Submission to the Senate Economics References Committee Inquiry into the 2016 Census

Dr David Glance
Director UWA Centre for Software Practice
The University of Western Australia

6th September 2016

Executive Summary

The ABS 2016 Census was a case study of poor communication and public relations together with poor design, planning, testing and execution of what should have been a relatively simple application.

The events leading up to and including the night of the Census on the 9th August 2016 highlight inexperience or naivety of the realities of running an Internet-based application at the scale the ABS was attempting.

This submission recommends that:

[1] An independent audit of the shortcomings and failures to adequately communicate the changes regarding retention to the public. Recommendations for future best practice communication strategies around these types of communication plans by government agencies in the future.

[2] An independent audit of the system architecture and operational design should be conducted and made public with a full analysis of the development of the application, its testing, its deployment and the particular events that lead up to the switching off and restoration of the service. This audit should be carried out by non-commercial experts because of the potential conflicts of interest that would be inherent in using one consulting firm to judge another.

[3] An audit of the security arrangements for the storage and protection of the names and addresses data. A review of the safety and degree of protection that is afforded by the use of statistical linkage keys by the ABS and end recipients of the data from the ABS using these keys.

[4] The results of the census should be independently audited to convince the public of the answer to this question. It is not sufficient to trust the ABS explicitly in this given the lack of transparency that has been demonstrated by ABS leaders so far.

[5] A committee or group consisting of a range of expertise to formulate minimum standards for the future secure deployment of systems to the Internet to avoid consequences such as these. The proceedings, findings and reports of this committee should be made available to the public.

Introduction

As a disclaimer, any response submitted regarding ABS Census 2016 will, in part, be based on a limited knowledge of the actual events that preceded and took place on and after the official Census night of the 9th August 2016.

To a large extent, this is not the issue at hand because the faults with the Census were principally about what was not done, rather than what was actually done.

[1] Poor communication

What was done badly from the outset was communication with the public. This started with the decision to retain names and addresses for a period of 4 years. The ABS conducted a “Privacy

Impact Assessment" [1] that was announced via a media release on the 11th November 2015 for submissions prior to 2nd December 2015. Just 3 weeks. This process was preceded by more wide scale discussion of the "possibility" of retaining names and addresses in earlier consultations in 2012 and 2013 but the explicit discussion of the realities of what this would mean in the context of what we know now about cybersecurity in 2016 was limited to the decisions made in December 2015.

Subsequent to that decision, media about the intentions and the possible benefits was muted with only a handful of articles [2] and [3]. There was certainly no concerted publicity campaign on behalf of the ABS to justify the retention of names and addresses to the public.

Consequently, the public furore that was sparked by Xenophon and the Greens senators declaring that they would not provide their names for the Census blew up into a avoidable PR disaster for the ABS [3]

Not only had they not fully "socialised" the idea of the increased data retention, but the public became aware of the use of statistical linkage keys (SLK) which would link names and addresses to the rest of the data.

Here again, the problem arose because these keys were being described by the ABS as "anonymous" [1] and "completely anonymous" [5]

The problem is, SLKs are *not* anonymous by any stretch of the imagination. The keys are made up of letters of a person's given name, family name, full date of birth and sex.

It is for this reason that SLKs are not used in health linkage databases exactly because they are not anonymous - the clearly and trivially identify data as belonging to an individual and it really doesn't take very much "collateral information" to be able to reverse engineer an SLK to an actual person.

Preserving privacy in linked data sets has been the subject of much ongoing academic and commercial research. Firms like Apple and Google are now going to extraordinary lengths to preserve privacy including the use of analytics techniques such as "Differential Privacy" [7].

The ABS however is openly using linkage techniques which would have resulted in derision and outcry if commercial companies adopted the same approaches to privacy.

Another element of poor communication was the representations made by statistician David Kalisch who repeatedly talked of how secure the ABS was and on the 3rd August 2016 said the ABS had "the best suecirty features [for which] you could ever ask". [8]. Clearly this was not the case (see below) but Kalisch has continued to make claims about the data being safe at the ABS despite the fact that we know that no system is unhackable [9].

It is the continued unrealistic and unsustainable claims about the nature of the systems and software used by the ABS that has exacerbated the concerns about the public's privacy and confidentiality especially after the spectacular failures of the Census night itself.

There was poor communication about when the online survey should be completed. Many people found accidentally that they could complete the survey as soon as they received their identification number. However, it was not clear whether this was expected or approved. It was not clear from the eCensus application itself that when completing it, it should be done for the situation that would or had taken place on the 9th August. The ABS could have actually communicated this better and actively encouraged people to complete the census forms early to avoid the situation of everyone trying to do this at the same time.

Recommendation:

[1] An independent audit of the shortcomings and failures to adequately communicate the changes regarding retention to the public. Recommendations for future best practice communication strategies around these types of communication plans by government agencies in the future.

[2] The system and IBM

The initial contract to IBM to build and operate the eCensus application [10] did not seem in any way underfunded or specified. The eCensus application itself is not a particularly complicated piece of software and if something similar had been commissioned within a university environment for example, would have been priced in the low 10s of thousands.

