Australian Privacy Foundation

p o s t: GPO Box 1196
Sydney NSW 2001

e m a i l: enquiries@privacy.org.au

w e b : www.privacy.org.au

11 July 2005

Evaluation of the privacy risks posed by the Census Enhancement proposal

Submission to the Australian Bureau of Statistics

Contents

About the Australian Privacy Foundation

The Australian Privacy Foundation is the leading non-governmental organisation dedicated to protecting the privacy rights of Australians. The Foundation aims to focus public attention on emerging issues which pose a threat to the freedom and privacy of Australians.

Since 1987 the Australian Privacy Foundation has led the defence of the rights of individuals to control their personal information and to be free of excessive intrusions. For further information about us see www.privacy.org.au

Introduction

This submission has been prepared by the Australian Privacy Foundation to set out briefly our views as to the privacy risks posed by the Census Enhancement proposal.

Our understanding of the nature of the Census Enhancement proposal (***the proposal***) has been informed by two key documents:

- the original ABS Discussion Paper which describes the proposal : *Enhancing the Population Census: Developing a Longitudinal View*, Discussion Paper 2060.0, 26 April 2005 (***the Discussion Paper***), and
- the subsequent publication of an independent Privacy Impact Assessment, commissioned by the ABS : *Census Enhancement Proposal : Privacy Impact Assessment Report*, Pacific Privacy Consulting[1], June 2005 (***the PIA Report***)

We welcome the opportunity to clarify why we remain opposed to the proposal[2].

Overview of the proposal

The proposal is to change the way information collected through the five-yearly Australian Census is stored, used, linked or matched with other information, and disclosed. If implemented, the proposal will commence in August 2006, when the next Census collection commences.

The primary objective of the proposed changes is to create a Statistical Longitudinal Census Dataset (***the SLCD***), which means a set of data about people which links information about them over time, from one Census to the next. By contrast, the current and long-standing practice is that Census data is not linked either over time, or with information from other sources.

The proposal is thus to convert the Census from being an anonymous 'snapshot' of Australians' lives once every five years, into a 'movie' highlighting changes in their lives.

The method by which Census data is to be linked over time (i.e. from one Census to the next) is described as 'probabilistic' linking[3]. This means using not name and address, but other information about people collected through the Census (such as their date of birth, sex, location, religion, language spoken, etc), to identify individuals to a high enough degree of certainty that the ABS can confidently link one record to another.

---

[1] We note here that the Principal of Pacific Privacy Consulting, Nigel Waters, is also a Board member of the Australian Privacy Foundation. However Mr Waters declared to the Foundation his interest as a consultant to the ABS from the time this possibility was first mooted, and has not participated in the Foundation's deliberations on this census proposal since that time. Nor was the Foundation privy to Mr Waters' views prior to the publication of his report by the ABS. The Foundation's responses to the census proposal have been primarily authored by Board members Anna Johnston and Dr Roger Clarke.

[2] An earlier submission, prepared prior to the publication of the PIA Report, was made by the Australian Privacy Foundation on 7 June 2005.

[3] An earlier proposal by the ABS to use names and addresses to link records from one Census to the next was withdrawn in March 2005 after opposition from organisations including the Australian Privacy Foundation.

The proposal to link Census data from one Census to the next includes retrospective data-matching from the results of the 2006 Census back to the results of the 2001 Census. The SLCD will thus be created in 2006 through the linking of 2001 and 2006 records, and then next added to in 2011, and so on.

However in addition to the *creation* of the SLCD, the proposal also includes various uses of the SLCD, including the linking of Census data with information from other sources – such as information about births, deaths, immigration and disease. In some cases names and addresses are proposed to be used as the key to achieving this linking, while in other cases probabilistic methods are proposed.

The rationale or driver behind the proposal is the creation of richer sources of data for use by researchers. Epidemiological researchers in particular are interested in examining changes in populations over time, which can be relevant to finding cause-and-effect trends in relation to public health.

Key elements of the proposal

The PIA Report summarises the proposal into its eight component parts and labels them from A to H. For ease of reference we will use the same system. We reproduce here Table 1 from the PIA Report, less its footnotes.

