2060.0

Australian Bureau of Statistics

Discussion Paper

Enhancing the Population Census: Developing a Longitudinal View

2006

100 years of STATISTICS

www.abs.gov.au

Discussion Paper

Enhancing the Population Census: Developing a Longitudinal View

2006

Dennis Trewin
Australian Statistician

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) TUES 26 APR 2005

ABS Catalogue No. 2060.0

Produced by the Australian Bureau of Statistics

INQUIRIES

- For further information about these and related statistics, contact the National Information and Referral Service on 1300 135 070 or Rosalie Butler by email rosalie.butler@abs.gov.au.

CONTENTS

page

PREFACE

The ABS is proposing to enhance the value of data from the 2006 Census of Population and Housing by combining it with future censuses. The central feature of the proposal is the creation of a Statistical Longitudinal Census Dataset (SLCD).

The ABS believes that creating the SLCD, and making it available for use with a limited number of datasets, would greatly expand the amount of data available to inform decision making among governments and the community. Creating this dataset would be an investment in the future, and the value of it would grow over time as more census information is added.

The ABS is seeking the views of the community on this proposal. I encourage you to read this discussion paper, consider the proposal outlined within it, and lodge a submission to the ABS containing your views.

This proposal is unrelated to the proposal to archive census forms for 99 years for those households that agree to do so. That proposal is under separate Government consideration.

Dennis Trewin
Australian Statistician

ABBREVIATIONS

ABS	Australian Bureau of Statistics
AIHW	Australian Institute of Health and Welfare
ASAC	Australian Statistics Advisory Council
DIMIA	Australian Government Department of Immigration & Multicultural & Indigenous Affairs
PIA	Privacy Impact Assessment
SLCD	Statistical Longitudinal Census Dataset

EXECUTIVE SUMMARY

INTRODUCTION

The ABS has been set up, under the *Australian Bureau of Statistics Act 1975*, as the central statistical authority for the Australian Government. Among its functions, it is required to:

- collect, analyse and disseminate statistics and related information,
- avoid the duplication of collection by official bodies, of information for statistical purposes, and
- achieve maximum possible utilisation, for statistical purposes, of information available to official bodies.

In undertaking its role, the ABS is required by the *Census and Statistics Act 1905* to maintain the confidentiality of information provided to it. ABS officers are required to sign an undertaking of fidelity and secrecy, and associated penalties are specified for any breach of this undertaking.

The ABS is currently celebrating 100 years of success in producing quality statistics. This success is built on the high level of community support for, and trust in, the ABS, which in turn is based on the sound reputation it has developed for consistently placing high priority on managing the privacy of information entrusted to it.

In recent years there has been recognition of the importance of data to provide an understanding of how people move through transitions in their lives and what factors influence these transitions (a longitudinal view), as well as data showing how things are at a particular point in time (a cross sectional view). Looking at patterns in individual histories, allows the circumstances that may influence particular outcomes for people to be identified, and appropriate interventions to be developed.

For example geographic location during childhood, or industry and occupation in early years in the work force, may predispose people to a particular disease, which only becomes apparent decades later. Patterns in the history of individuals with that disease, may indicate effective early interventions.

The five-yearly censuses provide a rich source of data for addressing these sort of issues by bringing it together over time, and using it in conjunction with other key datasets. This discussion paper outlines a proposal to enhance the value of the census data by developing a longitudinal view, the Statistical Longitudinal Census Dataset (SLCD). The proposal is in line with the legislated function of the ABS to maximise the use, for statistical purposes, of information available to official bodies. Safeguards are proposed as part of the proposal to ensure that the ABS, while filling this function, manages the privacy of personal information.

The SLCD would be used in its own right, and also brought together with specified datasets for approved statistical projects. The datasets considered for use in conjunction with the SLCD are:

- ABS household surveys

INTRODUCTION *continued*

- the 2005–06 Agricultural Census
- birth and death register data (including cause of death)
- long-term immigration data
- national disease registers.

While the real value of the SLCD would not become apparent for some years, the proposal would constitute an investment in our future statistical information, as data from the 2006 Census is brought together with data from 2011, 2016 and so on. Over time, the SLCD will develop as a significant national statistical asset, with the power to address not only national and state, but also local level issues. To achieve earlier value from the proposal, consideration is being given to bringing the 2001 Census into the SLCD. Views are also being sought on this aspect of the proposal.

ABOUT THE PROPOSAL

Under this proposal the ABS would create a SLCD through combining census data over time. The proposal does not require the retention of names and addresses from the census. Names and addresses will continue to be destroyed following processing of the 2006 Census.[1]

The proposal to enhance the value of the census contains three parts.

Creating the SLCD

- Creating the SLCD by combining the data provided in the 2006 census with the data provided in future censuses, and using statistical techniques (see Glossary), not involving name and address, to bring the census data together.
- The SLCD could be extended by including 2001 Census data, again using statistical techniques. This is an extension to the proposal and particular views are being sought.

Using the SLCD with other datasets

- Using statistical techniques to allow the SLCD to be used with other datasets, namely:
  - other collections conducted by the ABS, such as household surveys
  - birth and death register data, long-term immigration data, and national disease registers.
- In these cases, use would be for specific approved projects only.

Using name and address information at the time of census processing

- During the period of census processing, name and address information could be used to bring together census data and other selected datasets.
- Three distinct purposes are being considered, namely:
  - quality studies - to help the ABS understand and evaluate the quality of its statistical operations and outputs
  - analysis of the 2005–06 Agricultural Census data with Population Census data
  - specific statistical studies - where there is insufficient information to bring the data together with the SLCD, using statistical techniques, to produce a dataset of adequate quality for a significant statistical purpose.

1 For the 2001 Census, all people completing the census form were provided the option of having their name identified census information archived for the use of future generations. The information for those people who elected to have their 2001 information retained is in the custody of National Archives of Australia, and is not accessible for 99 years. A separate proposal to allow this archiving option for future censuses, including the 2006 Census, is under Government consideration.

ABOUT THE PROPOSAL *continued*

- The datasets under consideration for these purposes are:
  - other collections conducted by the ABS, such as household surveys and the agricultural census
  - birth and death register data, long-term immigration data, and national disease registers.
- For ABS household surveys and the agricultural census, the resulting combined dataset would be maintained within the ABS for use for statistical purposes. These datasets would not contain name and address or any other identifiers.
- In all other cases, the resulting combined datasets would be temporary and for use in the specific approved project only.

PRIVACY AND CONFIDENTIALITY PROTECTION

Respecting privacy and safeguarding confidentiality are fundamental components of the ABS business, and have been core to its success over the 100 years of its history. The ABS has a very sound reputation in this regard. Both legislative protection and a very high security capability enable the ABS to ensure strong protection for this proposal.

The ABS is obligated to comply with provisions in the *Census and Statistics Act 1905* and the *Privacy Act 1988* to respect the privacy of individuals and to protect the confidentiality of their data. The provisions outlined under both these Acts would govern the use and release of data under the proposal being discussed in this paper.

In particular, the ABS will ensure the full protection of the secrecy provisions of the Census and Statistics Act applies to any dataset created through this proposal. This Act requires the ABS to keep information provided to it confidential. Identifiable data created through this proposal will not be provided outside of the ABS.

The Australian Statistics Advisory Council (ASAC) is set up under the Australian Bureau of Statistics Act to advise the Australian Statistician and the Minister. It provides an independent annual report to parliament through the Minister. ASAC advice will be sought on the privacy aspects of this proposal, including the ABS response to the Privacy Impact Assessment and to the submissions received through this consultation process.

