Chapter 5
Information security

5.1 This chapter considers the security of the information kept by the ABS in order to undertake the census and associated activities. This chapter firstly discusses the logistical and administrative arrangements put in place to ensure information security, and then considers issues brought to the committee's attention regarding information security throughout this inquiry.

How the data will be stored and handled

5.2 Data provided through the eCensus application was encrypted during transmission and at rest within the IBM datacentre in NSW.[1] The ABS was the only organisation with the decryption keys to the census data.[2] As IBM explained to the committee:

> In terms of the primary security objective here of protecting respondent data, we had encryption mechanisms in place to ensure that the data was fully encrypted while it was in transit—in flight from the respondent to the census site—and that it was encrypted while at rest and stored within the backend of databases. IBM does not have the keys to be able to decrypt that data, so we have not and have never been at any point able to see any of the respondent data that is stored on our systems.[3]

5.3 Once the census data has been provided to the ABS it is decrypted and processed. The ABS proposes to store name and address information separately from one another, and separate from other census information.

5.4 The 2015 PIA gave an overview of how the information gathered in the census would be retained:

> After processing of the Census data, names and addresses would be separated from other personal and household information on the Census data set. Names and addresses would also be separated from each other. Names would not be brought back together with other information collected from respondents to the Census. Anonymised versions of names would be generated for data integration purposes and addresses geocoded.[4]

---

1 IBM Australia Limited, *Submission 87*, pp. 11, 19.

2 IBM Australia Limited, *Submission 87*, p. 19.

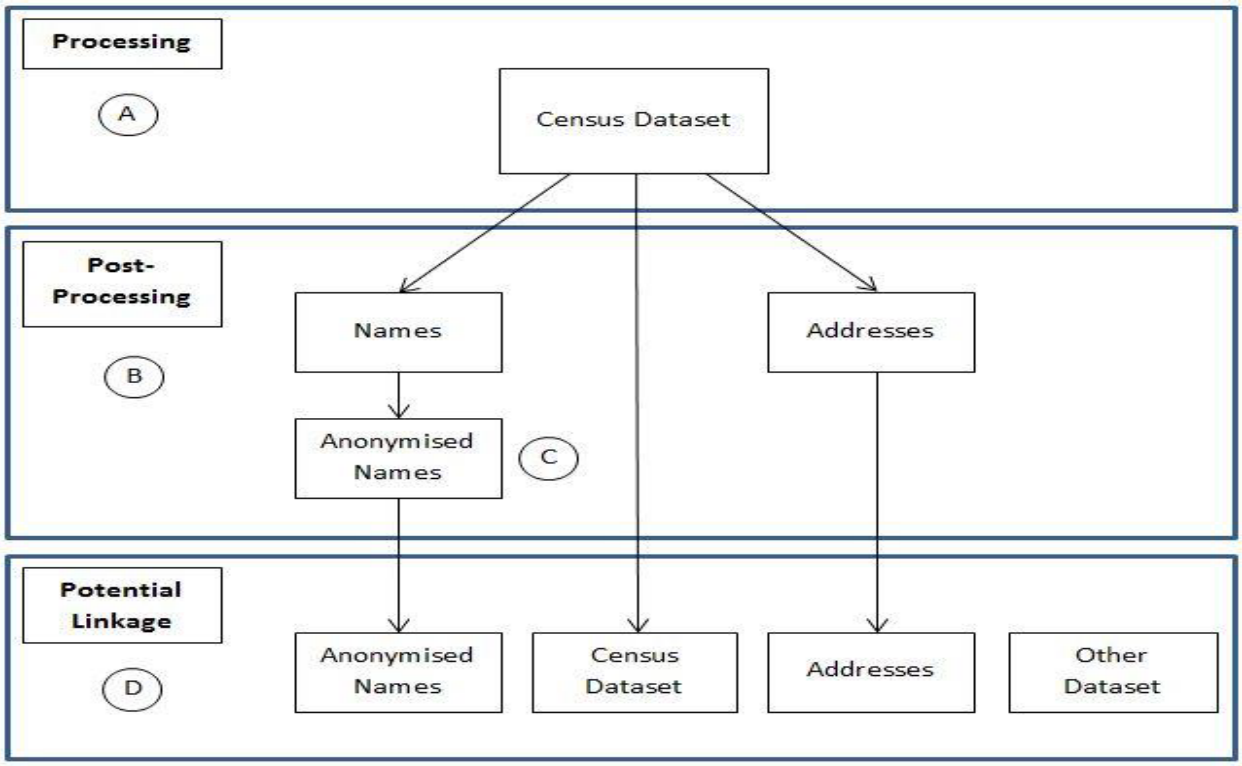
3 Mr Michael Shallcross, Distinguished Engineer for Global Technology Services, IBM Australia and New Zealand, *Committee Hansard* , 25 October 2016, p. 21.

