



CENSUS DATA ENHANCEMENT PROJECT - STATEMENT OF INTENTION

Background

The primary objective of the Census Data Enhancement (CDE) project was to enhance the value of the Census by combining it with future censuses and possibly other datasets held by the ABS. The central feature would be the Statistical Longitudinal Census Dataset (SLCD) involving all respondents to the Census. The confidentiality of the data would be fully protected by the secrecy provisions of the Census and Statistics Act and the secure environment the ABS maintains to ensure public trust.

A Discussion Paper including specific proposals was issued on 26 April 2005. Public submissions were called for and a large number were received by the ABS. We are very grateful for those who took the time to make a contribution.

A Privacy Impact Assessment (PIA) was commissioned. It was prepared by Mr Nigel Waters of Pacific Privacy Consulting. Mr Waters has had extensive experience in privacy matters including many years as a senior officer at the Office of the Privacy Commissioner.

The submissions and the PIA had a substantial influence on the revised ABS position.

For reasons set out below, the ABS will not proceed with the SLCD as proposed. As a consequence the CDE proposal has been substantially modified.

Summary of Submissions

The PIA stated that the CDE proposal could be seen in two very different lights.

"For many people, the Proposal is likely to be seen as no more than a marginal change in the way in which the Census data is processed and used. No additional data is to be collected or held, and the only uses will be for statistical purposes within the well-established highly protective legislative and procedural framework, with no disclosure of identifiable personal information.

However, some will see the Proposal as a radical departure from established practice, which will create a data resource so rich and valuable for administrative uses that the privacy and secrecy framework under which the ABS operates may come under great and possibly irresistible pressure, if not immediately, then at least in the medium to long term."

The submissions reflected this bipolar view.

A large number of submissions from potential users provided very strong support. This included the use of outputs derived from the SLCD. The submissions also outlined the considerable benefits that would arise from the analytical studies that would be possible under the CDE proposal.

Some proposals also argued that name and address be retained to enable more accurate matching than was possible with the CDE proposal in the Discussion Paper (which relied on statistical matching).

There was a view among many of those supporting submissions that the CDE proposal should not proceed in its current form if it put the accuracy of the 2006 Population Census at risk.

A number of submissions expressed deep concern at the privacy risks associated with the development of a rich longitudinal dataset that related to the whole population. Underpinning their concern was that a SLCD, even with the uncertainty provided by statistical matching, might provide a "honeypot" that might prove to be extremely attractive to future governments. Most recognised that current legislation and practice protects the confidentiality of the dataset and restricts identifiable data to the ABS. However there is concern about how attractive such a dataset might appear for non statistical or administrative uses, and the potential for future legislative changes to enable such use.

These concerns, and their potential to impact the Census, might be heightened with the recent debate on the pros and cons of a National Identity Card.

Privacy Impact Assessment

The PIA points out that non-statistical or administrative uses would only be possible if radical changes were made to the authorising legislation.

The PIA also points out that the framework within which the ABS operates in general, and conducts the Census in particular, is as protective of individuals' personal information as is possible. It also said that statistics legislation, subject to some recommended clarification, provides an assurance that no identifiable personal information will be disclosed by ABS, and that any uses of the information will be confined to statistical and not administrative purposes.

The PIA also notes the ABS is also subject to the *Privacy Act 1988*, which requires compliance with a set of Information Privacy Principles covering a wider range of issues than just use and disclosure. The Proposal could operate consistently with the Privacy Act, assuming the necessary attention by ABS to notices to individuals and to security.

However, despite the rigour of the legislative protections, and the ABS' track record both of procedural safeguards and of defence to the principle of confidentiality, it identifies the residual privacy risk of future changes in legislation to allow administrative or other non-statistical uses.

The PIA states that the magnitude and significance of this risk is ultimately a matter of subjective judgement, which depends largely on levels of trust in future governments not to overturn the longstanding principles underlying statistics legislation.

It points out that the history over the last 100 years has been one of strengthening privacy protection in the ABS legislation rather than the reverse.

The residual risk exists because the five yearly Census results in a uniquely comprehensive data base about all Australians.

The Office of the Privacy Commissioner, in its submission, argues for a legislative framework for the CDE proposal to increase public confidence and reduce this residual risk. In particular, establishing a legislative framework would confirm the intentions of the current Government that it has no intention of using a SLCD in a way feared by those with privacy concerns.

ABS Position

Our prime consideration has always been that the CDE proposal would not affect the quality of the 2006 Population Census. The PIA recognises that the current legislative framework, the secure environment in which the ABS operates, and our history should cause no privacy concern. The privacy concern is about possible future changes not the present.

The ABS believes that such changes are extremely unlikely. It would require a Government that would want to use Population Census data in this way, both Houses of Parliament to approve the change in legislation and a compliant Australian Statistician. Changes of this nature would be in clear breach of the United Nations Fundamental Principles of Official Statistics.