In addition, the following list of protective mechanisms would be put in place for this proposal:

- a Privacy Impact Assessment covering all aspects of the proposal will be undertaken by a non-ABS privacy expert, and a report published
- any use of the SLCD with data from non-ABS sources would also comply with any privacy, access and other legislative requirements associated with these other sources, including approval through relevant ethics committees where appropriate
- any dataset created under this proposal would be used for statistical purposes only, and not for administrative or client management purposes relating to a specific individual
- datasets created for particular projects would include only the necessary data items for the approved purpose, and no identifiable information would be provided outside the ABS
- in bringing together datasets under this proposal there will be functional separation within the ABS between those responsible for bringing datasets together and those accessing either the source or the resulting datasets

PRIVACY AND CONFIDENTIALITY PROTECTION *continued*

- independent audits and reviews of the usage and management of the SLCD, and associated datasets, would be regularly undertaken

The ABS will continue its practice of destroying census forms containing names and addresses following processing of the 2006 Census.

DATA ACCESS AND AVAILABILITY

Use of the SLCD would be for approved statistical purposes only, as outlined in Section 7 of this paper.

Access to potentially identifiable data would be restricted to ABS employees, with strict access control limits. Any access by researchers to non-identifiable unit records from the SLCD would be through ABS data laboratory (see Glossary).

If the proposal proceeds and 2006 Census data are brought together with 2001 Census data, then the SLCD would be available for analysis from early 2008. From approximately 2012 the SLCD including the 2011 Census data would be available for analysis.

CONSULTATION PROCESS

Users of census data and other interested members of the community are invited to comment on any aspect of this proposal. Submissions may be lodged electronically using the questionnaire on the internet at <www.abs.gov.au>. A hard-copy questionnaire is also available by phoning the National Information and Referral Service on 1300 135 070. The closing date for submissions is 10 June 2005.

Following the release of this discussion paper, the ABS will hold seminars in all capital cities. Anyone wishing to attend a seminar should contact the ABS person in their state or territory (see table in Section 9) for further information on dates, times and locations. The seminars are open to members of the public who are encouraged to attend to obtain more information prior to lodging a submission.

DECISION ABOUT THIS PROPOSAL

At the conclusion of the submission period, all the submissions will be carefully considered and the proposal will be reviewed in the light of feedback received. Recommendations from the Privacy Impact Assessment will also be considered.

The Australian Statistician will make an in-principle decision on whether to proceed with the proposal, or any part of the proposal, in late June 2005. People who have lodged submissions will be notified of the decision.

EVALUATION AND REVIEW

The full benefits of the SLCD would not be realised for some years, as more census data is combined and as the dataset becomes more widely used. Under this proposal, it is intended that the ABS would conduct a review of the management and use of the SLCD. This would include evaluation of the effectiveness of the statistical techniques used to bring the data together.

It is envisaged that this review would take place after 2011 Census data are incorporated into the SLCD.

SECTION 1 INTRODUCTION

INTRODUCTION

The ABS has been set up, under the *Australian Bureau of Statistics Act 1975*, as the central statistical authority for the Australian Government. Among its functions, it is required to:

- collect, analyse and disseminate statistics and related information,
- avoid the duplication of collection by official bodies, of information for statistical purposes, and
- achieve maximum possible utilisation, for statistical purposes, of information available to official bodies.

In undertaking its role, the ABS is required by the Census and *Statistics Act 1905* to maintain the confidentiality of information provided to it. ABS officers are required to sign an undertaking of fidelity and secrecy, and associated penalties are specified for any breach of this undertaking.

The ABS is currently celebrating 100 years of success in producing quality statistics. This success is built on the high level of community support for, and trust in, the ABS, which in turn is based on the sound reputation it has developed for consistently placing high priority on managing the privacy of information entrusted to it.

In recent years there has been recognition of the importance of data to provide an understanding of how people move through transitions in their lives and what factors influence these transitions (a longitudinal view), as well as data showing how things are at a particular point in time (a cross sectional view). Looking at patterns in individual histories, allows the circumstances that may influence particular outcomes for people to be identified, and appropriate interventions to be developed.

For example geographic location during childhood, or industry and occupation in early years in the work force, may predispose people to a particular disease, which only becomes apparent decades later. Patterns in the history of individuals with that disease, may indicate effective early interventions.

The five-yearly censuses provide a rich source of data for addressing these sort of issues by bringing it together over time, and using it in conjunction with other key datasets. This discussion paper outlines a proposal to enhance the value of the census data by developing a longitudinal view, the Statistical Longitudinal Census Dataset (SLCD). The proposal is in line with the legislated function of the ABS to maximise the use, for statistical purposes, of information available to official bodies. Safeguards are proposed as part of the proposal to ensure that the ABS, while filling this function, manages the privacy of personal information.

The SLCD would be used in its own right, and also brought together with specified datasets for approved statistical projects. The datasets considered for use in conjunction with the SLCD are:

INTRODUCTION *continued*

- ABS household surveys
- the 2005–06 Agricultural Census
- birth and death register data (including cause of death)
- long-term immigration data
- national disease registers.

While the real value of the SLCD would not become apparent for some years, the proposal would constitute an investment in our future statistical information, as data from the 2006 Census is brought together with data from 2011, 2016 and so on. Over time, the SLCD will develop as a significant national statistical asset, with the power to address not only national and state, but also local level issues. To achieve earlier value from the proposal, consideration is being given to bringing the 2001 Census into the SLCD. Views are also being sought on this aspect of the proposal.

This proposal does not involve the retention of name and address information from the 2006 Census. Names and addresses held by the ABS will continue to be destroyed following processing of the 2006 Census. In addition, this proposal is different from census 'archiving', where for the 2001 Census, all people completing the census form were provided the option of having their name identified census information stored for the use of future generations. The information for those people who elected to have their 2001 information retained is in the custody of National Archives of Australia, and is not accessible for 99 years.

OVERVIEW OF PUBLICATION

The next Australian Census of Population and Housing will be conducted on 8 August 2006. A significant amount of planning is devoted to each census to ensure that the output reflects the information needs of Australians. *Enhancing the Population Census: Developing a Longitudinal View* continues the series of papers in which the ABS will publish its views and seek comments in the lead up to the 2006 Census.

This discussion paper is designed to inform community consultation about the proposal to enhance the value of the census. The ABS is seeking to obtain comments on views expressed here and further information from potential users on the types of projects for which the SLCD could be used.

Section 2 discusses what the SLCD is, how it would be created and the benefits of this approach.

Section 3 discusses the use of the SLCD in conjunction with other ABS statistical collections, such as household surveys, where the data would be brought together using statistical techniques.

Section 4 outlines how the SLCD could be used in conjunction with a small number of key statistical datasets. In these cases the data would be brought together using statistical techniques.

Section 5 discusses how name and address information, only available during the period of census processing, could be used to bring together other datasets with the SLCD or data from a single census. This would be for studies to understand and evaluate the quality of ABS statistics or for specific statistical studies. The section also outlines how the 2006 Census data could be used with the 2005–06 ABS Agricultural Census.

OVERVIEW OF PUBLICATION *continued*

Section 6 discusses confidentiality and privacy as it relates to the data held by the ABS, and additional arrangements that would be put in place for this proposal.

Section 7 discusses what data would be available from the SLCD and how it would be accessed.

Section 8 discusses some data quality issues for the SLCD.

Section 9 outlines the public consultation process, including what type of information is being sought in submissions.

Appendix 1 lists the topics being considered for the 2006 Census.

SECTION **2**

THE STATISTICAL LONGITUDINAL CENSUS DATASET

THE STATISTICAL LONGITUDINAL CENSUS DATASET

The ABS is proposing to create a Statistical Longitudinal Census Dataset (SLCD), by combining data from the 2006 Census with data from future censuses. An extension to this proposal would be to also combine data from the 2006 Census with data from the 2001 Census.

