2001 CENSUS: ANCESTRY - DETAILED PAPER
(Census Paper No. 03/01b)

2001 CENSUS: ANCESTRY - DETAILED PAPER
(Census Paper No. 03/01b)

Chris Kunz & Liz Costello

Population Census Evaluation
June 2003

INQUIRIES

For further information about this paper, contact the Assistant Director, Census Evaluation by telephone: (02) 6252 5611 or email: joanne.healey@abs.gov.au

SUMMARY OF FINDINGS

The 2001 Census Papers on Ancestry, this *Ancestry - Detailed Paper (03/01b)* and the separate *Ancestry - First and Second Generation Australians (03/01a)*, evaluated the data quality of the Ancestry question in the 2001 Census. Overall, the quality of Ancestry data has improved over 1986 Census results.

- Nearly 21 out of every 22 persons responded to the Ancestry question in the 2001 Census (see *Section 6.2 Non-response*). The non-response rate for Ancestry was 4.6% (down from 6.8% in 1986). For those who had both parents born in Australia, the level of non-response decreased to 4.1% (from 7.0% in 1986).

- The number of people stating Australian Ancestry increased, from 3.4 million in 1986 (20% of total persons enumerated), to 6.7 million (35.5%) in 2001.

- 22.1% of the population recorded multiple Ancestries, up from 12.6% in 1986 (see *Section 6.3 Multiple Response*). However, while 21.5 million ancestry responses were captured in 2001, a further estimated 1.9 million written responses were ignored. A decision to code only the first two Ancestries encountered (while not stating this on the Census Form nor in the Census Guide), resulted in the loss of an estimated 8.1% of all ancestry responses written on forms. The issue of lost Ancestries was common to both 1986 and 2001 Censuses.

- Question design virtually precluded the prioritisation of multiple responses - except where none of the seven response options listed on the form were considered appropriate. The 8.1% of lost ancestries may have included the most important ones from an individual's perspective (see *Appendix B: The Impact of Lost Ancestries* for the estimated loss by Ancestry).

- The majority of people who identified as indigenous in the 2001 Census claimed Australian ancestry, as opposed to Aboriginal or Torres Strait Islander ancestry. Indigenous Ancestry counts appear to have been significantly affected by the different forms used in 2001. *Section 6.5 Ancestry and Special Indigenous Personal Forms* provides more detail.

- Overall, the quality of 2001 Census Ancestry data is high, and an improvement over 1986. This assessment is based on the recorded improvement in response rates, the introduction of a more comprehensive coding classification, and the increased propensity of individuals to identify multiple Ancestries.

- Recommended improvements for processing an Ancestry question in a future census include: increasing the minimum number of Ancestries coded to four; stating the coding limit on the form; redesigning the response area on the form; increasing coder support during DPC processing; and extending the ASCCEG to include a dual Ancestry classification listing.

CONTENTS

LIST OF TABLES AND FIGURES

1 INTRODUCTION

1.1 About Census Papers

The ABS has a corporate objective to provide for informed and increased use of statistics. This Paper is part of a series produced after each Census by the Australian Bureau of Statistics' Population Census Evaluation team, whose role is to review the data quality of the 5-yearly Census of Population and Housing. The aim of Census Papers is to inform users of issues that have been identified as impacting on the quality of the census data, which they should keep in mind when utilising the data. Analyses such as this are a critical factor in the continuous quality improvement of the Census Program. The ABS welcomes your feedback and suggestions.

1.2 This Paper

The focus of this *Detailed Paper* is Ancestry - a question that has only been asked in two Australian Censuses: in 1986 and 2001. Between those years, significant research, testing and refinement have resulted in the 2001 question content and format shown in *Section 2.3 2001 Census question format.*

This *Detailed Paper* analyses Ancestry data quality in terms of question design, field operations, and processing issues, with a particular focus on areas that underperformed in 2001 and require further improvement. To provide a comparative measure, this paper makes regular statistical references to 1986 data. Together, information from the two snapshots provide an insight into the changing backgrounds of the Australian population. Differences between question phrasing, coding, or the classification structure have been noted.

A complementary Census Paper, titled *2001 Census: Ancestry - First and Second Generation Australians (03/01a)* is also available from the ABS.