4 Australian Bureau of Statistics, *Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing*, December 2015, p. 11.

5.5 The ABS reports that the structural separation of names, addresses, and other data will mean that authorised ABS officers will only have access to the information required to support their role. Additionally, only a limited number of ABS staff would have access to the retained information.[5]

5.6 The 2015 PIA included an information flow diagram (figure 1) outlining how the ABS would handle census data.

Figure 1 Map of Information flows[6]



5.7 The ABS informed the committee that personal information is heavily protected with high-restricted access controls. Officers only have access to the specific data elements that they need to complete their research, not the entire dataset. Access

5 Australian Bureau of Statistics, *Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing*, December 2015, p. 11.

6 A: Census staff will collect and process data from the 2016 census.

B: Names and addresses are permanently separated from the remainder of the census dataset, and stored securely in separate files with restricted access.

C: Anonymised versions of names would be generated from the names. These are stored separately from both files of names and the census dataset.

D: For approved data integration projects involving 2016 census data, demographic and anonymised name information is recombined on an as-needed basis to allow the census dataset to be used for statistical data linkage.

See: *Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing*, p. 15.

to data will vary depending on the role the officer is performing, with no one staff member having access to both identifying and analytical information from datasets during the linking process.[7]

5.8 The ABS highlighted for the committee its strong track record in information management, noting that:

> The ABS has strong legislative protections founded in the *Census and Statistics Act 1905* that safeguard the identity of a particular person or organisation, and it has a proud history of more than 100 years of maintaining community trust in the way it safely collects, uses, discloses and stores statistical information about people and businesses.[8]

5.9 The ABS began investing in a dedicated data integration facility in 2005 which builds upon and extends the internal mechanisms that the ABS uses to keep personal information secure. The facility was independently accredited as a Commonwealth data integration facility in 2012 satisfying the National Statistical Service accreditation requirements relating to the preservation of privacy.[9] The ABS further assured the committee that data integration projects are closely managed so that privacy is protected:

> The ABS requires all data integration project proposals to go through a rigorous assessment and approval process to ensure the project provides a significant public benefit and takes a privacy-by-design approach. In addition, staff members assigned to a project are never able to see all of an individual's information together at any point of the data integration process and data access rights are only provided on a 'needs to know' basis – this is known as the 'separation principle'.[10]

5.10 The ABS reports constantly improving its safe data dissemination capabilities. These advances have enabled improved access to data held by the ABS by organisations and researchers for statistical and research purposes while protecting privacy. The committee heard that the Australian Census Longitudinal Dataset has been used by over 8000 registered users without a single data breach.[11]

Security concerns about data retention

5.11 Concerns were raised that by storing the name and address information—as well as future datasets that are created from the linkage of census information—the ABS is creating a 'honey pot' or target.[12] It was suggested that the nuanced datasets

---

7 Australian Bureau of Statistics, *Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing*, December 2015, p. 13.

8 Australian Bureau of Statistics, *Submission 38*, p. 24.

9 Australian Bureau of Statistics, *Submission 38*, p. 31.

10 Australian Bureau of Statistics, *Submission 38*, p. 31.

11 Australian Bureau of Statistics, *Submission 38*, p. 32.

12 Dr Monique Mann & Dr Matthew Rimmer, *Submission 75*, p. 21; Name withheld, *Submission 45*, p. 2; Name withheld, *Submission 52*, p. 2; Mr Gary Lord, *Submission 27*, p. [1].

resulting from linking census data would be very tempting to criminal organisations and foreign governments, as well as susceptible to misuse by Australian government and security agencies.[13]

5.12 It was pointed out that due to the nature of digital information a single unauthorised disclosure can release huge amounts of information, and once that information is public there is no way to recover it.[14] Furthermore, the longer the information is held the greater the risk of eventual exposure.[15] It was highlighted that if the data is not collected, then it cannot be exposed.[16]