However, we have to recognise that concerns about future changes might get some traction with the media. The Population Census is dependent on strong community support, and a strong negative media campaign at the time of the census, particularly one drawing on issues relating to the national ID Card has the potential to confuse and concern the community and reduce the necessary support for the census. If such a campaign were mounted at the time of the census, the approach taken by the ABS must be such that the issues could quickly be clarified and community confidence regained.

Taking account of this, and the views of those with privacy concerns, and recognising that the opportunity to develop the Census as a statistical resource only occurs at five yearly intervals, the ABS proposes the strategy set out below. This strategy removes the root cause of concern, the comprehensive dataset, that gives rise to the residual risk identified by the PIA. At the same time it does enable many, although certainly not all, of the identified uses.

The ABS Proposal for 2006 Census

Note in relation to all the elements below:

- As has been past practice, names and addresses will not be retained in the census for longer than the period required for census processing, and will be used only in relation to census processing and for ABS quality studies. Names and addresses will be destroyed at the end of census processing.
- As prescribed by statistics legislation, no identifiable data from the census will leave the ABS.
- The exception is those persons who have consented to have their Census records retained by the National Archives to be made available after 99 years.

SLCD:

The SLCD will be based on a 5% sample of the population. Records will be brought together from each census via a statistical match. The sample based nature of the dataset, and the absence of name and address or other identifying information, and the degree of uncertainty introduced by statistical matching, will together make the dataset unsuitable for administrative and other non statistical uses.

The reduction of the dataset to a 5% sample will reduce the utility of the dataset, particularly in relation to understanding issues affecting small areas (eg environmentally related health issues) and small groups (eg better understanding of mortality issues relating to Indigenous populations). This is a real cost to the community in terms of the potential value of the dataset.

On the other hand, the 5% sample will enable many of the proposed uses to be undertaken.

Using 2001 data:

The 5% SLCD will only be created for 2006 and successive censuses. 2001 data will not be included in the SLCD, although some limited bringing together of the data will be undertaken for the ABS quality studies mentioned below.

This will mean that the first longitudinal view of the 5% sample SLCD will not be available until after 2011.

Using the 5% SLCD with other ABS datasets:

This will not go ahead as most ABS datasets themselves relate to samples, and the overlapping sample with the 5% SLCD will be too small to be useful.

Use of 5% SLCD with specified non- ABS datasets:

This will proceed as set out in the Discussion Paper, but in relation to the 5% sample and with the agreement of the custodians of the non-ABS datasets. While the use of the sample will preclude some important uses, there are a number of uses that will be valid even with the sample.

2006 Census data with 2005-06 Agricultural Census Data:

This will not proceed. However a proposal to request consent of Agriculture Census respondents to link their records to their 2006 Population Census records is being considered. This will only proceed if an adequate consent rate is able to be achieved.

ABS Quality Studies:

ABS quality studies using census data and selected datasets at the time of processing of the Census, using name and address will proceed as proposed. Datasets will not leave the ABS, or be accessible to anyone other than ABS officers. They will be destroyed after use. This will allow important studies relating to the quality of ABS outputs to be undertaken. These studies include the Census Post Enumeration Survey (PES), as has been done for previous censuses, evaluation of indigenous mortality estimates based on indigenous identification on birth and death registers, and the evaluation of the quality of statistical matching.

In addition, after census processing is complete, and name and address information has been destroyed, ABS quality studies will proceed using the Census dataset, where statistical matching is adequate to support such use. Datasets will not leave the ABS, or be accessible to anyone other than ABS officers. They will be destroyed after use.

A description of these ABS quality studies will be in the public domain.

Census data and selected datasets for specific statistical studies (at the time of the Census, using name and address, and 100% census):

This will not proceed.

Studies with consent

Seeking consent through the census, for use of census data in conjunction with other data is considered unworkable, and the resulting quality of the data would be unfit for most purposes.

However there are situations where a consent based approach may be possible and the resultant quality of data fit for its intended purpose, for example in relation to the use of census data in conjunction with ABS household survey data. ABS would seek the consent of the individuals involved in the household surveys to use their data with the census information itself. The benefits of the particular use, and safeguards, could be explained more clearly to respondents. In these cases the link would be to the census information itself, not the SLCD.

Similarly, for the selected non ABS datasets outlined in the Discussion Paper, a study of a sample of respondents may be set up on a consent based approach, bringing together the information from the dataset with census information on the basis of respondent consent to do this.

Comparison with UK proposal

Since 1971, the United Kingdom Office of National Statistics has developed an anonymised linked census and vital event data file for one per cent of the population of England and Wales. It covers all UK censuses since 1971. The vital events include births, deaths, cancer registrations, migration, and entries to long stay hospitals.

This longstanding data base has proven to be non-controversial.

There are three main differences with the ABS proposal.

- (i) Name and address matching is used in the UK Longitudinal Study, but will not be for the Australian proposal.
- (ii) Fewer datasets are proposed for use in conjunction with the (sample) longitudinal file in Australia than the UK.
- (iii) The sample size in Australia (expected to be about 1 million) is larger than that for the UK study (nearly 600,000).

Dennis Trewin
Australian Statistician

18 August 2005

—