Proposal	**Description**
Base or core proposal	
A	The creation of the SLCD by combining data from 2006 Census with data from future censuses. While the SLCD itself will not exist until after the 2011 Census (subject to Proposal B below), the preliminary establishment of the data framework would commence after the 2006 Census
Other proposals (involving probabilistic matching but not name and address)	
B	Extending the SLCD by including 2001 Census data
C	Using the SLCD in conjunction with other ABS datasets for specific approved projects.
D	The use of selected data from the SLCD with specific third party datasets for specific approved statistical research projects. Three distinct groups of third party data sets are specified: • Birth and death register data • Long term immigration data • National disease registers
Other proposals (using name and address information during the period of Census processing to bring together certain datasets for specific uses)	
E	2006 Census data with the 2005-06 Agricultural Census data for analysis

F	Census data and other selected ABS datasets to undertake quality studies
G	Census data and other selected ABS datasets for specific statistical studies where probabilistic matching techniques will not produce a dataset of adequate quality for a significant statistical purpose. ABS has no immediate plans to use name and address for any such studies, but it is a possibility for the future
H	Census data and other selected third party datasets (as listed in option D) for specific statistical studies where probabilistic matching techniques will not produce a dataset of adequate quality for a significant statistical purpose

In addition to the eight components of the proposal identified in the above Table, another three features of the 2006 Census (not the subject of the consultation on this proposal) must be highlighted, to complete the context in which this proposal would operate:

- for the first time names will be captured electronically by the ABS and converted to an electronic text file, and stored for the processing period in the database with the other information from the Census form
- for the second time people will be asked if they consent to have part of their Census forms transferred to the Australian Archives in an imaged, identifiable format, for closed storage for 99 years (the first time this occurred was in 2001)
- the geographic unit used for measuring data is to be made smaller, with the adoption of the 'mesh block', which may contain as few as 30 households

What is extraordinary about this proposal?

First, it is unique in the world. The PIA Report notes that "no other countries currently do exactly what the ABS is proposing to do". For example the UK only links data on identifiable individuals for 1% of the population; but here the ABS proposes to link information about every single person in Australia[4].

The SLCD will be a database holding richer data on Australians than has ever been available before. As the PIA Report notes, that alone means "a more intrusive database than has been held before"[5].

Second, the proposal includes using names and addresses to link Census data to information from other sources, some of which are compulsory collections of personal information from people for a purpose entirely unrelated to this proposal (such as registering a birth or travelling overseas). This would represent a breach of trust by the government agencies supplying that data to the ABS.

[4] See the PIA Report paras 4.2, 4.5, and 4.12.
[5] PIA Report para 11.3.

And third, the proposal is retrospective. The "systematic use of 2001 Census data in ways not envisaged or explained at the time of collection" represents a significant breach of trust by the ABS[6].

How privacy risks can be minimised

Privacy is not only protected by laws prohibiting the disclosure of identifiable personal information, or by the use of secure technologies to minimise the risk of unauthorised access to the information.

While these are important features of any framework of privacy protection, they are by no means enough. 'Fair information' principles adopted around the world to protect privacy also stress the importance of limiting the collection and storage of personal information in the first place, to only what is absolutely necessary[7].

In short, sometimes the best way to protect privacy is to not collect information about people at all, or to destroy information as soon as possible after its primary purpose has been met.

For decades Australians have trusted the Census collection because it has incorporated practical features to protect our privacy - like destroying all the paper forms once they were processed, and even during the time before that point of destruction, not entering names and addresses into the database alongside the 'statistical' information we expected to be kept.

The safeguards for this proposal are not enough

The ABS has responded to the PIA Report's identification of the various privacy risks by pointing to "a number of safeguards within the Proposal"[8]:

- the law : legal obligations under the *Census and Statistics Act 1905* and the *Privacy Act 1988* which prevent the release of identifiable data by the ABS
- technology : high standards of security in the ABS technological environment, including an audit program
- staff integrity : a track record of defending the principle of confidentiality
- the destruction of name and address information held by the ABS after 15 months

The law is not enough

The law can be changed.

---

[6] PIA Report para 10.37.

[7] In 1980 the OECD issued *Guidelines for the Protection of Privacy and the Trans-Border Flow of Personal Data*. The 'fair information principles' contained therein form the basis of privacy laws to protect personal information around the world, including Australian privacy laws.