1.3 Background to the inclusion of an Ancestry question

1.3.1 The 1986 and 1991 Censuses

A question on each person's Ancestry was asked for the first time in the 1986 Census. This resulted from an investigation in 1984 by the Population Census Ethnicity Committee of the need for data on ethnicity other than Language, Birthplace of Individual, or Birthplace of Parents. The question was designed to identify the person's origin or ancestry, rather than the ethnic group with which that person identified.

The aim of the 1986 question (see *Section 2.4 1986 Census question format*) was to measure the ethnic composition of the population as a whole. Evaluation showed that it was not useful for this purpose as there was a high level of subjectivity and confusion about what the question meant. Very little use was made of the data from the 1986 Census and as a consequence, Ancestry was not included in the 1991 Census. Refer to the ABS publication *Census 86: Data Quality Ancestry* (Cat. no. 2603.0, 1990) for more detail.

1.3.2 The 1996 Census

In the lead up to the 1996 Census, two questions on Ancestry were tested to determine the extent to which the results were compatible with, and augmented, data collected in existing questions. One aspect of the assessment was the degree of compatibility with the 1986 question results. The testing program and results of the August 1993 Census Test for Ancestry can be found on the ABS Website (see www.abs.gov.au/census/ under Working Papers, Census Working Paper 94/4 - Ancestry). The analysis indicated that data quality for both question formats was unacceptable and that the results were difficult to interpret due to:

- an unacceptably high non-response rate for the second question format;
- the proportion of responses for Australian-only ancestry for both question formats was more than two times greater than for 1986 Census results. Analysis of Australian-born respondents with one or both parents born overseas indicated that many of these people, who clearly had an Ancestry that was not Australian, did not either report, or identify as such;
- more people responded in the affirmative to the question format which asked if the Ancestry they *identified with* was different to their country of birth, than those who responded to the question format which asked whether their Ancestry was different from their country of birth; and
- telephone follow-up of respondents revealed that, while most claimed they knew what Ancestry meant they were uncertain how to answer the question on either question format. Some respondents indicated that given the chance they would probably have responded differently from the way in which they had.

As a consequence of the above concerns, it was recommended that a question on Ancestry not be included in the 1996 Census.

1.3.3 The 2001 Census

As a result of user demands, the ABS established a Census Consultative Committee on Ancestry in 1995 to seek user input and to identify user requirements for the data; research international practices, as well as develop and test questions which may provide acceptable and accurate data at a reasonable cost. In 1996, the Consultative Committee recommended that an Ancestry question should be tested along the lines of the 1986 question but with some pre-coded response categories - utilising intelligent character recognition (ICR) technology.

During the 2001 Census consultation process users had indicated the primary population group of relevance for policy purposes consisted of persons born overseas or those who had parents who were born overseas. Analysis of 1986 Census data and data from the Census Testing Program showed that acceptable identification of these groups could be achieved with a combination of an Ancestry question and a question on whether a person's parents were born in Australia or overseas. This information, in conjunction with a person's own birthplace, would provide a good indication of the ethnic background of first and second generation Australians.

Consequently, a question about Ancestry was included in the 2001 Census to enable identification of those groups which cannot be identified adequately through the Census questions on Language, Religion, Birthplace of Individual, Birthplace of Parents and Aboriginal/Torres Strait Islander origin. *Census Paper 02/03 - 2001 Form Design Testing Paper* (on the ABS Website) contains information regarding the testing program for Ancestry and other 2001 variables.

2 QUESTION DESIGN

2.1 Defining Ancestry

Research completed prior to the 1986 Census found that there was common understanding of what the word Ancestry meant. The Population Census Ethnicity Committee, formed to advise the Australian Statistician on, among other matters, ways in which information could be obtained in a census to satisfy unmet requirements for data on ethnicity, found that 'forefathers/forebears', 'our origins', 'family tree', and 'where you came from' were frequently suggested descriptions of Ancestry (*The Measurement of Ethnicity in the Australian Census of Population and Housing*, Cat. no. 2172.0, 1984, p27).

