that's not identifiable. In addition, innovation is required such as engaging more readily with the public so they understand how their information is being used – social licence is a key piece of the equation. The OAIC is working with the Department of Human Services and other agencies on how to better convey how personal information is being dealt with in government agencies.

2.4   Ms Falk indicated that feedback on the draft guide was diverse and positive overall, with 35 submissions ranging across government and private entities. Next steps include:

- Updating the guide to consolidate feedback by providing more examples and best practice;
- Responding to other issues of concern raised by stakeholders e.g. cross-jurisdictional information sharing, the interplay of different privacy laws, individual agency secrecy laws, application of the Privacy Act to define de-identification;
- Exploring the governance around de-identification and the ability to test whether something is de-identified. Data61 has good expertise in this regard; OAIC are exploring this with PM&C's Data Branch.

2.5   OAIC has de-identification guidance on their website now and Ms Falk encouraged members to utilise this as a tool for enhancing privacy. OAIC intends to check their guidance in terms of risk based framework in light of international developments e.g. work in the United Kingdom on anonymization and a risk based framework, and Canada's risk based framework to de-identification. The OAIC intends to engage further on this guidance, including with AGSF members and a workshop as part of the GovInnovate series.

2.6   Dr Gould reiterated the need to get definitions in order so that communication on this topic is clear.

Agenda Item 3 – *Protecting confidentiality while making data available for research and policy analysis (presentation circulated)*

3.1   Dr Christine O'Keefe (CSIRO) delivered a presentation that outlined methods for protecting confidentiality while making data available for research and policy analysis.

3.2   Dr O'Keefe outlined two popular risk models used to protect information: anonymise raw data or confidentialise outputs. Current CSIRO work related to data confidentiality includes the National Data Linkage Centre; anonymisation of statistical analysis outputs; remote analysis (automated methods for remote analysis and assessment of existence/applicability of differential privacy methods); virtual data centres – checklist to assist researchers in how they treat the data.

3.3   Dr O'Keefe also addressed the means of data linkage without sharing names and addresses. Whilst this provides privacy preserving linkage with custodians able to pre-process data then share encrypted data with others, there are also scalability and implementation issues.

3.4   The key messages of Dr O'Keefe's presentation were that privacy performance of population data linkage initiatives has been good; the landscape is

changing with increasingly distributed governance and decision making, and the public are becoming active on privacy concerns; need for sophistication in anonymization models and frameworks.

3.5   Members then shared their perspectives on related issues.

- Mr Geoff Niedeck recognised that having a range of different approaches to protecting the privacy of data is good. He also noted the recent high profile data breaches from data.gov.au. Mr Niedeck emphasised the importance of working with technical issues, understanding the source data, and having expert knowledge to understand data disclosure risk management – without these three things, breaches can and do occur.
- Mr Glen Malam asked what CSIRO are doing on the iterative nature of their work e.g. corporate memory, making what's already released doesn't conflict with something else. Dr O'Keefe indicated that they release same or very different data and employ checking algorithms in remote access.
- Ms Milly Lubulwa enquired about synthetic datasets, specifically the method for creating them, how useful they are for researchers, and how distorted the data is when it gets to the researcher.

Agenda Item 4 – The Five Safes Framework: How the ABS is supporting use of public sector data (presentation circulated)

4.1   Mr Andrew Webster (ABS) delivered a presentation that outlined how the Five Safes Framework is being employed by the ABS to support use of public sector data.

4.2   The 'five safes' are part of clearly articulating a framework for an approach to data disclosure – precise controls around the data are used by applying the five dimensions. Whilst the ABS has traditionally focused on the safe data angle (confidentialise then release), it now takes a wider approach – ABS has a remote access data laboratory that supports researchers to access fine level data.

4.3   Whilst this framework will enable better access to ABS data, it also provides opportunity for other agencies to consider how they can use these dimensions to further release their data. The ABS is positioned to engage with discussions about hosting data because it has the secure environment i.e. remote data laboratory.