A true longitudinal dataset is one which contains information collected for the same person over a number of different points in time. As the ABS is proposing to bring together the data from successive censuses through statistical techniques, rather than matching based on name and address, the dataset is being refered to as a 'statistical longitudinal dataset'.

This section discusses the proposal to create the SLCD based on the 2006 Census. It highlights the benefits of enhancing the census data in this way and describes what data are involved and how the SLCD would be created.

BENEFITS OF THE STATISTICAL LONGITUDINAL CENSUS DATASET

The five-yearly census provides considerable information about Australian people and households (see Appendix 1). But there are certain questions of great interest to policy makers, researchers and the wider Australian community that census data alone cannot answer. Over time, the SLCD would build a picture of how society moves through various changes: which groups are affected by different types of change, and in what way. It would show up patterns of how the physical, social and economic environment in which people live at different stages in their life influences their later outcomes.

The value of the SLCD would grow as data are brought together for future censuses. The SLCD by its nature is an investment in the future, and its value will be in answering the research questions of the future.

However, the richness of the census lies in the information it provides on topics such as:

- family structure, and changes in family structure
- education and qualifications
- work, including hours worked, occupation and industry
- income
- housing

Patterns in individual experiences over time in these areas, and factors that might influence these experiences, would be those where the SLCD on its own would provide most insight. The SLCD would have particular power in providing a picture of how regional effects might interact with patterns in individuals experiences over time.

BENEFITS OF THE STATISTICAL LONGITUDINAL CENSUS DATASET *continued*

Example 2.1 - How do parents' education and labour force participation affect their children's subsequent participation?

A current hypothesis is that children growing up in a household where no-one has a job, are less likely as adults, to have jobs themselves. In contrast, children whose parents have higher levels of education and are in stable employment may be more likely to obtain qualifications and achieve stable employment themselves. The SLCD would provide information on these social and economic aspects of family life, and how they flow through to the next generation.

Longitudinal analysis could assist in investigating a range of issues related to employment and education.

Research questions include:

- Does the mobility of a family affect the education and work outcomes for the next generation? and if so, what factors are associated with positive and negative effects?
- Do upgraded qualifications for individuals over time relate to changes in employment history, and what factors are associated with positive effects?
- Do some regions experience movement of professional workers, for example health service workers or teachers, to and from the area? and if so, what factors are associated with this, and is there an impact on longer term outcomes for the communities involved?

Example 2.2 - How does economic restructuring in an area influence long-term outcomes for individuals and their families?

Changes to industry in a region can occur due to changing economic conditions or due to government policy. Such changes have immediate consequences for people in the region, they also have the potential for long-term impacts for both the people directly affected and the community in which they live. While the current five-yearly snapshot provided by the census can highlight changes to the community over time, the SLCD would show some of the dynamics that underlie the bigger picture and the patterns of impact for different groups.

People may leave the area and experience a range of outcomes. They may move to areas with more job opportunities and settle into stable employment, or they may continue to move between areas and not regain a stable work environment. Alternatively, the type of work done by people remaining in the area may change. People might start to work in different more sustainable industries with higher incomes, or they might be unemployed or work in jobs which do not fully use their skills.

Understanding the factors that are associated with positive and negative outcomes for individuals and their families, as a result of economic restructuring, would allow better informed policy development, and effective targeting of support and assistance.

WHO IS COVERED BY THIS PROPOSAL?

All people counted in the census are covered by this proposal.

The scope of the census is all people in Australia on census night, other than foreign diplomats and their families. Visitors to Australia are counted regardless of how long they have been in the country or how long they plan to stay. Australian residents out of the country on census night are not included.

DATA INVOLVED IN THE STATISTICAL LONGITUDINAL CENSUS DATASET

The creation of the SLCD itself would only involve the use of data from the Census of Population and Housing. Other aspects of this proposal involve the use of SLCD with selected other datasets.

All data collected in the 2006 Census, for all individuals and households counted in Australia on census night, would be included.

For the 2011 Census the data collected would be combined with the data provided in the 2006 Census. Data for people covered in 2011 but not covered in 2006 would be added to the SLCD. This would continue for future censuses. This means the SLCD would only have value from 2011 on, and would become more valuable over time. However, as an extension to the proposal, and in order to achieve earlier value, consideration is being given to bringing data from the 2001 Census into the SLCD. Views on this aspect are being sought.

CREATING THE STATISTICAL LONGITUDINAL CENSUS DATASET

Under this proposal the ABS would create the SLCD by combining data from the 2006 Census with data from future censuses, using statistical techniques to bring the data together. An extension to the core proposal is to use similar techniques to combine the data from the 2006 Census with the data from the 2001 Census.

One way to bring the data together over time would be to use personal identifiers, such as name and address information, to bring together records for the same person. However, this would require the retention of name and address information from one census to the next. This raises significant privacy issues. The ABS is not proposing to retain name and address past the time needed to process each census and would not be using name and address information to bring the census data together over time. The ABS will continue its practice of destroying census forms containing names and addresses following census processing, as it has done in previous censuses.

The proposed method of bringing the census data together over time is often referred to as 'probabilistic record linkage'. This involves bringing together data from two different datasets using a number of characteristics such as age/date of birth, sex, geographic region, and country of birth. All possible linkages based on these data items are evaluated. The records for which the linkage is most likely to be correct would be brought together.

For some individuals, the linkage would be correct while for others it will not, but there is no certainty associated with the linkage for any particular individual. This promotes privacy and data protection. Analyses based on the SLCD would have a level of statistical error caused by this probabilistic record linkage.

Within the ABS, the processes of bringing together two datasets would be separated. A special unit would be responsible for the process of determining the best linkages between the two datasets, based on the characteristics, and then providing a key that would enable the full two datasets to be brought together. This unit would not have access to full original data sources, nor access to the resulting combined dataset.

HOW WOULD THE DATA BE USED?

The SLCD would be for statistical purposes only and not for administrative or client management purposes. For a particular project requiring use of the SLCD, a dataset which included only the data items necessary for the analysis would be brought together. Approval of proposals and access to the SLCD would be controlled as outlined in Section 7.

SECTION 3

USE OF THE STATISTICAL LONGITUDINAL CENSUS DATASET WITH ABS HOUSEHOLD SURVEYS

While the Statistical Longitudinal Census Dataset (SLCD) would be a very useful data source on its own, information from ABS household surveys could be used in conjunction with the SLCD for specific statistical purposes.

This section discusses this use of the SLCD with ABS household surveys, including the benefits of this approach.

BENEFITS OF THIS USE

The ABS conducts a number of household surveys which provide a more detailed range of statistical data, but for a sample of the population and at a single point in time. Using the SLCD with household survey data provides a richer environment for researchers to address questions like 'what are the predisposing factors for people to have certain survey measured outcomes?', or 'what are the subsequent impacts on individuals and families of people having certain survey measured outcomes?'

The examples discussed in Section 2 relate to the use of the SLCD (that is, census data combined over time) on its own. The following provides an example of how the SLCD could be used with data from the National Health Survey.

Example 3.1 - How might people's past social, economic and physical environment affect their current health?

The ABS runs health surveys which collect information on health status and conditions, health-related behaviours and health risk factors, in the general population. For people included in this survey, their health data could be combined with the SLCD. Relationships between a person's past social, economic and physical environment, and their current health status or health risk factors, could be studied. As well, relationships between the presence of a particular health condition and subsequent life events, such as educational attainment, labour force participation, income and family formation, could be studied.

Such statistical analyses could be used to assist in the development of health related policy, health education programs, or the planning of services.