The Information Paper *Census 86: Data Quality Ancestry* (Cat no. 2603.0, 1990) summarised the Ethnicity Committee's analysis of two approaches to measuring ethnicity as the 'self-perceived identification approach' and the 'ancestry approach':

> 'The *self-perceived identification approach* is concerned with establishing the ethnic group with which a person identifies, and is based on the person's current perceptions, irrespective of origin. People could identify with any ethnic group or groups, irrespective of their background. Thus they could identify with an ethnic group through being closely associated with the lifestyle and culture of that group even if they were not of that group.'
>
> 'Under the *ancestry approach*, people would be asked to base their ancestry on the ethnic group from which they and their ancestors had descended. This is irrespective of whether they continue to be associated with the lifestyle or culture of that group.'

The Committee opted for the *ancestry approach*, feeling that the alternative did not satisfy the criteria for inclusion as a question in the 1986 Census.

The August 1993 Census Test (see Section 1.3.2), demonstrated that an element of confusion regarding the definition and application of the term 'ancestry' still remained. Focus Group testing in October 1998 revealed that, although people generally gave the same response to questions including terms 'ancestry' and 'cultural background', discussion arrived at the conclusion that the former was easier to understand and respond to, than the latter.

2.2 Identifying Ancestry

As explained earlier, a question on Ancestry was included in the 2001 Census to help identify the ethnic backgrounds of first and second generation Australians. Responses to the question may have been influenced, however, by the Census Guide, question design, personal perspective, aspirations, or through third party intervention.

2.2.1 Census Guide

Page 7 of the 2001 Census Guide (see *Appendix A: Ancestry-related information in the 2001 Census Household Guide*) outlined how respondents should answer the Ancestry question. 'Count your ancestry back as far as three generations. For example, consider your parents, grandparents and great grandparents.' While feedback from the Census Inquiry Service (CIS) indicated that many people did not read the Census Guide, and therefore missed this instruction, it may have affected the way some people responded to the question.

2.2.2 Other factors affecting the reporting of Ancestry

Some respondents may have been influenced by the presence of the option boxes to select a response that was not fully representative of their ancestry (see *Section 6.3.4 Ancestry multiple response and the List Effect*). Others (immigrants or children of immigrants) may have marked 'Australian' as a statement of their decision or intent to align with their chosen country of residence. Regardless, any response to Ancestry is based on personal perspective, depending on the importance (or otherwise) of a variety of individual and historical characteristics and traits.

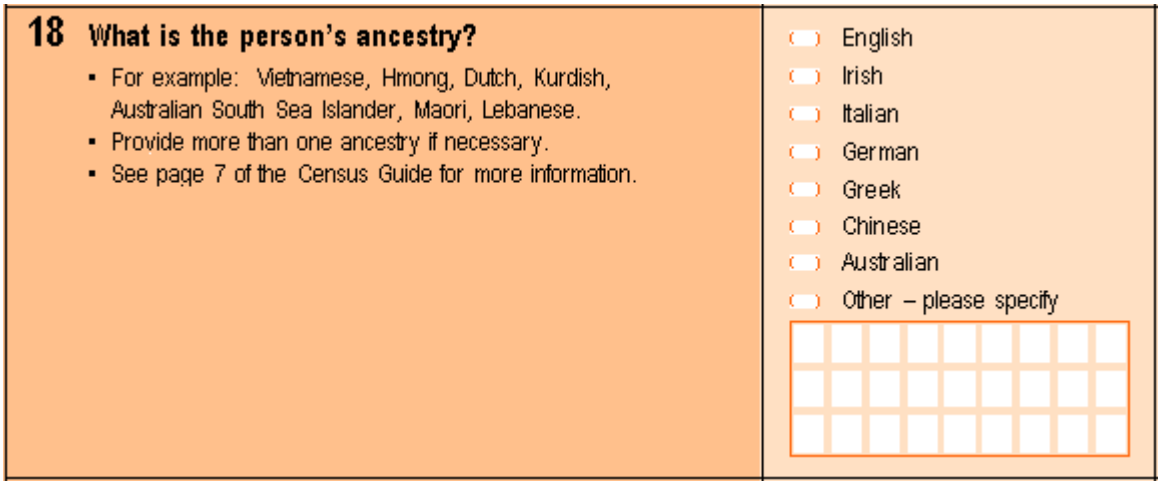
It has been noted that, often, one person completes the Census Form on behalf of (but not necessarily with the active input of) others in the household. For Ancestry, this practice can introduce reporting bias - the responses may be ascribed only partially, or incorrectly, or not what others in that household may have chosen to report for themselves.

