12 September 2016

To whom it may concern,

I am writing in response to the Senate Inquiry into the preparation, administration and management of the 2016 Census by the Australian Bureau of Statistics (ABS).

From the outset, it is my opinion that holding a census is of great value for public policy makers and researchers alike. It is only by the appreciation of "how the land lies" that policy can be developed to overcome issues or even predict where they may occur. A census process should allow this sort of insight about our nation.

But I have been driven to write this submission by some regrettable decisions made by the Australian Bureau of Statistics in the lead up to this year's census. My background is as a software engineer.

My submission will cover 3 areas.

- Data Security and Privacy
- Data Integrity
- Online help when filling in the paper form

Data Security and Privacy

In the week before Christmas 2015, the ABS quietly announced on their website their intention to retain and use names and addresses in the census. As far as I can tell, no major news publication picked up this story.

It is breathtaking that such a fundamental change would be made when without extensive public consultation when that very same idea was proposed and rejected not once but twice in the past decade. Former head of the ABS Bill McLennan makes this very same point in a much clearer way here.

https://www.privacy.org.au/Papers/ABS-Census_2016_and_Privacy_v8.pdf

As mainstream media picked up on the story in the lead up to August 9, many privacy advocates sought answers from the ABS about how they could possibly ensure data security and privacy with this change. These questions were entirely reasonable given the changes, yet the responses from David Kalisch and Michael McCormack MP could only be described as dismissive towards those genuine concerns.

No attempt was made to explain in concrete terms what benefits may flow. We instead heard condescending statements like "people share more on Facebook". Whilst that may be true, people are sharing that on a voluntary basis in exchange for something that they value. One cannot be fined for not using Facebook. One can sign up to Facebook using a nickname without sharing their address, income, family arrangements. It is self-evidently not the same thing.

Writing like this about data security is pretty abstract. To try to better explain what that means in concrete terms, I have compiled details below of some recent data breaches. Many of these organisations have budgets that far exceed that of the ABS, but they couldn't keep the data secure. Many of these leaks were from departments that unlike the ABS would be anticipating cyber-attacks from nation-state actors, but they couldn't keep the data secure. Some of these breaches were rogue employees or contractors. Some were carelessness in disposal of old equipment. Some were misconfigurations. Some we just don't know. Where possible, I have included the official report or a mainstream news report or expert in the field.

In many cases, these breaches exposed information which is orders of magnitude more significant than that which is collected by the census leaving real people at risk of identity theft or worse. My point is to show that even the most competent and well-meaning organisations can fail to secure data.

United States Office of Personnel Management (OPM)

Hack by foreign entity

This is the agency responsible for the management of security clearances for the United States Government. In July last year, they discovered that through two separate events, details of between 21 and 22 million people, including 5.6 million fingerprints had been stolen

https://www.opm.gov/cybersecurity/cybersecurity-incidents/

Philippines' Commission on Elections (COMELEC)

Hack by online group

A 338 GB database containing the personal information for 54.28 million voters was leaked online, including nearly 16 million fingerprints and passport numbers and expiry dates of over a million overseas voters.

http://www.comelec.gov.ph/?r=CyberSecurity

Australian Department of Immigration

Misconfiguration

In February 2014, the personal details of 10,000 asylum seekers held in Australia were unintentionally published online.

https://www.theguardian.com/world/2014/feb/19/asylum-seekers-identities-revealed-in-immigration-department-data-lapse

Australian Department of Human Services

Unauthorised access by employees

A Senate estimates hearing in July 2013 uncovered 63 investigations where bureaucrats in the Department of Human Services used their high-level access to pry on private files, with 170 cases of potential unauthorised entries in the past financial year.

http://www.theaustralian.com.au/national-affairs/investigations-found-to-be-in-breach-of-australian-public-service-code-of-conduct/story-fn59niix-1226688378866?nk=6f1106212d7ab600b00c014e3becabc5-1473249806

NSA

Unauthorised access by contractor

In May 2013, a contractor for the United States National Security Agency named Edward Snowden leaked tens of thousands of documents from the NSA and co-operating intelligence services including Australia's ASD to the Guardian and others.

https://www.theguardian.com/us-news/the-nsa-files

British Ministry of Defence

Careless disposal or theft of hardware

Personal data from up to 1.7 million people may have been compromised after a hard drive went missing from a firm contracted by MoD.

http://news.bbc.co.uk/2/hi/uk_news/politics/7667507.stm

Turkish Citizenship database

Poorly secured

Personal information about 50 million Turkish citizens were leaked in April 2016.