HOW WOULD THE DATA BE USED?

The data from the SLCD and the ABS household survey would be brought together using the statistical techniques described in Section 2. There are a number of common data items, such as age/date of birth, sex, geographic region, that are collected in both the census and ABS household surveys which would facilitate bringing the data together.

HOW WOULD THE DATA BE USED? *continued*

Any use of ABS surveys in conjunction with the SLCD would be for statistical purposes only. A dataset would be brought together for the specific purpose, and would include only the necessary data items for the analysis. Approval of proposals and access to the data would be controlled as outlined in Section 7.

SECTION 4

USE OF THE STATISTICAL LONGITUDINAL CENSUS DATASET WITH OTHER SPECIFIED DATA SOURCES

Further use could be made of the Statistical Longitudinal Census Dataset (SLCD) if it were brought together using statistical techniques, with a small number of other datasets, besides ABS surveys. Using the SLCD with other datasets could significantly enhance the statistical value of these sources.

This section describes what other datasets are proposed as appropriate for use in conjunction with the SLCD and some examples of the benefits of such an approach.

BENEFITS OF THIS USE

Section 3 discussed the use of the SLCD with ABS household surveys. Some important data able to be used for statistical purposes are not collected by the ABS, for example birth and death register data. In performing its functions under the Australian Bureau of Statistics Act, the ABS could bring this data into the ABS and its use within the ABS would be protected by the Census and Statistics Act.

Example 4.1 - What did the people who died from a particular type of cancer have in common?

The social, economic and physical environment in which people live can impact on their physical health. Over time, these environments will change as people move around or go through different life stages. The SLCD could establish a picture of these environmental changes over time. Using the SLCD together with cause of death data would allow statistical research to identify the common environmental factors shared by a group of people who died of a certain type of cancer. This might be in terms of the geographic location in which they spent different periods of their life, or it might be in terms of their social and economic environment, including their occupation, industry, income and education level.

Analyses of the SLCD in conjunction with cause of death data could lead to a better understanding of the possible linkages between environment and cause of death. It may also be able to identify, based on common characteristics, those population groups who are at higher risk of the particular type of cancer, and enable more targeted education campaigns, services or other interventions.

Example 4.2 - How long does it take migrants to settle productively in Australia, and what determines the success of such settlement?

As the Australian population ages, and as our populations become more mobile, Australia will become increasingly dependent on the successful settlement of migrants into Australia, and managing the loss of skilled Australians to overseas countries.

BENEFITS OF THIS USE *continued*

People migrate to Australia under various programs, such as the migration program, including the skill and family streams, and the humanitarian program. By combining information relating to year of arrival and the immigration category (from immigration datasets) with the SLCD, statistical analysis of the settlement patterns and experiences of different categories of immigration over time could be undertaken.

Patterns in the pathways taken and outcomes achieved by people after their migration could be studied. For example, analysis of the SLCD and long-term immigration data could answer questions such as:

- How quickly do migrants secure work? Is it long-term work? Is this different for those who arrived under the different programs?
- Can migrants find appropriate housing, and in what time frame?
- What are the educational and labour market outcomes for the children of migrants? How do they compare to their parents' outcomes? What factors are associated with successful outcomes?
- Are migrants staying in Australia? What factors are associated with migrants leaving Australia to resettle elsewhere?

Information from such studies could help inform decisions on the numbers of people brought to Australia under each program, where they might be encouraged to settle, the amount of assistance they require and the level and location of services required to support these people.

WHAT DATASETS WOULD BE INVOLVED?

Use of the SLCD with other sources would be restricted to a small number of key statistical datasets, namely:

- birth and death register data, including cause of death
- long-term immigration data (see Glossary)
- national disease registers (see Glossary).

The ABS currently uses birth and death register data and overseas arrivals and departures data, from immigration datasets, for compiling demographic statistics. The use of these data sources in conjunction with the SLCD would be an extension of the use already made of these data. The national disease registers that would be considered for use under this proposal are those administered by the Australian Institute of Health and Welfare (AIHW) for statistical and research use.

HOW WOULD THE DATA BE USED?

Again, the data would be brought together with the SLCD using the statistical techniques described in Section 2.

Use of the SLCD with another data source would be for statistical purposes only. The datasets would only be brought together after a proposal to conduct a particular statistical research project had been approved. The dataset created would only include the necessary data items for the particular project. Approval of proposals and access to the data would be controlled as outlined in Section 7.

Any use of the SLCD with data from these non-ABS data sources would be required to comply with the privacy, access and other legislative requirements associated with these sources. For example, statistical use of national disease registers held by the AIHW are subject to a strict scientific and ethical review process, which requires approvals from the AIHW Ethics Committee, and also approval from relevant state and territory registries

HOW WOULD THE DATA BE USED? *continued*

including their ethics committees. The use of these registries together with the SLCD would also be subject to these approval procedures.

In the case of these datasets, the limited information available to bring the dataset together with the SLCD using statistical techniques may restrict the uses that can be made of the combined data. So that users can make appropriate use of the data for these key non-ABS data sources, it is proposed that the quality of the linkage achieved through the statistical techniques be evaluated against a name and address based linkage undertaken by the ABS at the time of census processing. These quality studies would be done within the ABS, and quality measures would be released from the studies. This is described further in Section 5.

SECTION **5**

USE OF NAME AND ADDRESS INFORMATION AT THE TIME OF CENSUS PROCESSING

During the period of census processing, which for the 2006 Census is expected to be from August 2006 to November 2007, the ABS uses name and address information collected in the census to assist in processing the census data. After census processing is complete, all name and address information held by the ABS is destroyed.

The previous sections have focussed on the use of the Statistical Longitudinal Census Dataset (SLCD) either in its own right or in conjunction with other data sources, where statistical techniques would be used to bring the data together, and name and address information would not be used.

This section describes how the most recent census data might be used with other datasets at the time the census is being processed, using name and address information to bring the datasets together.

WHAT PURPOSES WOULD BE CONSIDERED?

Three distinct purposes are being considered where name and address would be used to bring together census data and another dataset at the time of census processing, namely:

- quality studies - to understand and evaluate the quality of ABS statistical operations and outputs
- analysis of ABS Agricultural Census data with Population Census data
- specific statistical studies - where there is insufficient information to bring the data together with the SLCD, using statistical techniques, to produce a dataset of adequate quality for a significant statistical purpose.

These are described in more detail below.

QUALITY STUDIES

Benefits of this use

To assist in informed decision making the ABS produces a wide range of statistics. In addition, the ABS provides information to users of these statistics about aspects of their quality, for example, for survey based data, the ABS describes how sampling error impacts on survey estimates.

Census data in combination with other datasets can be used to help the ABS understand and evaluate the quality of its statistical operations and outputs. Such a study is the study of census undercount that has been conducted as part of every census since 1966, and which will continue.

Other quality studies that would use name and address at the time of census processing, that are being considered are:

- studies of the quality of the statistical techniques used to bring together the SLCD and the selected other datasets (as outlined in Section 4), to inform the ABS on the most suitable technique
- studies of the quality of the datasets created using these statistical techniques

Benefits of this use continued

- studies of non-response bias for ABS surveys conducted around the time of the census.

Example 5.1 - What error rate can be expected when Cause of Death information is brought together with the SLCD using statistical techniques, and what impact would this have on research findings?

Example 4.1 (in Section 4) outlined how the SLCD might be used with cause of death data to investigate a particular cause of death. The proposed method to bring these data sources together was by the application of statistical techniques, which has some uncertainty around the quality of the linkages made between the two sources.