[8] See *ABS response to the Census Data Enhancement Privacy Impact Assessment*, June 2005, at http://www.abs.gov.au/websitedbs/d3310114.nsf/Home/Census

The PIA Report notes an example from Australian history of the willingness of a past government "to put short term administrative needs ahead of principle" in relation to access to Census records[9]. As the PIA Report notes, how we measure the magnitude and likelihood of the privacy risks "depends largely on levels of trust in future governments not to overturn the longstanding principles underlying the *Census and Statistics Act*".

While it will not always be so, the current Australian Government has a majority in both Houses of Parliament. There is therefore little in our Parliamentary processes to prevent a change of law even before the 2006 Census collection date.

Nor can our judicial system protect the privacy of Census data. Now unique amongst democratic countries, Australia does not have the benefit of a bill of rights, which would allow judges to overturn laws inconsistent with human rights, or provide relief for any person harmed as a result[10]. A bill of rights can also function to temper political pressures or temptations, such as the temptation to access or use Census or SLCD information (including the imaged forms held by the Australian Archives) for new purposes.

The PIA Report concludes[11] that:

> ***in the absence of any constitutional protection of privacy, (legislative safeguards) are ultimately vulnerable to the decisions of the government of the day***.

Technology is not enough

Everything from the Pentagon to Paris Hilton's address book has been hacked into in recent years. No system is 100% secure.

The integrity of staff is not enough

The ABS has pointed to an organisational history and culture of staff respecting privacy, noting there has been no prosecution of staff for leaking personal information in the ABS's 100 years of operations[12].

That may be so, but a lack of prosecutions does not prove there have been no leaks.

Nor is an unblemished record a reliable indicator of the future behaviour of staff, because in the past, the richer data simply wasn't there to leak in the first place. The 2006 proposal means both richer data (in the SLCD), and more immediately useful data (because of the linking of name and address with Census data in the same database), will be held by the ABS, which adds up to the creation of information *worth* leaking.

---

[9] See the PIA Report, paras 11.14 – 11.16.
[10] George Williams, "Balancing national security and human rights", Fulbright Public Lecture, University of Melbourne, June 2005, available from *Australian Policy Online*, www.apo.org.au
[11] PIA Report para 11.19.
[12] Paul Williams, head of the census program, quoted in Cherelle Murphy, "Quarrel over use of census database", *Australian Financial Review*, 27 June 2005, p.6.

Other government agencies holding large databases of personal information have been corruptible[13], and there is no reason to believe that ABS staff are any less subject than the rest of the population to the human frailties of self-interest, corruption, misplaced priorities or natural curiosity[14].

Furthermore, by even suggesting the retrospective aspect of this proposal – to use data collected in 2001 "in ways not envisaged or explained at the time of collection"[15] - the ABS is demonstrating that past assurances about their use of data ring hollow.

The destruction of names and addresses after 15 months is not enough

That names and addresses will be deleted after 15 months is not enough to protect against privacy abuses – the damage will have been done.

For that first 15 month period every five years, the risks of privacy breaches will be greatest, because the data will be:

- easiest to obtain (the data will be most accessible in terms of finding the information desired, because names and addresses will be included with the other Census information), and
- at its most valuable (both because the data is as its 'freshest' immediately after collection, and because names and addresses will be included in the information)

The risks associated with the first 15 month period will therefore be greater than in previous Census collections, for two reasons:

- because of the ABS's intention to actually record and store names and addresses with other Census information (i.e. in the same database) for the first time, and
- because of the proposals to use names and addresses in ways not done before (i.e. to link Census information with information from other sources to create even richer datasets)[16].

But the risks don't stop once the names and addresses have been destroyed.

The creation of the SLCD using probabilistic methods to match data on individuals is intended to have, as much as possible, the same utility as if names and addresses were permanently retained on the database along with all the other data – from identifying information such as sex, date of birth, geographic region and country of birth, to particularly sensitive information such as income, religious affiliation and family relationship.

---

[13] See for example the 1990-1992 NSW Independent Commission Against Corruption investigation into the unauthorised release of government information, known as the Roden Report.

[14] Indeed we note a recent example in which a senior officer within the ABS misused his computer access to alter his office footy-tipping competition results, and that attempts by the ABS to maintain its reputation of honesty and integrity by dismissing the officer were undermined by his reinstatement by the Industrial Relations Commission; Leonie Lamont, "Sacked office footy tipster gets a good result, again", *Sydney Morning Herald*, 7 June 2005, at www.smh.com.au

[15] PIA Report para 10.37.