"As well as the national ID numbers for all of the entrants in the system, it also contains a large amount of other personal information, including full name and parents' names, full address, and date of birth."

https://www.theguardian.com/technology/2016/apr/04/database-allegedly-containing-id-numbers-of-50m-turks-posted-online

Greek Government

Unknown

In November 2012, a man was arrested attempting to sell identification card data, addresses, tax id numbers amounting to about 2/3 the population of Greece.

http://uk.reuters.com/article/greece-theft-idUKL5E8MK80020121120

Companies don't get a free pass either. Some notable recent examples are:

Mossack Fonseca

Hacked

Law firm behind the "Panama Papers" scandal had 11.5 million documents about 214 thousand organisations leaked to ICIJ

http://www.theregister.co.uk/2016/04/05/email_server_hack_led_to_mossack_fonseca_leak/

Anthem

Stolen credentials

Anthem is one of the largest health insurers in the US. In February 2015, they discovered that hackers had gained access to the health records on up to 80 million current and former members.

http://www.latimes.com/business/la-fi-mh-anthem-is-warning-consumers-20150306-column.html

Dropbox

Unknown

Dropbox is a commercial cloud file storage and sync website. In August 2016, 60 million account credentials were released. These credentials have been held by unknown parties for 4 years.

https://www.troyhunt.com/the-dropbox-hack-is-real/

Does the ABS honestly believe that their IT are more adept at securing their systems than the NSA was?

Do they believe that their employees are better at password management than other organisations; picking strong passwords, not writing them down; avoiding using the same credentials for other sites? I put it to you that most people pick poor passwords because they can't get their head around just how quick a dictionary attack is. With consumer level hardware, common password hashes can be computed at a rate measured in 10s of billions per second. For reference there are somewhere around a million words in the English language, so all of them could be attempted in somewhere around 1 millisecond. Let that sink in. If any of their employees so much as had a Dropbox account using the same password, you can assume that it has compromised.

Do they believe that their employees and contractors are without exception more honest than in other Australian Bureaucracies?

The answer to all of these questions should be pretty obvious.

I will conclude with two adages:

- You cannot lose data in a hack that you did not collect
- There is no undo button once data has been breached

Data Integrity

The value of the census is in the fact that a policy maker can rely upon the questions being truthfully answered by virtually everyone. There is a social contract that exists between the ABS and Australians. Australians have historically played their part providing world leading data for analysis. On the other side of the contract is the anonymity that has historically been granted to the responses by the ABS. This social contract is very important to understand because I strongly suspect that the behaviour of the ABS in this year's Census has broken this social contract. This may take decades to repair.

By silently changing the way that names are retained, it allows longitudinal analysis on people's responses to the same questions over time. On one hand that has some benefits, and I acknowledge that. On the other hand, that has substantial privacy implications. Those who do not agree with those changes must either:

- Omit their names from the census, and risk being fined; OR
- Lie in the answers that cause them discomfort; OR
- Answer honestly, suppressing their concerns about privacy

There is a real danger to the census that a significant number of people have done that second option; to lie in the answers that caused them discomfort. This option is the one that is hardest to detect and has the least consequences to the person filling it in, yet it is simultaneously the worst outcome for anyone trying to come up with policy or research based upon the data.

Online help when filling in the paper form

My final piece of feedback is about accessibility when filling out the paper form. There isn't a particularly neat segue into this, but it also needs addressing. It seems that so much focus has been put into getting everyone to do it online that a pretty key characteristic of the paper form was overlooked.

As a background to this, there was a particular question that I found particularly emotional to answer due to an unfortunate coinciding of date. I understand that this is unavoidable; that if you are going to have any census then some people are going to be hurt at some times by some of those questions, so I resolved to answer and move on.

There was however a problem. My answer would depend on how the ABS was interpreting one of the words in the question. Whilst there were some examples beneath the question, I needed further clarification. The census paper states "(i) Go to census.abs.gov.au for more information."

I did that, and I spent the next 15 minutes trawling through that mess of a website trying to find anything that resembled further information on this question. I searched using the search box for the particular term, for the question number. I checked the FAQs. There was nothing I could find. In fact, I challenge you to locate any of the questions labelled in this way; 18, 23, 25, 27, 28, 32, 33, 34, 41, 42, 48 and 60 are all questions that allegedly have more information online.

The URL should not have been the census landing page, but a specific help page. The online census is a great convenience for whoever wants to fill it out that way, but the paper version should not have been treated as a lesser census in this regard.

Conclusion

The census has tremendous value, but only if questions are answered honestly. I call on this committee to acknowledge the overreach in the non-statistical, personally identifiable and long term trackable data requested in 2016 census under threat of financial penalty. I call on the committee to apply pressure to the ABS to not financially penalise those who genuinely attempted to answer every question accurately but withheld personally identifiable information on data security and privacy grounds. I do not know whether it is in the remit of this committee, but if at all possible, I believe that the ABS should be ordered to destroy the collected names (as has been practice in previous censuses) until a proper public debate can be had about the benefits and downsides of holding that information. Whilst the move to online form submission is convenient for some, I urge the committee to demand the ABS provide thorough documentation for each question in a way that is easily accessible for those using the paper form.

Adam Gardner