In the months following the census it would be possible to bring together cause of death data and census data using both statistical techniques and using name and address information. Through comparing the results of these two processes, a measure of the statistical impact of the statistical techniques used to bring the data together can be developed. This would help inform users about the quality of the combined dataset.

Example 5.2 - Understanding non-response in ABS household surveys

The census achieves amongst the highest response rate of all the ABS household statistical collections. For a number of reasons, response can be lower in particular surveys. By using addresses of the households selected in the survey together with census data, patterns in the census characteristics of non-respondents in these ABS surveys could be compared to the patterns for those who do respond. This would provide a better understanding of the quality of the statistics collected in the survey, and may also help to adjust survey outputs for non-response bias, resulting in better quality statistics.

How would the data be used?

For this section of the proposal, name and/or address information would be used to bring together a combined dataset. The use of name and address would be for internal ABS purposes only. This would only be possible during the period of census processing, when name and address information is held by the ABS for use in processing of the census.

Use of the census data for such quality studies would be for statistical purposes only. The combined dataset would be deidentified on creation, and once the results of the study were available the dataset that was created for the study would be destroyed. Census name and address information held by the ABS would be destroyed after census processing. Approval of proposals and access to the data would be controlled as outlined in Section 7.

Any use of census data with data from non-ABS data sources would be required to comply with the privacy, access and other legislative requirements associated with these other sources.

2005–06 ABS AGRICULTURAL CENSUS

Benefits of this use

The ABS will be conducting an Agricultural Census in 2006, at a similar time to the Census of Population and Housing. A large number of smaller agricultural businesses included in the Agricultural Census correspond to households in the Population Census. For many of these, there is a strong relationship between the business activities and the family involved in operating the business. Bringing these two datasets together for 2006

Benefits of this use continued

would enhance the understanding of relationships between person and family characteristics and the operations of smaller agricultural businesses.

Example 5.3 - Understanding the impacts of restructuring within a specific agricultural industry

There are a wide range of programs and policies affecting agriculture in Australia which would benefit from improved data sources which provide perspectives on both social and economic impacts. For example, targeted industry restructuring programs might be necessary in response to long term commodity price slumps or the need to improve efficiency in an industry as a result of global competition.

Where restructuring programs are being considered for a particular agricultural industry, information from the Population Census about people in farm households that are involved with that industry, such as the demographic profiles, education levels, English proficiency and household income, would assist in understanding the impacts of the proposed reform. Population information would help inform both the type of policies and programs that need to be developed, as well as the methods of communicating these programs to those affected. This information would also help inform possible impacts on rural communities in which these industries operate.

How would the data be used?

Residential address would be the key item that would be used to bring together data from the Population Census and the Agricultural Census, as there are very few common data items collected in both censuses.

Use of residential address would be possible only during the period of processing for the Population Census, when name and address information is used to assist in processing of the census.

Use of the Agricultural Census with the Population Census would be for statistical purposes only. The new dataset created would be deidentified and contain only the data items necessary for analysis from both censuses. Once the data had been brought together, residential address from the Agricultural Census would be destroyed. Population Census name and address information held by the ABS will be destroyed once census processing is complete. Approval of proposals and access to the data would be controlled as outlined in Section 7.

SPECIFIC STATISTICAL STUDIES

Benefits of this use

Bringing together the SLCD and other datasets using statistical techniques (as outlined in Section 4) requires a minimum number of data items common to both sources. There may be specific statistical studies where the dataset involved does not have sufficient information for statistical techniques to produce a combined dataset of adequate quality for a significant statistical purpose.

In these cases consideration would be given to name and address information being used, at the time of the census processing, to bring the dataset together with the current census, and through it the SLCD, for specific approved statistical studies. These projects would address specific significant research questions and result in new statistical outputs.

Benefits of this use continued

This use is being considered for the following datasets only:

- national disease registers
- birth and death register data
- long-term immigration data.

These studies would bring together the datasets at the time of the census processing, as name and address will be destroyed following census processing.

Example 5.4 - What influences survival rates for different types of cancer?

National cancer registries provide information about people with specific types of cancer, including information about first diagnosis, that is necessary to determine survival times and rates. However, national cancer registries have limited information common to the census, and the statistical error associated with statistical techniques to combine the data, may prohibit important studies for uncommon types of cancer.

For example, for a particular type of cancer, analysis of survival rates, as derived from cancer register data, for people with different socioeconomic backgrounds, or from different regions, as indicated by SLCD data, may point to factors that affect survival. If factors affecting survival rates can be identified it may be possible to design interventions to improve them.

How would the data be used?

For these studies, name and address information would be used to bring together a combined dataset. This would only be possible during the period of census processing, when name and address information is held by the ABS for use in processing of the census. At the conclusion of census processing all name and address information collected in the census is destroyed.

Use of the SLCD with another data source would be for statistical purposes only. The datasets would be brought together only after a proposal to conduct a particular statistical research project had been approved. The dataset created would be deidentified and include only the necessary data items for the particular project. Projects would have a time limit and once the project was complete, the dataset would be destroyed. Approval of proposals and access to the data would be controlled as outlined in Section 7.

Any use of census data with data from these non-ABS sources would be required to comply with the privacy, access and other legislative requirements associated with these other statistical sources. For example, statistical use of national disease registers held by the AIHW are subject to a strict scientific and ethical review process, which requires approvals from the AIHW Ethics Committee, and also approval from relevant state and territory registries including their ethics committees. The use of these registries together with the SLCD would also be subject to these approval procedures.

SECTION 6 CONFIDENTIALITY AND PRIVACY

A fundamental aspect of this proposal is the management of privacy. The ABS is obliged to safeguard the confidentiality of data it holds under the *Census and Statistics Act 1905* and the *Privacy Act 1988*. The provisions outlined under both these Acts would govern the use and release of data under the proposal being discussed in this paper.

This section describes the processes that the ABS will put in place to manage privacy in relation to this proposal. As privacy management is fundamental to the ongoing business of the ABS, relevant current ABS practices and procedures in maintaining the confidentiality of the data collected are also described.

MANAGING PRIVACY

Community consultation

The ABS is undertaking this consultation to inform the community about the proposal to enhance the value of the census, and the privacy management steps it proposes to take. The consultation process is seeking feedback about all aspects of the proposal, including steps to manage privacy and safeguard confidentiality.

Legislative protection

The ABS is obligated to comply with the provisions of the Census and Statistics Act and the Privacy Act and to respect the privacy of individuals and protect the confidentiality of their data. The provisions outlined under both these Acts would govern the use and release of data under the proposal being discussed in this paper.

In particular, the ABS will ensure the full protection of the secrecy provisions of the Census and Statistics Act applies to any dataset created through this proposal. This Act requires the ABS to keep information provided to it confidential. Identifiable data created through this proposal will not be provided outside of the ABS.

All ABS officers are legally bound to secrecy under the Census and Statistics Act. Officers sign an undertaking of fidelity and secrecy to ensure that they are aware of their responsibilities. Section 19 of the Act forbids past or present ABS officers from directly or indirectly divulging information collected under this Act, under a penalty of a fine of up to $5,000 or imprisonment for a period not exceeding two years or both.

Destruction of census forms and name and address information

The ABS maintains a longstanding practice of destroying the individual census forms once processing of the census information is complete. Name and address information collected in the census is not retained by the ABS, after completion of census processing.

In the 2001 Census all people completing the census form were provided the option of having their name identified census information stored for the use of future generations. The information for those people who elected to have their 2001 information retained is in the custody of National Archives of Australia, and is not accessible for 99 years.

Access to ABS information

All aggregate outputs from data held by the ABS are checked to ensure they are unlikely to enable the identification of an individual.