2.3 2001 Census question format

2.3.1 Household and Personal Forms

Identical Ancestry questions appeared on the 2001 Census Household and Personal Forms. Instructions on the forms included examples of what could be written, but also offered seven Ancestries in an initial 'mark box' selection range. A Census Guide, handed out with each Census Form, encouraged respondents to mark or write the Ancestries with which there was closest identification, and to go back to their great grandparents (three generations), if known (see Appendix 1 for relevant images of the Census Guide).

FIGURE 1: THE 2001 CENSUS ANCESTRY QUESTION (HOUSEHOLD & PERSONAL FORMS):



The mark boxes facilitated and encouraged a response (and provided some of the most likely responses), while the Write-in section allowed for any other Ancestries to be recorded. Information from the mark boxes was captured by the automated Data Capture (DC) process. Information contained in the Write-in response boxes was automatically coded, where possible, by the Intelligent Character Recognition (ICR) system. If set tolerances for recognition were unable to be met during ICR, the response was passed to a coder for manual assignment to a particular Ancestry code. See *Section 4 Processing Issues* for a fuller description of Census processing procedures.

2.3.2 Special Indigenous Personal Forms

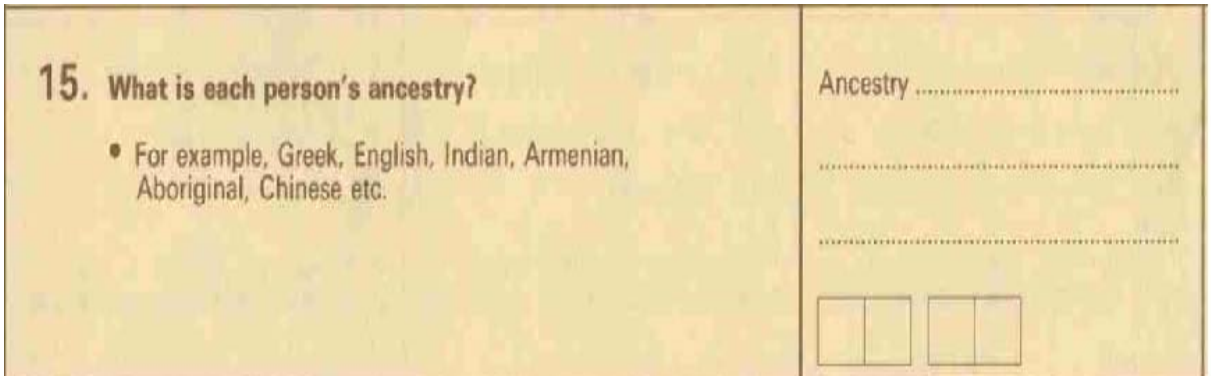
People of Aboriginal or Torres Strait Islander descent living in nominated discrete Indigenous areas in 2001 had their responses to the census recorded on a Special Indigenous Personal Form (SIPF) by an interviewer.

While the examples given on the SIPF were the same as those on the Household and Personal Forms, the range of mark box options was limited to two ('Aboriginal', and 'Torres Strait Islander'), plus an 'Other - please specify' option. The extremely low occurrence of Australian Ancestry amongst those enumerated on Special Indigenous Personal Forms is discussed in *Section 6.5 Ancestry and Special Indigenous Personal Forms*.

2.4 1986 Census question format

Six examples were included on the single form used in 1986 Census, but there were no mark box options for respondents to complete, merely lines for recording Ancestry. As in 2001, only the first two Ancestries were coded, though again there was no mention of this on the form.

FIGURE 2: THE 1986 CENSUS ANCESTRY QUESTION:



A separate Guide was distributed with Census Forms. It stated:

> *'Ancestry' means the ethnic or national group from which you are descended. It is quite acceptable to base your answer on your grandparents' ancestry. Persons of mixed ancestry who do not identify with a single group should answer with their multiple ancestry. Persons who consider their ancestry to be Australian may answer 'Australian'.*

Throughout this paper, comparative data from 1986 is provided to illustrate change. Care should be taken in deducing reasons for any intercensal variation between 1986 and 2001, as the basis of information collection (examples given and mark boxes options in 2001) was different.