5.13 Supporters of the changes to the 2016 census emphasised, however, that the changes do not fundamentally alter the security situation:

> That threat is real and is there whether names are retained for 12–18 months or 4 years, and must be countered by appropriate measures. The appropriate response is to take adequate measures to protect data, not to shut down useful and productive applications.[17]

5.14 It was argued to the committee that security experts have begun seeing data as a new 'toxic asset' in that it always poses a risk to those who guard it. The easiest way to protect information is not to have that information in the first instance.[18] One submitter related an allegory from a conference on Big Data:

> The correct way to think about data collection is to treat it as the digital analogue of nuclear waste: a by-product of useful processes that is very difficult to handle safely.[19]

5.15 The committee was provided details of recent unauthorised data releases from a variety of government agencies such as the Department of Immigration and Border Protection, the Bureau of Meteorology, the Department of Human Services, the United States' National Security Agency and the United States Office of Personnel Management, and the United Kingdom's Ministry of Defence, among other private enterprises.[20] It was observed:

> 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

13 Mr John Denham, *Submission 23*, p. [3]; Salinger Privacy, *Submission 24*, p. 11.

14 Mr Adam Gardner, *Submission 4*, p. 3.

15 Salinger Privacy, *Submission 24*, p. 11.

16 Dr Cassandra Cross, *Submission 66*, p.6.

17 Professor Ian Ring, *Submission 9*, p. [2].

18 Australian Privacy Foundation, *Submission 74*, p. 6.

19 Name withheld, *Submission 49*, p. [1].

20 Mr Adam Gardner, *Submission 4*, pp. 2–4; Name withheld, *Submission 18*, p. [4].

> disposal of old equipment. Some were misconfigurations. Some we just don't know.[21]

5.16 These examples highlight that even organisations that believe they are doing everything possible to secure their information can be vulnerable to breaches from a variety of vectors.[22] It was noted that the ABS itself has reported 14 data breaches since 2013.[23]

5.17 In responding to these security concerns, the ABS highlighted the strong institutional framework they have in place to protect personal information. Many people expressed concerns regarding the security of data collected as part of the census. The ABS has, for an organisation of its size and complexity, a very strong track record of treating the information it collects with the utmost of care. The ABS informed the committee that:

> The *Census and Statistics Act 1905* secrecy provision requires that all information, including personal information, provided by the ABS remains strictly confidential and is never released in a manner which is likely to enable an individual to be identified. All ABS staff are legally bound never to release identifiable statistical information collected by the ABS to any external individual or organisation – including courts and law enforcement agencies. This is a lifelong obligation which carries heavy penalties for breaches, including fines of up to $21,600 or imprisonment for up to two years, or both. [24]

5.18 The Australian Institute of Family Studies explained how the ABS provides data and training to research organisations:

> The ABS provides these data in a form that protects the identity of individuals, yet contains sufficient detail to enable research to be undertaken. There are strict protocols about how these data are to be stored, how they can be used, what they may be used for, and who can access these data. The ABS provides training and support to ensure data users have a very thorough understanding of their responsibilities in using Census or other ABS data.[25]

5.19 The committee heard that the ABS' security policies that restrict access to data are sufficiently robust to frustrate some researchers' work. It was pointed out to the committee that there are regular concerns that the ABS does not have the internal resources to process all the data they acquire, but that outside researchers are limited in accessing that information held by the ABS on security grounds.[26]

---

21 Mr Adam Gardner, *Submission 4*, p. 2.

22 Dr Cassandra Cross, *Submission 66*, p. 6.

23 Dr Cassandra Cross, *Submission 66*, p. 7.

24 Australian Bureau of Statistics, *Submission 38*, p. 24.

25 Australian Institute of Family Studies, *Submission 8*, p. 3.

26 Dr Leonard Robert Smith, Visiting Academic, School of Demography, Australian National University, *Committee Hansard*, 25 October 2016, p. 53.

Anonymity and Statistical Linkage Keys

5.20 The committee heard many concerns regarding the use of statistical linkage keys (SLKs) which serve as unique identifiers for projects allowing the ABS to link census information to other datasets. Adding an SLK to each record in each individual dataset allows different datasets to relate to each other so that they can then be brought together into a consolidated, new dataset linked by the unique SLKs.