Access to ABS information continued

The Census and Statistics Act allows the ABS to make non-identifiable unit record data available for statistical use, at the discretion of the Australian Statistician. When this does occur, tight procedures and protocols govern such access. A high-level ABS committee scrutinises the proposal to ensure that the data are not likely to enable the identification of an individual. This committee advises the Australian Statistician about the disclosure risk of the proposed file. All users who request access to these non-identifiable datasets must state what their statistical purpose is for using the data, and sign an undertaking to keep the data secure and not to attempt to identify an individual. Should a user breach the conditions of the undertaking and deliberately attempt to identify an individual, they are subject to prosecution under the Census and Statistics Act.

Non-identifiable unit record data is released to researchers, under the above conditions, in two ways: on stand alone files; and through the ABS data laboratory (with either on site or remote access to the data laboratory).

The data laboratory access can be more limiting to a researcher, but provides further protection for more detailed data by eliminating any opportunity for a researcher to breach their undertaking and attempt to identify an individual, for example by matching to external files.

Access to non-identifiable unit record data from the SLCD, and the SLCD used in conjunction with other data sources, would be subject to all the above procedures and would be restricted to access through the ABS data laboratory.

Access to any information from the SLCD, and the SLCD used in conjunction with other data sources, would be subject to the approval processes set out in Section 7.

Privacy impact assessment

A Privacy Impact Assessment (PIA) of all aspects of this proposal to enhance the use of the census, will be undertaken by an expert in privacy management. This will ensure the privacy issues surrounding the proposal are fully identified. The PIA will be published and the recommendations from the assessment used to refine the proposal, and help inform the decision on whether to proceed with it.

Oversight of ABS privacy management

The Australian Statistics Advisory Council (ASAC), constituted under the *Australian Bureau of Statistics Act 1975*, has been set up to advise the Minister and the Australian Statistician in relation to the improvement, extension and co-ordination of statistical services provided for public purposes in Australia. It reports independently to parliament through the Minister.

The ABS has sought, and will continue to seek, ASAC advice on this proposal. In particular, the ABS will seek ASAC views on:

- the submissions received through this consultation process and its respones to them
- the report from the PIA
- the ABS response to the recommendations made in the PIA.

The ABS will continue to seek ASAC input on issues arising in its privacy management strategy in relation to the proposal, if the proposal is implemented.

Oversight of ABS privacy management continued

In its role in providing an independent report to parliament through the Minister, ASAC provides a mechanism for high level oversight of the proposal, and the ABS implementation of it. The ABS will support ASAC in this role by providing information as requested, on all aspects of the proposal including information from the consultation process and from the related ongoing review, audit and evaluation processes.

Access control relating to SLCD data within the ABS

Operational arrangements for managing data flows within the ABS, including restricting access to personal information through functional separation of roles, would be developed. They would minimise access within the ABS to personal information to that access which is necessary to achieve the intended outcomes.

In particular, the processes of bringing together two datasets would be separated. A special unit would be responsible for the process of determining the best linkages between the two datasets and then providing a key that would enable the full two datasets to be brought together. This unit would not have access to full original data sources, nor access to the resulting combined dataset.

Data Security

The ABS maintains a high standard of practices to ensure the security of all information it holds. Features of the ABS environment are:

- strong security arrangements for all ABS information technology systems. ABS conforms with IT Security arrangements set out in the *Australian Government Information Security Manual ASCI 33*
- strict control of access to all ABS premises in accordance with the *Commonwealth Protective Security Manual* to ensure compliance with legislative responsibilities
- appropriate personnel security arrangements. Upon appointment all ABS staff undergo security checks and are required to sign an undertaking of fidelity and secrecy
- a secured Internet gateway which is reviewed annually by Defence Signals Directorate
- regular Protective Security risk reviews to ensure that security arrangements continue to be effective
- an ongoing program of security audits and reviews of computer systems and the physical environment.

In addition the ABS induction and training strategy for its staff places strong emphasis on the importance of security in safeguarding confidentiality, and on the appropriate use of the technology environment.

Audit and review

A regular program of independent audits would be set up to cover the processes associated with managing, combining and accessing data involved in this proposal. Usage of the datasets created under the proposal would be reviewed annually.

An evaluation of the proposal, and whether it is achieving adequate privacy management as well as community benefit, is intended, following the addition of 2011 data to the SLCD.

Notifications of use of the data

A list of all research and statistical projects that used the SLCD would be published regularly by the ABS on the ABS website.

Notifications of use of the data continued

A summary of these projects would be included in the ABS Annual Report. This report is tabled in the Australian Parliament.

Under the Privacy Act, the ABS has an obligation to record in the Privacy Information Digest a list of ABS data holdings which contain personal information. The digest is published annually by the Privacy Commissioner. The ABS will ensure that these obligations continue to be met.

SECTION 7 DATA AVAILABILITY AND USE

Researchers would be able access the datasets resulting from this proposal through one of two means:

- access to customised statistical outputs
- access to non-identifiable unit records held within the ABS, through data laboratories (either on site or remote access, see Glossary).

This section discusses what data would be available as a result of this proposal to enhance the use of the census and how the data would be accessed, including approval processes for its use.

ACCESS TO DATA FROM THE STATISTICAL LONGITUDINAL CENSUS DATASET

If the proposal proceeds and 2006 Census data were brought together with 2001 Census data, the Statistical Longitudinal Census Dataset (SLCD) would be available for analysis from early 2008. From approximately 2012 the SLCD would be available for analysis including 2011 census data.

To access data from the SLCD, researchers would specify the intended use of the data. This would be assessed by the ABS to ensure:

- the use was statistical
- the use was likely to be of community benefit
- the SLCD was an appropriate dataset to address the research question
- the use would not require access to identifiable information.

This process may require ABS subject matter statisticians to work with researchers to clarify the proposal.

A tailored dataset, containing only the relevant data items for the research project, would be produced. Any required statistical outputs would be produced and checked to ensure that they would not enable individuals to be identified. If access to non-identifiable unit record data was required, the dataset would be reviewed to ensure access through the data laboratory conditions would not enable individuals to be identified. Access to unit record data would be restricted to access through the ABS data laboratory, either via remote or on site access.

The data items within tailored datasets could theoretically be drawn from the full list of data items collected in the census (see Appendix 1). However, in practice highly detailed data can be identifying. Therefore the detail of information available (for example geographical or occupational detail) would be constrained to ensure no data were released which would be likely to enable the identification of an individual.

Each project would be recorded on the ABS list of projects using the SLCD and published on the ABS website.

There would be costs associated with the customised services required in supporting research projects, and charges would be expected to apply.

ACCESS TO DATA FROM THE STATISTICAL LONGITUDINAL CENSUS DATASET *continued*

The ABS would also consider publishing a standard set of statistical outputs from the SLCD to support frequently researched issues.

ACCESS TO COMBINED DATA FROM THE SLCD AND ABS COLLECTIONS

To access combined data from the SLCD and ABS collections, the process would be the same as for accessing data from the SLCD. The project would be specified and assessed by the ABS to ensure:

- the use was statistical
- the use was likely to be of community benefit
- the data proposed to be used was appropriate to address the research question
- the use would not require access to identifiable information.

A tailored dataset containing only the necessary data items for the research project would be produced. Any required statistical outputs would be produced and checked to ensure that they would not enable individuals to be identified. If access to non-identifiable unit record data was required, the dataset would be reviewed to ensure access through the data laboratory conditions would not enable individuals to be identified. Access to unit record data would be restricted to access through the ABS data laboratory, either via remote or on site access.

Each project would be recorded on the ABS list of projects using the SLCD and published on the ABS website.

There would be costs associated with the customised services required in supporting research projects, and charges would be expected to apply.

The ABS would also consider publishing standard statistical outputs from particular sets of combined data from the SLCD and ABS collections.