[16] See components E, F, G and H of this proposal.

We believe the ABS can't have its cake and eat it too.

If records for the entire population are to be 'linkable' to a high enough degree of certainty to satisfy researchers seeking longitudinal data on individuals, then the records will be rich enough to make re-identification so much easier too[17].

The privacy risks for the SLCD will therefore be more significant than for traditional Census data, and those risks will continue to increase with each successive Census – because once you add longitudinal data together, the chances of re-identification rise.

In short, it is easier to identify a person when you are looking at a movie, instead of a single snapshot.

The net effect : how privacy protections will be eroded by this proposal

The net effect of the proposal examined here, and other features being implemented for the 2006 Census, is to erode a number of features which in the past have proven effective means by which to ensure the privacy protection of Australians.

The Australian Privacy Foundation is not reassured by promises about how the technology will be secure, or how the law will prevent unauthorised access or non-statistical uses. We have seen too many examples over too many years to believe that either the law or a computer can stand in the way of human nature, political imperatives or market forces[18].

We believe that in order to protect the privacy of all Australians, we must maintain the current system, which uses a mix of practical techniques, as well as the law and technology, to protect our privacy.

Method of protecting privacy		**Will it still exist in 2006?**
Destroy all Census forms (originals and copies) after processing	✘	No. Imaged copies of some completed Census forms will be kept by the Government indefinitely[19]
Do not keep names or addresses once all Census forms have been collected	✘	No. Names and addresses will be kept by the ABS[20]

[17] See advice from the Victorian Privacy Commissioner (para 12 of his submission on this proposal), and from the Federal Privacy Commissioner (referred to in para 10.8 of the PIA Report).
[18] In addition to monitoring privacy abuses on behalf of the Australian Privacy Foundation, a number of members have previously worked in various Privacy Commissioners' offices or equivalents around the world, and have thus seen hundreds of formal investigations into alleged privacy breaches, and thousands of enquiries from people believing their privacy to have been breached.
[19] See the PIA Report para 3.7. We understand the conditions of this to be the same as for 2001, which is that only people who mark that they consent will have their forms kept. The imaged copies will be sent to the Australian Archives on the condition that they not be used for 99 years.
[20] We understand this will only be for the period of 'census processing', but that is 15 months, or 25% of the time before the next Census is conducted.

Do not keep names or addresses linked to other Census information on a database	✘	No. Names will be kept in the same database as (and linked to) other Census information[21]
Do not link Census information over time	✘	No. Census information will be linked over time, from one Census to the next[22]
Do not link Census information with any other source of information	✘	No. Census information will be linked with information from other sources[23]
Do not use names and addresses to link Census information with other information	✘	No. Census information will be linked with information from other sources, using names and addresses as the key to linking[24]
Do not use Census information for purposes not intended or notified at the time of collection	✘	No. Information from the 2001 Census will be used 'retrospectively' in ways not intended or notified in 2001[25]
Do not use other information for purposes not intended or notified at the time of collection	✘	No. Information from other sources will be used 'retrospectively' in ways not intended or notified when the data was collected[26]
Keep data aggregated in large cell sizes	✘	No. The geographic unit used for measuring data is to be made smaller, with the adoption of the 'mesh block', which may contain as few as 30 households[27]
Have legislation prohibiting the disclosure of identifiable information by the ABS		Yes – but it is not enough. The law can be changed.

[21] See the PIA Report para 3.6. We understand this will only be for the period of 'census processing'.
[22] This is the core of the SLCD, i.e. components A and B of this proposal, and will involve creating a unique identifying 'key' for each person so that Census records can be linked together over time using the 'probabilistic' technique.
[23] We understand that components C and D of this proposal will involve using a unique identifying 'key' for each person so their Census record can be linked with other sources of information about them using the 'probabilistic' technique. We also understand from the PIA Report that this aspect of the proposal does not involve adding information from these other sources *into* the SLCD. It involves the linking of data from the SLCD with data from other sources to create a new set of data, for use in particular research studies. This was not clear to us from the Discussion Paper.
[24] This is components E, F, G and H of the proposal. It involves linking records using names and addresses as the primary 'key' to linkage. We understand from the PIA Report that this aspect of the proposal does not involve adding information from these other sources *into* the SLCD. It involves the linking of data from the SLCD with data from other sources to create a new set of data, for use in particular ways. Again, this was not clear to us from the Discussion Paper.
[25] This is component B of the proposal.
[26] For example privacy guarantees given by other government agencies to people registering birth or deaths, or passing through immigration controls at Australian borders, will not have anticipated the linking of such information with Census data.
[27] See the PIA Report para 3.4. We understand that data is nonetheless not intended to be published at these smaller cell sizes.