5.21 Although SLKs appear to provide some level of data security, it was put to the committee that SLKs still contain personal information:

> The use of an SLK would appear to bypass the need to use personal information (e.g. Name and Address) as the key to relate two data sets – something that is very problematic when working between two government departments both governed by the Privacy Act.
>
> …
>
> But there are also problems with SLKs – they are not simply 'random identifiers' such as a Tax File Number that have no intrinsic meaning – they *contain embedded fragments of personal information* – and in fact the more personal information they have embedded, the better they perform. An an SLK is relatively easy to break – even if it is obscured (or 'hashed') using encryption techniques, it can typically be broken at very modest cost, in hours or even minutes.[27]

5.22 It was further put to the committee that SLKs are not sufficient to protect privacy:

> However SLKs do not offer anonymity. At best, they create a pseudonym…[It] is important to note that SLKs do not offer anonymity, let alone privacy. The very purpose of an SLK is to be able to disambiguate between individuals, and thus to link data between datasets, and draw conclusions about the individuals in those datasets.[28]

5.23 A number of submissions raised the specific concern that the ABS would use an algorithm called SLK581 to anonymise records for use in statistical linkages.[29] SLK581 uses a person's name, date of birth and gender to create an identifier. It has been shown that SLK581 does not provide robust anonymity, and is simple to reverse engineer.[30] The ABS has confirmed that it does *not* intend to use SLK581 to create statistical linkage keys.[31]

---

27 Name withheld, *Submission 42*, p. [8].

28 Salinger Privacy, *Submission 24*, p. 7.

29 Salinger Privacy, *Submission 24*, p. 8; Australian Privacy Foundation, *Submission 74*, pp. 1, 8; Name withheld, *Submission 7*, p. 8.

30 Castan Centre for Human Rights Law, Monash University, *Submission 48*, p. [4].

31 Australian Bureau of Statistics, *Submission 38*, p. 81.

5.24 The ABS explained that 'names would be used to generate anonymised versions of names to use as linkage keys in statistical and research projects'.[32] Some submissions pointed out that that ABS has not explained how they intended to generate these 'anonymised versions' of names.[33] The ABS' submission reports that they are working with international experts to arrive at the optimal solution:

> The ABS will use a cryptographic hash function to anonymise name information prior to use in data linkage projects. This function converts a name into an unrecognisable value in a way that is not reversible. There are a number of cryptographic methods that could be used, and the ABS is currently in discussions with international experts in cryptography to determine the most appropriate cryptographic method ahead of the 2016 Census Data Enhancement program commencing in mid-2017.[34]

Statistical linkage keys as unique digital identifiers

5.25 Concerns were raised that SLKs will be used as a way of creating a unified national dataset of personal information.[35] The APF labelled this prospect as the 'Australian Card for big data by digital stealth'.[36] The APF argues in its submission that:

> In the past Australians comprehensively rejected the introduction of the Australia Card. The ABS is using and promoting the SLK, and has the most comprehensive store of data on Australians. The extended use of the SLK is in fact a form of digital 'Australia Card', and one which has new dangers in the context of 'Big Data'.[37]

5.26 The ABS emphasised that it is not creating 'permanent virtual identifiers' that are comparable to a unique identifier for everyone in Australia. Each data linking project will use its own set of SLK, as explained by the ABS:

> The ABS will be creating anonymised linkage keys on a project-by-project basis to allow Census data to be anonymously and safely connected with other existing datasets by the ABS.[38]

5.27 The ABS further confirmed:

> This anonymised version of name will be used with other linkage variables to produce an anonymised linkage keys. Anonymised linkage keys will

---

32 Australian Bureau of Statistics, *Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing*, December 2015, p. 12.

33 Salinger Privacy, *Submission 24*, p. 7; Name withheld, *Submission 42*, p. [8]; Mr Bill McLennan, *Submission 37*, p. [10].