The apparent architecture that the ABS and IBM ultimately deployed in IBM's SoftLayer data centre showed a fixed structure of servers running IBM software hosting the eCensus site. Although not an ideal configuration for a site that needs to scale massively for only a few hours of its lifetime, the ABS maintains that scalability testing was done and that it could handle the expected loads. However, the testing seemed inadequate and due to the fixed structure of the servers, there was no room to dynamically change the number of servers to scale with unforeseen load [11].

Everything about this configuration suggests that the ABS and IBM were simply not prepared to run a modern web site in the real-world environment that we now find ourselves in. Distributed Denial of Service (DDoS) protection should have been baked into the design at the outset and assumed as being a certainty.

The blanket blocking of traffic from outside Australia was another knee-jerk reaction that not only didn't end up protecting the eCensus site but stopped legitimate users from

If the allegations that firewalls needed to be rebooted as a result of a DDoS are true, again, this should never have happened. It was alleged [12] that in this process, a backup firewall device was rendered inoperative because it had not been synchronised with the firewall rules of its pair. If this is true, this would have been an unforgivable and basic error that should never have happened.

It is not clear what the role of IBM was in the night in question. It is also not clear under whose responsibility the particular devices and software that failed belonged. The bigger point was the lack of preparation and planning that lead to these failing was clearly the fault of both parties.

There has been a suggestion that cuts to the ABS' budget may have contributed to the poor running of the Census. It is hard to see how this was the case unless the ABS had used the issue of cuts to remove critical staff who had the experience with networking and systems development and operations. Even if this were the case, that would still reflect poorly on the ABS who seemingly did not take seriously the challenge of running an online census for the entire country.

Recommendation:

[2] An independent audit of the system architecture and operational design should be conducted and made public with a full analysis of the development of the application, its testing, its deployment and the particular events that lead up to the switching off and restoration of the service. This audit should be carried out by non-commercial experts because of the potential conflicts of interest that would be inherent in using one consulting firm to judge another.

[3] An audit of the security arrangements for the storage and protection of the names and addresses data. A review of the safety and degree of protection that is afforded by the use of statistical linkage keys by the ABS and end recipients of the data from the ABS using these keys.

[3] The fallout

The immediate consequences of the poor communication and shutdown of the eCensus application are that:

[1] The overall quality of responses and the actual reaching of the necessary response rates will be threatened.
[2] The public will have less trust in the ABS, its ability to keep data protected and how it does its job generally
[3] The cause of eGovernment generally and in particular the goal of pursuing online electronic voting has been set back substantially by the poor performance of an Australian Government agency.

Given the negative publicity around the question of retention of names and addresses prior to the Census and the events of the Census night itself, it would seem optimistic to trust the integrity and non-biased sampling of the data that has been submitted.

Recommendation:

[4] The results of the census should be independently audited to convince the public of the answer to this question. It is not sufficient to trust the ABS explicitly in this given the lack of transparency that has been demonstrated by ABS leaders so far.
[5] A committee or group consisting of a range of expertise to formulate minimum standards for the future secure deployment of systems to the Internet to avoid consequences such as these. The proceedings, findings and reports of this committee should be made available to the public.

References

[1] http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Privacy+Impact+Assessment
[2] https://theconversation.com/benefits-of-the-census-retaining-names-and-addresses-should-outweigh-privacy-fears-57223
[3] https://www.crikey.com.au/2016/01/27/govt-to-store-a-trove-of-highly-personal-data-putting-you-at-risk/
[4] https://www.theguardian.com/australia-news/2016/aug/09/census-controversy-shows-abs-needs-to-do-better-says-statistical-society
[5] http://www.abs.gov.au/websitedbs/censushome.nsf/home/privacy?opendocument&navpos=130
[6] http://www.abs.gov.au/AUSSTATS/abs@.nsf/lookup/70A712119ED1C2E1CA2579B9000D1585?opendocument
[7] https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/
[8] http://www.smh.com.au/business/consumer-affairs/australias-top-statistician-david-kalisch-says-abs-is-ready-for-census-2016-20160803-gqk4by.html
[9] https://theconversation.com/censusfail-the-abs-hasnt-convinced-the-public-their-privacy-is-protected-63702
[10] http://www.itnews.com.au/news/ibm-wins-96m-to-host-ecensus-in-2016-397613
[11] http://www.lifehacker.com.au/2016/08/what-organisations-can-learn-from-the-abs-census-fail/
[12] https://theconversation.com/root-of-census-failures-say-badly-done-ibm-and-abs-still-down-for-some-63845

Appendix A Experience and Qualifications of Dr David Glance

Dr David Glance has spent 15 years in industry working for companies such as Tibco, IONA Technologies, Microsoft, HSBC and Salomon Brothers.

His experience has principally been in scalable distributed architectures including scaling web architectures at Microsoft.

At UWA, the Centre for Software Practice focusses on research and development of software relating to health but has involved the development and operations of commercial systems deployed in the cloud.

Dr Glance is a columnist for The Conversation, covering the interface of technology and society. He has written over 350 articles that have been republished regularly on sites such as the ABC, SBS, The Age, SMH, Washington Post, Lifehacker, Gizmodo and others.