Have high security protocols around access to databases	Yes – but it is not enough. Technology can be hacked into.
Rely on the integrity of ABS staff	Yes – but it is not enough. Staff are subject to human frailties, and organisational culture can change.

How might privacy be breached?

This erosion of the many practical methods for protecting privacy will leave all Australians' personal information at higher risk of misuse or disclosure for various purposes, unrelated to the original purpose for which it is collected: the public interest in statistics for research and the development of public policy.

Simply by creating a database of richer data than has ever been available before, the PIA Report concludes that ABS will be increasing "its attractiveness to a range of potential users beyond those who would only be interested in statistics"[28].

The SLCD will be like dangling a tempting carrot in front of those who would seek greater access to personal information for a range of purposes: tracking welfare fraud, identifying tax evasion, or locating people wanted for reasons stretching from a lawful public interest purposes (e.g. law enforcement) through lawful but private purposes (e.g. debt collection) to the unlawful (e.g. tracking down individuals to do them harm).

As noted above, these risks will be most acute for the first 15 months after each Census collection, during which the data will be most valuable, and easiest to obtain.

Possible privacy breaches could come from three sources:

- hacking into the system by people who want access to the data,
- unauthorised disclosure of information by ABS staff, whether accidental or deliberate, and
- pressure to use Census / SLCD data for new purposes (political or commercial)

The PIA Report concludes that "by far the greatest concern in relation to the Proposal … is the attraction of the dataset to other official bodies in pursuit of other public interests"[29].

Taking other identity management schemes and proposals over the past two decades as a guide, agencies most likely to be interested in Census / SLCD / other data-matched data will include the Tax Office, Centrelink, DIMIA, the Health Insurance Commission, law enforcement and national security agencies, and State and Territory driver licensing authorities.

[28] PIA Report para 11.3.
[29] PIA Report para 11.13.

The identification of these risks to privacy are no more hypothetical than the touted research benefits of the proposal; they are based on the lessons of history.

In particular the risk of profiling and locating people using Census data (which includes country of birth, language spoken and religion) is significantly increased by this proposal.

The worst excesses involving misuse of census data in the past have included facilitating genocide (the Holocaust in Europe, possibly in Rwanda), systemic racial discrimination (apartheid in South Africa), and internment (American citizens of Japanese descent during WWII)[30].

However the pressure to use information for new purposes may not only come from political responses to fear and hatred of people of 'other' races or religions. The most mundane yet potent source of pressure could be money.

The Victorian Privacy Commissioner has for example warned of the attractiveness of the data to commercial data brokers, who specialise in data mining, matching and profiling for a range of business and government interests[31]. It is not too difficult to foresee a time of budgetary pressures on government agencies, such that the ABS is expected to recoup its costs and/or generate profits by becoming a data supplier to other agencies, business or interest groups.

The 'key' – the creation of a unique national identifier?

The strength of the proposal (in terms of enabling research) is also its weakest link. That weak point is the creation of the 'key' by which information is to be linked.

Whether using name and address or other detailed information about people, the key will be, as the Victorian Privacy Commissioner has advised, "a universal and (relatively) unique identifier". A universal, unique identifier is the basis of any national identification scheme. Even if not intended by the ABS, this type of metadata is extremely valuable, because it is an enabler for other uses beyond those which were originally intended.

The result of this proposal will be a series of databases, each set representing a tempting 'honey-pot' of data, containing the broadest range of the most sensitive personal data, in a form highly convenient for profiling and data mining.

If this proposal were to proceed, we therefore have no doubt that personal information in the various collections would quickly come to be used in ways, and for purposes, additional to the starter-set mentioned in the Discussion Paper.