34 Australian Bureau of Statistics, *Submission 38*, p. 81.

35 Mr Stephen Howell, *Submission 78*, p. 2; Ms Rosie Williams, *Submission 85*, p. [4].

36 Australian Privacy Foundation, *Submission 74*, p. 4.

37 Australian Privacy Foundation, *Submission 74*, p. 5.

38 Australian Bureau of Statistics, *Submission 38*, p. 118.

> therefore vary from project to project depending on the characteristics of the datasets to be linked and the variables in those datasets that are available for linkage.[39]

5.28 The 2005 PIA prepared for the 2006 census noted that the privacy risk does not come from creating identifiers, but 'from the creation of the linked unit records, independently of any administrative record number'.[40] The report goes on to note that there is nothing to prevent a third-party creating their own identifier keys if they were able to obtain the data, potentially recreating individual records.[41]

Risk of re-identification from linked datasets

5.29 A number of submissions raised concerns with the potential for datasets created out of the census data being re-identified; that is, individual records from a dataset being directly linked to an individual in the community.[42]

5.30 Improvements in technology and digital archiving have been one of the key driving forces behind statistical linkages and data retention. While improvements in this field have opened up new avenues of research and knowledge, improved computing power can also increases the ability of an adversary re-identifying a dataset. Digital Rights Watch (DRW) argued that constant vigilance is required to ensure security is maintained:

> Updates and developments of technology used to anonymise and store data should be subject to rigorous analysis as to their fitness for purpose. This process should include documented testing, bug bounties and de-anonymisation efforts to demonstrate the veracity of the ABS's claims with some confidence. Best practice will involve taking steps to determine the level of risk of re-identification. This includes an assessment which takes into account the content and value of the original data and the availability of other data that can be linked to this.[43]

5.31 The APF argued that re-identification of anonymised datasets is always a risk, and that the only way to guarantee that re-identification cannot be completed is to not store personal information:

---

39 Australian Bureau of Statistics, *Submission 38*, p. 81.

40 Pacific Privacy Consulting for the Australian Bureau of Statistics, *Census Enhancement PIA Report*, 17 June 2005, http://www.abs.gov.au/websitedbs/D3110124.NSF/f5c7b8fb229cf017ca256973001fecec/fa7fd3e58e5cb46bca2571ee00190475/$FILE/ATT1UQCI/Privacy%20Impact%20Assessment%20report_1.pdf (accessed: 10 October 2016), p. 31.

41 Pacific Privacy Consulting for the Australian Bureau of Statistics, *Census Enhancement PIA Report*, 17 June 2005, http://www.abs.gov.au/websitedbs/D3110124.NSF/f5c7b8fb229cf017ca256973001fecec/fa7fd3e58e5cb46bca2571ee00190475/$FILE/ATT1UQCI/Privacy%20Impact%20Assessment%20report_1.pdf (accessed: 10 October 2016), p. 39.

42 Name withheld, *Submission 49*, p. [2].

43 Digital Rights Watch, *Submission 51*, p. 5.

> In terms of the linkage keys, the issue is that re-identification is a real and pressing problem...So the only way to properly protect people from being re-identified with personal information is to not have that personal information, like names, in there in the first place. That really is the bottom line. If you want to protect Australians from being re-identified through unique identifier keys, it absolutely has to not include sensitive personal identification.[44]

5.32 DRW argued that an appropriate test of whether a dataset is adequately de-identified is the motivated intruder test: whether a reasonably competent motivated person with no specialty skills could succeed in re-identifying the data.[45]

5.33 Salinger Privacy pointed out that statistical disclosure risk—where re-identification is achieved through identifying anonymised records using known information—would increase along with the size and complexity of datasets.[46] The more granular the image, the greater the risk that someone can identify an individual.

5.34 It was pointed out to the committee that there have been examples since the 2016 census of Australian Government agencies releasing datasets that were supposedly de-identified being re-identified. The Department of Health and the Australian Public Service Commission both released datasets that were later able to be re-identified.[47]

5.35 The ABS assured the committee that no information will be released in a way that can be re-identified:

> Under the Census and Statistics Act 1905, the ABS cannot and will not release information in a manner that would enable an individual to be identified. The ABS has built up considerable methodological expertise and capability to meet this requirement and manage the safe dissemination of statistical information.
>
> A range of procedures and techniques are used to ensure an individuals' identity is protected, including removing identifiable information such as name and address; by controlling and limiting the amount of detail available in datasets released to researchers; by slightly modifying or deleting data from datasets released to researchers where that data may enable identification of individuals or businesses; and by requiring individual researchers and their employing organisations to sign legally enforceable undertakings that restrict how they use the data.[48]