ACCESS TO COMBINED DATA FROM THE SLCD AND SELECTED NON-ABS DATA SOURCES

To access combined data from the SLCD and the selected non-ABS data sources included in this proposal, the specific requirements of the custodians of the non-ABS data sources, as well as ABS requirements, would need to be met. For example, statistical use of national disease registers held by the AIHW is subject to a strict scientific and ethical review process, which requires approvals from the AIHW Ethics Committee, and approval from relevant state and territory registries, including from their ethics committees. The use of these registries together with the SLCD would also be subject to these approval procedures.

Once all requirements of the non-ABS data custodian have been met, the project would still need to be assessed as appropriate by the ABS.

Usually, such projects would combine the data using statistical techniques, rather than name and/or address. In these cases the project would be assessed by the ABS to ensure:

- the use was statistical
- the use was likely to be of community benefit
- the data proposed to be used was appropriate to address the research question
- the use would not require access to identifiable information.

ACCESS TO COMBINED DATA FROM THE SLCD AND SELECTED NON-ABS DATA SOURCES *continued*

In the exceptional cases where use of name and/or address information was proposed, at the time of census processing, and assuming all requirements of the non-ABS data custodians were met, the project would be vetted by a specially constituted senior level ABS committee.

This committee would evaluate the proposal to ensure the project:

- was statistically appropriate
- could not be done without name and/or address to bring the data together
- demonstrated significant potential community benefit
- would not require access to identifiable information.

The committee would make a recommendation to the Australian Statistician for final approval.

Once the project was approved, a tailored dataset containing only the relevant data items for the research project would be produced. Any required statistical outputs would be produced and checked to ensure that they would not enable individuals to be identified. If access to non-identifiable unit record data was required, the dataset would be reviewed to ensure access through the data laboratory conditions would not enable individuals to be identified. Access to unit record data would be restricted to access through the ABS data laboratory, either via remote or on site access.

For datasets combining the SLCD and a non-ABS data source, where name and/or address information was used at the time of census processing to combine the data, the date for completion of the project and destruction of the dataset would be specified.

Each project would be recorded on the ABS list of projects using the SLCD, indicating the particular non-ABS data source which was used and whether statistical techniques or name and/or address information had been used to bring the data together.

There would be costs associated with the customised services required in supporting research projects, and charges would be expected to apply.

WHAT DATA ITEMS WOULD NOT BE AVAILABLE?

Consistent with the statistics legislation, any data likely to identify an individual would not be available. The current ABS confidentiality practices and procedures would be applied to any output from the SLCD.

No unit record datasets created from the SLCD, or by bringing together the SLCD with other ABS or non-ABS data sources, would be provided outside the ABS. Access to such datasets would be through the ABS data laboratory. Only statistical outputs using the information from these datasets would be available outside the ABS. These outputs would also be checked to ensure that no data likely to enable the identification of an individual left the ABS.

SECTION 8 DATA QUALITY

As with any statistical collection there are a number of quality considerations associated with the data brought together under this proposal. The quality of the Statistical Longitudinal Census Dataset (SLCD) is primarily dependent on the underlying quality of the data collected in each census, and when the SLCD is used in conjunction with other datasets, the underlying quality of those datasets also needs to be considered.

This section outlines some of the main issues relating to the quality of SLCD, namely the coverage of the SLCD and quality limitations arising from the methodology that would be used to bring the data together.

COVERAGE LIMITATIONS

The main limitation that would impact on the use of the SLCD is that the census is only conducted every five years.

People who move to Australia after the 2006 Census would be included only for subsequent censuses. People who left Australia after the 2006 Census would not have additional data added with subsequent censuses.

Only information collected in the census would be included in the SLCD and changes that occurred between censuses would not be reflected.

ACCURACY OF PROBABILISTIC RECORD LINKAGE

For many cases, the two records brought together using probabilistic record linkage would reflect the data for the same individual. In other cases there might be many plausible linkages and a decision would be made to determine which records to bring together. The sub-populations that fall into the latter category are more likely to have changes in characteristics between censuses, and are often of more interest for analysis.

A further limitation of the record linkage process is that when there is insufficient or missing information for the data items being used to bring the census data together, census records may not be able to be brought together.

Both these factors would impact on the accuracy of analyses done using the SLCD. At the time of bringing the census data together, the ABS would investigate and report on the overall quality of the SLCD. This would assist in users understanding how to use the dataset. It would also feed into the proposed evaluation and review of the proposal outlined in Section 9.

The level of detail available from the 2001 Census, to bring the data together with the 2006 Census, is broader than what will be available from 2006. This would constrain the quality of the dataset combining 2001 and 2006 Census data. Again, these aspects of quality would be reported on.

ACCURACY OF PROBABILISTIC RECORD LINKAGE *continued*

Similar types of limitations would apply for use of other datasets in conjunction with the SLCD. Other datasets have more limited information available that could be used to bring the data together with the SLCD. This would result in more uncertainty in the linkage of records between the datasets and therefore higher levels of statistical error associated with the combined dataset.

Again, it is proposed to conduct quality studies, at the time of processing the 2006 Census as outlined in Section 5, to gain an understanding of the statistical error associated with the statistical techniques to combine each dataset with the SLCD. Further, this aspect of quality would form part of the proposed evaluation and review (Section 9).

SECTION 9 CONSULTATION PROCESS

This discussion paper outlined a proposal by the ABS to enhance the value of the population census. The ABS has not yet made a decision on whether this proposal, or any part of it will proceed.

As stated earlier, users of census data and other interested parties are invited to comment on all aspects of the proposal.

This section outlines the consultation process and details feedback topics included in the submission form. It also states how a decision about this proposal will be made.

SEMINARS

Following the release of this discussion paper, the ABS will hold seminars with groups of interested people in all capital cities. The seminars are open to members of the public, who are encouraged to attend to obtain more information prior to lodging a submission.

These seminars will:

- introduce the proposal
- discuss examples of the benefits of the proposal
- discuss how confidentiality and privacy will continue to be maintained by the ABS
- detail how submissions can be raised and lodged with the ABS.

To find out about the time and location of seminars in your capital city, or to register your interest, please contact the relevant person in the table below.

	Name	*Email*	*Phone*
Sydney	Jill Tomlinson	jill.tomlinson@abs.gov.au	(02) 9268 4123
Melbourne	Maxine McDermott	maxine.mcdermott@abs.gov.au	(03) 9615 7080
Brisbane	Greg McNamara	greg.mcnamara@abs.gov.au	(07) 3222 6155
Adelaide	Monica Moss	monica.moss@abs.gov.au	(08) 8237 7433
Perth	Wavne Rikkers	wavne.rikkers@abs.gov.au	(08) 9360 5385
Hobart	Bill Trethewie	bill.trethewie@abs.gov.au	(03) 6222 5851
Darwin	Michael Bourchier	michael.bourchier@abs.gov.au	(08) 8943 2123
Canberra	Victoria Allen	victoria.allen@abs.gov.au	(02) 6207 0277

SUBMISSION PROCESS

Submissions may be lodged electronically using the questionnaire on the internet at <www.abs.gov.au>.

A hard-copy questionnaire is also available by phoning the National Information and Referral Service on 1300 135 070. Once completed, the hard-copy questionnaire should be sent to Rosalie Butler, Population Census Development and Field Organisation, Australian Bureau of Statistics, Reply Paid 10, Belconnen, ACT, 2616.

The closing date for submissions is 10 June 2005.

FEEDBACK TOPICS

The submission form seeks views on all aspects of the proposal described in this discussion paper.