[30] Paul Chadwick, Victorian Privacy Commissioner, *Submission to the Australian Bureau of Statistics on its proposal – Enhancing the Population Census: Developing a longitudinal view*, 9 June 2005, para 41.
[31] See para 7 in the Victorian Privacy Commissioner's submission on this proposal.

Who believes their privacy is most at risk?

Attitudes towards privacy differ throughout the community, and may reflect both past experience and current vulnerabilities in terms of the fear of abuse of personal information by governments.

In the 2001 Census, almost half of all Australians took their privacy so seriously that they refused to allow the Australian Archives to keep their imaged Census forms, even with the promise that they would be kept in closed storage for 99 years before being used.

However the response rate differed across the population, according to country of origin and income. People born in England were most likely to agree, while people born in Vietnam were least likely. People on higher levels of income were more likely to agree than those on lower incomes.

Indigenous people, and people currently reliant on government support (welfare payments and/or public housing) have also expressed particular concern at the likelihood of their Census information being leaked back to the Government if Census information were linked to other sources of data[32].

How might research be affected?

The Discussion Paper appears to accept without question the notion that privacy acts as a significant barrier to conducting important research projects, and that therefore this barrier must simply be overcome. Yet privacy and research share a significant value - data quality.

If people believe a question being asked of them is overly intrusive, or that their answer may possibly be used to affect them negatively, the quality of their answers becomes less reliable.

The Victorian Privacy Commissioner has warned of public wariness over Census data possibly affecting its quality. He cites by way of example the 1991 Census in the UK, where the fallout from the poll tax debate affected the Census response rate to such an extent that the results were rendered useless[33].

The ABS's own research from focus groups indicated that around 20% of people objected to the idea of using name and address data to establish the link between different censuses, and around 10% felt strongly enough that they would likely react by not completing the Census, or by giving false answers. The research further found that around 5% could be described as "serious resisters" who would object to the linking of their data using probabilistic methods, even if name, address and date of birth were not included. The proposal for creation of the SLCD sits somewhere between these two scenarios, as the proposal is to include date of birth, but not name or address.

[32] See the PIA Report para 5.7.
[33] See para 42 in the Victorian Privacy Commissioner's submission on this proposal.

As a rough indicator then, we may expect somewhere between 5% and 10% of the population to object strongly enough to this proposal that they would give deliberately false answers to the Census, or not answer at all.

Our conclusion

As this proposal is unique in the world, we can only try and predict what the costs – and the benefits – will be.

We believe the ABS is proceeding on an unfair and possibly blinkered basis. The consultation process on this proposal suggests to us that the possible research benefits of the proposal have been taken as 'given' by the ABS, but the privacy risks must be quantified[34].

The rationale for this proposal is better longitudinal data for research purposes, and more particularly, medical research purposes. Medical research is (usually) a noble aim, and like anybody else, privacy advocates would love to see marvellous advances in both medicine and social policy – a cure for cancer, or the eradication of poverty.

But the promised social benefits upon which this proposal rests are unquantified. There is no guarantee that any social benefit will come from this proposal; even if a researcher were to pinpoint the cause of some disease or disadvantage, it does not automatically follow that governments will have the resources or political will to intervene or address the problem.

This is not to say that we do not believe there could be research benefits to the proposal. Nor is it to suggest that if only those benefits could be proven up-front, we would withdraw our objection to the proposal. We do believe, however, that research and privacy must *both* be recognised as important public interests.

There is a delicate symbiotic relationship between the interests of research and privacy, which must be respected and protected.

When privacy is at risk of too great an intrusion, people instinctively protect themselves with one of two responses: they will subvert the system, or disengage from the system entirely. In a research context, that means bad data, or no data at all.

We believe the ABS has under-estimated the risk to the privacy of all Australians posed by this proposal, and has also under-estimated the value placed on privacy by Australians.

Our belief is that the predicted risks to both privacy *and* research are so significant that they outweigh the predicted benefits, and for that simple reason the proposal ought not proceed at all.

[34] It should not pass without notice that there is no equivalent 'Research Impact Assessment'. Indeed in its response to the PIA Report, the ABS noted that it must "weigh up the public benefits of the proposal … against privacy issues, risks and benefits". It then effectively dismisses the privacy risks identified in the PIA Report because they relate "largely to perceptions of 'what might happen' in the future".