4.4   Mr Webster drew members' attention to international examples of secure data release. In the United Kingdom, Office of National Statistics data is entrusted to the UK Data Service, run by the University of Essex, for release of data to researchers/academics.

4.5   Mr Webster indicated that the 'safe projects' aspect of the framework is an area for further work for the ABS – understanding how to judge the intended use of

data. The ABS aspires to expanding the degree of microdata access for researchers/academics but can only do so in compliance with legislation.

4.6   A number of members provided observations:

- Ms Falk expressed interest in the context of the Productivity Commission Inquiry into Data Availability and Use. Ms Falk also queried the practical issue of providing data to researchers through 'five safes' and Mr Webster confirmed that this is commercial arrangement with no fees for 2016 trial.
- Mr Niedeck raised issues around implementation, emphasising the need for considerable judgement on how the single framework is brought together. Mr Webster acknowledged the importance of data governance.

4.7   Mr Webster acknowledged recent data.gov.au breaches and upcoming amendments to privacy legislation introducing new procedures for publishing data on data.gov.au. A special meeting of data champions will be held to discuss the new process.

Agenda item 5 – SURE: Balancing access and confidentiality in the use of administrative data (presentation circulated)

5.1   Mr Robert Wells (Sax Institute) delivered a presentation that explained the Secure Unified Research Environment (SURE) and how it balances access and confidentiality in the use of administrative data.

5.2   SURE is a high performance remote access computing environment funded by NSW government as part of National Collaborative Research Infrastructure Strategy (NCRIS) implemented for collaboration across health and human services in Australia. SURE is part of the Population Health Research Network (PHRN) and makes available data that has been linked for use by trained, appropriate people. Mr Wells described how SURE facilitates safe and secure research collaborations across disciplines, organisations and countries.

5.3   SURE's technical infrastructure and controls have prevented data breach since its implementation. SURE provides shared access to project workspaces remotely over encrypted internet connections (using Citrix). The Curated Gateway controls movement of data and prevents the taking data out or printing.

5.4   All projects hosted by SURE have approval of data custodians and Human Research Ethics Committee. Custodians include Australian Institute of Health and Welfare, Department of Social Services and State and Territory health departments. Approvals specify datasets, variables and level of confidentialisation required. Researchers and their institution must have agreement with Sax Institute prior to accessing SURE.

5.5   A number of members provided observations:

- Mr Andrew Lalor enquired about how SURE is funded. Mr Wells outlined that NCRIS funded the setup of SURE through PHRN; users pay access fees that are subsidised by NCRIS (these can be included in grant processes). Ms Khoo explained further that SURE is transparent about access fees: they are available on their website along with information for grant applications.
- Mr Matt Power asked how outputs are taken away from SURE. Mr Wells explained, once approved by the curator, the researcher can download it and custodians can place whatever requirements on the release of their data.
- Dr Gould enquired about the typical number of concurrent users of SURE. Mr Wells indicated that SURE manages 300 users at any one time, usually 20 active and 60 offline.

Agenda item 6 – Managing Disclosure Risk at AIHW (presentation circulated)

6.1   Mr Geoff Niedeck (AIHW) spoke of balancing the competing priority of risk of releasing confidential data and breaching legislation with reducing the value of data holdings by confidentialising. Within AIHW, all Confidentialised Unit Record Files (CURF), linked datasets and data releases with lowly aggregated data and small cell sizes are subject to disclosure risk management.

6.2   Mr Niedeck informed members that AIHW is an accredited Integrating Authority and explained their separation principle whereby identifying information is always held separately to data. The Statistical Linkage Key (SLK) 581 was designed as a simple method for state/territories to provide data to AIHW for linkage and is only used on incoming data to AIHW.

6.3   The definition of 'identifiable' was raised in the context of the attribute disclosure approach, requiring judgement and knowledge of risk disclosure policies i.e. when identifiable data is released, as well as being able to identify an entity, it may also reveal previously unknown details.