It asks whether or not each main aspect of the proposal is supported:

- the creation of the Statistical Longitudinal Census Dataset (SLCD) by combining 2006 Census data with data from future censuses (as described in Section 2)
- combining 2001 Census data with data from the 2006 Census (Section 2)
- use of the SLCD with ABS household survey data (Section 3)
- use of the SLCD with other datasets, where statistical techniques would bring the data together (Section 4)
- use of name and address information during the census processing period to bring together 2006 Population Census data with 2006 Agricultural Census data (Section 5)
- use of name and address information during the census processing period to enable quality studies and other statistical studies (Section 5)

The form asks whether any additional privacy procedures and protocols are considered necessary.

For users of statistical data, it calls for a description of the types of projects that could be done using the SLCD.

DECISION ABOUT THIS PROPOSAL

At the conclusion of the submission period, all the submissions will be carefully considered and the proposal will be reviewed in the light of feedback received. In assessing the submissions the ABS will consider:

- both positive and negative public reaction as expressed by the nature of responses
- any reasons for concern
- any suggested ways to improve the proposal.

Information from the submissions will be used to inform the Australian Statistician about the reaction from the Australian public and census data users, in order to make a decision about the proposal. Recommendations from the Privacy Impact Assessment will also be considered.

The Australian Statistician will make an in-principle decision on whether to proceed with the proposal, or any part of the proposal, in late June 2005. People who have lodged submissions will be notified of the decision.

EVALUATION AND REVIEW

The full benefits of the SLCD would not be realised for some years, as more census data is combined and as the dataset gets more widely used. Under this proposal, it is intended that the ABS conduct a review of the management and use of the SLCD. This would include evaluation of the effectiveness of the statistical techniques used to bring the data together.

It is envisaged that this review would take place after the 2011 Census data are incorporated into the SLCD.

APPENDIX 2006 CENSUS TOPICS AND DATA

The following topics are being considered for the 2006 Census.

TOPICS RELATING TO EACH PERSON	Name
Basic Demography	Sex
	Age last birthday (collected through date of birth question)
	Residential status (for people in non-private dwellings)
	Registered marital status
	Address on census night
	Usual residence at census time
	Family relationship
	Internal migration
	Number of children ever born
	Need for assistance indicator
	Unpaid Work
	Religious affiliation
Ethnicity	Australian citizenship
	Country of birth
	Year of arrival in Australia
	Indigenous status
	Main language spoken at home
	Proficiency in spoken English
	Country of birth of parents (Australia/overseas)
	Ancestry
Education	Attendance at an educational institution
	Type of institution being attended
	Highest year of schooling completed
	Level of qualifications
	Field of qualification
	Income (including family and household income)
Labour force	Labour force status
	Occupation
	Industry
	Government/Non-Government Employer Indicator
	Community Development Employment Project Indicator (special Indigenous form only)
	Hours worked
Transport	Workplace address
	Mode of travel to work

TOPICS RELATING TO HOUSEHOLDS AND DWELLINGS	Address on census night
	Family relationship (for people in private dwellings)
	Income (family, household)
Household	Number of motor vehicles garaged
	Number of bedrooms
	Type of tenure (nature of occupancy)
	Housing loan repayments
	Rent
	Landlord type
Dwelling	Dwelling Internet connection
	Structure of private dwelling (classification)
	Location of private dwelling (classification)
	Non-private dwelling type (classification)

GLOSSARY

Birth and death register data Data obtained from the registration of births and deaths. The registration of births and deaths is the responsibility of the individual state and territory Registrars of Births, Deaths and Marriages. As part of the death registration process, information as to the cause of death is supplied by the medical practitioner certifying the death or by a coroner. Other information about the deceased is supplied by a relative or other person acquainted with the deceased, or by an official of the institution where the death occurred. For births, the data are based on the information form completed by the parents of the child.

Census processing period The period of time immediately after the conduct of the Census of Population and Housing during which the census forms are processed to produce statistical outputs. For the 2006 Census, this period is expected to extend from August 2006 to November 2007.

Data laboratory The ABS provides a service for use of unit record data through an access system involving relevant analytic software. Statistical analyses specified by researchers are submitted and run against the dataset to produce statistical outputs. The system prohibits certain types of analyses that would have the potential to provide identifiable data, while allowing most standard analytic tools to be used against the unit data. Non-identifiable outputs of the analyses are provided to the researcher. The unit record data remains within the secure ABS environment.

Dataset A file containing the individual responses from a statistical collection, administrative records or register of information (for example disease register). Datasets are used to generate statistical output.

Identifiable In this publication unit record data is considered identifiable if it is possible to use the data available in the record to identify the specific individual to whom it refers.

Longitudinal dataset A dataset which contains information collected for the same unit over a number of different points in time.

Long-term immigration data Statistical data held by the Department of Immigration and Multicultural and Indigenous Affairs (DIMIA) from the administration of immigration programs. This includes overseas arrivals and departures data, where the period of duration in over 12 months, and visa grant data, including type of visa.

National disease registers Registers containing information about people diagnosed with certain diseases have been established in the states and territories. These registers allow research about the particular disease. To ensure a coordinated approach to use of these registers, the Australian Institute of Health and Welfare (AIHW) administer national disease registers, consolidating the information contained in the state and territory registers.

Currently, national disease registers administered by the AIHW include:

- National Cancer Statistics Clearing House - all new cases of malignant cancer since 1982
- National Diabetes Register - new cases of Type 1 diabetes since 1999 for which consent has been given for inclusion on the Register.

Non-response bias Error introduced to statistics due to non-response to a survey or question, which are derived from a systematic difference between the characteristics of respondents and non-respondents.

Statistical purposes Functions related to the compilation, analysis and dissemination of statistics. Statistical purposes precludes use of a dataset for administrative or client management purposes, where there ia an impact on specified individuals.

Statistical techniques In this publication, statistical techniques refer to the method that would be used to bring together census data over time to form the SLCD, and to bring the SLCD together with other datasets. The proposed method is often referred to as probabilistic record linkage, which involves bringing together data from two different datasets using a number of characteristics such as age/date of birth, sex, geographic region, and country of birth. All possible linkages based on these data items are evaluated. The records for which the linkage is most likely to be correct are brought together. In this proposal, the statistical techniques used do not involve the use of name and address information to bring together data from two different sources.

FOR MORE INFORMATION . . .

INTERNET	**www.abs.gov.au** the ABS web site is the best place to start for access to summary data from our latest publications, information about the ABS, advice about upcoming releases, our catalogue, and Australia Now—a statistical profile.
LIBRARY	A range of ABS publications is available from public and tertiary libraries Australia-wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.
CPI INFOLINE	For current and historical Consumer Price Index data, call 1902 981 074 (call cost 77c per minute).
DIAL-A-STATISTIC	This service now provides only current Consumer Price Index statistics call 1900 986 400 (call cost 77c per minute).

INFORMATION SERVICE

Data already published that can be provided within five minutes will be free of charge. Our information consultants can also help you to access the full range of ABS information—ABS user pays services can be tailored to your needs, time frame and budget. Publications may be purchased. Specialists are on hand to help you with analytical or methodological advice.

PHONE	1300 135 070
EMAIL	client.services@abs.gov.au
FAX	1300 135 211
POST	Client Services, ABS, GPO Box 796, Sydney NSW 2001

WHY NOT SUBSCRIBE?

ABS subscription services provide regular, convenient and prompt deliveries of selected ABS publications and products as they are released. Email delivery of monthly and quarterly publications is available.

PHONE	1300 366 323
EMAIL	subscriptions@abs.gov.au
FAX	(03) 9615 7848
POST	Subscription Services, ABS, GPO Box 2796Y, Melbourne Vic 3001

2206000001063

RRP $0.00