---

44 Ms Katherine Lane, Vice-Chair, Australian Privacy Foundation, *Committee Hansard*, 25 October 2016, p. 57.

45 Digital Rights Watch, *Submission 51*, pp. 5–6.

46 Salinger Privacy, *Submission 24.1*, p. 8.

47 Salinger Privacy, *Submission 24.1*, p. 7.

48 Australian Bureau of Statistics, *Submission 38*, p. 118.

5.36 Seemingly in response to the aforementioned recent re-identified data releases by government agencies, the Turnbull Government has proposed introducing legislation that would make it a crime to re-identify data that has been de-identified:

> …[With] advances of technology, methods that were sufficient to de-identify data in the past may become susceptible to re-identification in the future.
>
> The amendment to the Privacy Act will create a new criminal offence of re-identifying de-identified government data. It will also be an offence to counsel, procure, facilitate, or encourage anyone to do this, and to publish or communicate any re-identified dataset.[49]

Mandatory reporting of unauthorised disclosures

5.37 It was suggested to the committee that the ABS should institute a mandatory reporting requirement to ensure that in the case of a data breach involving census data all affected individuals would be notified.[50]

5.38 The committee heard that Australia does not currently have any mandatory data breach notification reporting laws. As was explained in one submission:

> In practice, this means that any organisation who is aware that their system has been compromised in some way (by external or internal factors) is not required to notify affected individuals about the extent of the compromise and what, if any, of their personal data has been exposed.[51]

5.39 Notifying affected individuals of the exposure of their information would allow them to take pre-emptive measures to defend against identity theft and misuse of their personal information.[52] The APF suggested that 'mandatory data breach notification laws, creating enforceable rights for individuals' could help restore trust in the ABS.[53]

5.40 The PIA which prepared the ground for the decision to retain name and address information considered how the ABS should respond to data breaches. These risk management strategies included the notification of affected individuals.[54]

Committee View

5.41 The committee is cognisant that the community wants to know how its information will be protected and used. It notes that no system is entirely secure, to

---

49 Senator the Hon George Brandis QC, Attorney-General for Australia, 'Amendment to the Privacy Act to further protect de-identified data', *Media release*, 28 September 2016.

50 Digital Rights Watch, *Submission 51*, p. 6.

51 Dr Cassandra Cross, *Submission 66*, p. 9.

52 Dr Cassandra Cross, *Submission 66*, p. 9.

53 Australian Privacy Foundation, *Submission 74*, p. 4.

54 Australian Bureau of Statistics, *Privacy Impact Assessment: Proposal to Retain Name and Address Information from Responses to the 2016 Census of Population and Housing*, December 2015, pp. 21–22.

say otherwise is either disingenuous or ignorant. There will always be a risk that data will be exposed: this could come from carelessness; a disgruntled employee wishing to cause harm; a malicious actor; or a change in the legislation governing the use and release of information. The committee is aware that the Australian Government already maintains a large amount of information on the community necessary to provide essential services. And that this information is secure and is only used for its intended purpose.

5.42 The retention of additional information from the 2016 census in the form of name and address information does represent a small additional risk. Previously name and address information was securely stored by the ABS for the period of census processing, approximately 18 months. From an information security perspective, increasing the time that this information will be held to four years does not seem a fundamental change from previous practice which has shown to be secure. However, the committee notes that ABS has failed in objectively arguing its case to the Australian public.

5.43 The use of statistical linkages to gain greater insights into data, when managed properly, is a powerful tool. Although data linking in not a new concept, the scope of application of data matching across the entire Australian population does represent a significant expansion on previous work. The committee believes that the ABS needs to bring the community along with them in this endeavour by honestly explaining how the process will work, what data will be linked, and why it is important.

5.44 The natural inclination of organisations may be to assure people that their data is safe, and that there is no risk. These guarantees cannot be made. The ABS needs to explain that there is a risk that private information may be released or that a dataset could be re-identified. The committee notes that these risks are small however, in comparison to the improvements in government services and economy wide transitions that can be realised through the judicious application of data linking techniques.

5.45 To build community confidence and buy-in in this initiative, the ABS will have to be open with the community regarding how the data is protected, the way data linkages work, and also inform the community immediately when data has been compromised.

Recommendation 3

5.46 The committee recommends that the ABS publicly commit to reporting any breach of census related data to the Office of the Australian Information Commissioner within one week of becoming aware of the breach.