6.4   By managing disclosure risk, AIHW were able to release more data whilst maintaining confidentiality, report on rare cases and apply data provider's confidentiality requirements in addition. Mr Niedeck explained that significant judgement is needed to devise strategies and make disclosure decisions and that there is a need for a balanced approached based on understanding of risks and methods for mitigating risks.

6.5   A number of members provided observations:

- Dr Gould provided that it is not just about our appetite for risk but the appetite of the public (social licence). Mr Niedeck added that community

needs to be able to understand the explanation of how we deal with risk mitigation.

- Ms Jeannette Corbitt enquired whether any entity has researched community attitudes. Dr O'Keefe indicated that Data61 and the OAIC had undertaken sentiment testing and that the UK also conducted public consultation.

Agenda Item 7 – Statistical Matters of Significance

7.1   Dr O'Keefe (CSIRO) raised the topic of the key barriers that agencies are facing in their confidentialisation of data. A number of members provided advice:

- Ms Neuendorf advised that DSS were grappling with a lack of in house expertise and difficulty explaining their model to their Minister's office.
- Ms McGuigan indicated that DIBP are revisiting their data on data.gov.au since Data61 had flagged some issues leading to products being removed. DIBP will explore their approach with AIHW on sensitive topics, such as visa status. DIBP need a strong policy on what can/cannot be done since they have an Act that prevents them from sharing/releasing some data.
- Mr Lalor advised that Industry is one of the grant hubs and therefore other agencies' legislation affects them. Industry is working with ABS to bring business data together for analysis and set up a data lab. ABS officer outpostings have been very valuable both ways, as has working with Data61.
- Mr Wells (Sax Institute) stated that researchers are looking for a broader range of datasets. Ms Neuendorf raised the question of how other entities differentiate between academic researchers and Phd students. Mr Wells indicated that SURE has special arrangement for students.
- Mr Niedeck said the key barriers for AIHW were establishing enduring datasets and reducing unnecessary rigour to get access.
- Ms Lubulwa advised that ABARES is setting up a large data warehouse, the challenge being accessibility with high a volume of sensitive data that cannot be released.
- Mr Malam explained that BITRE surveys are for businesses rather than individuals therefore legislation makes their collections compulsory and allows release of identifiable data if the owner agrees. Their concern is with commercial-in-confidence rather than personal data. Mr Malam queried whether the PC inquiry was appropriate place to deal with the siloed nature of enabling legislation; an earlier and easier step for sharing data across government possibly being to change legislation.
- Ms Cavallaro reported that NCVER had a review of data release one year ago which raised four approaches to suit different products (rounding, suppression, signing of undertaking, perturbation). NCVER has been

collecting a unique student identifier from 2015 but it's not released without permission. NCVER has had no data breaches.

- The Department of Employment has released cohort 4 of the stepping stones dataset. The Australian Data Archive (ADA) will host the data. Data will be available for general access (frequency tables), medium level access (mini datasets – requires singing of ADA confidentiality agreement), restricted access (based on approved proposals). Ernst &Young helping to develop data strategy.
- Mr Power advised that the ATO are looking to reduce the amount of confidentialisation; traditionally only a small sample of highly confidentialised tax file records have been released. The ATO are going to release a new product through the SURE environment (data going back to the 1990s).

Agenda Item 8 – Other Business

8.1   Dr Gould confirmed the next meeting is currently scheduled for May 2017 however there was interest in the group meeting more frequently in 2017. The agreed topic for the next meeting in 2017 was 'data governance' including its links to social licence; how agencies are doing risk assessment; the growing problem of how to keep aware of all the data that's been released; and implications of recommendations from the PC inquiry. It was agreed the PC inquiry would be discussed at the next meeting.

Meeting Close

Action items from 6 October meeting

Action Item	Description	Responsibility	Time-frame	Minute reference
1	*Investigate including a third meeting in 2017*	AGSF secretariat	Before end of 2016	8.1
2	*Include an item on the PC inquiry on the agenda for the next meeting*	AGSF secretariat	Before next meeting	8.1