2.5 Reduced Country of Birth for Parents in 2001

A factor that the Census has to consider is form size, which impacts on both respondent load and processing costs. This was a major factor in the decision to reduce the Birthplace of Father/Mother options to 'Australia' and 'Overseas' only, while reintroducing the Ancestry question for individuals. Such a combination was considered by the Committee to provide data of sufficient quality.

This decision restricts the further analysis of Ancestry by parents' country of birth in detail. In cases where the Ancestry of an individual has been coded to a generalised grouping, such as Fiji Indian to Indian, 'Overseas' provides no more detail on where the parents originated.

The reintroduction (in future censuses) of specific Country of Birth, for both Father and Mother, would not only provide additional clarification on Ancestry, but also support further cross-analysis of language, income and other Census variables.

Such a change would impact on form design (substituting a Write-in Response Box, or list of options, for the current 'Overseas' mark box for each of Father and Mother). Automated processing in 2001 coded 93.0% of responses to Birthplace of Individual (a nearly identical concept): these efficiencies are now proven. Utilising the classification and coding instructions already in existence for Birthplace of Individual, each parent's country of birth should be able to be automatically coded, at reduced costs and processing time when compared with 1996 (when a Write-in Response Box was last offered for these questions).

3. COLLECTION ISSUES

3.1 Background

The Ancestry question was self-enumerated by 97.3% of people counted in Australia on Census Night. A further 0.4% were enumerated on interviewer-based Special Indigenous Personal Forms. The balance (2.3%) reflects a combination of non-contact with individuals in households and non-private dwellings, and information sourced from administrative records.

During the collection phase of the 2001 Census, collectors reported increased difficulty contacting some householders. Access to secure small and large apartment buildings, gated communities, and growing community concerns about security, make it increasingly difficult to judge whether the residents of a dwelling are absent or not. System Created Records are manufactured during census processing for people for whom a census form has not been received but where the collector believes the dwelling was occupied on census night. System Created Records have values imputed for age, sex, marital status and usual residence only; values for other variables are set to Not Stated or Not Applicable, depending on the imputed value for age.

An increase in non-response (Not Stated) rates is apparent for many census variables in the 2001 Census. Most of the change can be attributed to the increase in the proportion of System Created Records. A Fact Sheet - *Effect of Census Processes on Non-Response Rates and Person Counts* - has been produced and stored on the ABS Website that discusses the factors that may have contributed to the increase in System Created Records for 2001, and the percentage of records affected by state. Please refer to this for further details. An analysis of non-response rates for Ancestry can be found in *Section 6.2 Non-response*.

3.2 Frequently Asked Questions (FAQ)

Staff at the Census Inquiry Service (a telephone service run by the ABS during the delivery and collection phase of the 2001 Census) referred to the following prompts to answer questions from callers:

TABLE 1: FREQUENTLY ASKED QUESTIONS

Why do you need to know my ancestry?	We need to know your ancestry to further the understanding of the origins of Australians. Together with other census information, this information will provide a comprehensive picture of the ethnic background of Australians. This helps in the development of policies and services that better reflect the needs of our diverse society.
What if my ancestries are not listed on the form?	You should mark, 'Other-please specify' and write your ancestry(s) in the boxes provided.
How far back do I have to go back to?	You should count your ancestry back as far as three generations, if known. You should consider your parents, grandparents and great grandparents when answering this question.
Which side of the family do I count back to, my father's side or my mother's side?	You should count back your ancestries on both your father's and mother's side, if known.
What if I only know the ancestry of one parent?	You should mark your ancestry(s) known for one parent.
What if I am unsure of both parents ancestries?	You should leave this question blank or write 'Not known' in the 'Other - please specify' boxes.
What if I am adopted?	If you were adopted, you should answer for your natural parents, if

	known. If not known, leave this question blank or write 'Not known' in the 'Other - please specify' boxes.
What if I have more than one ancestry?	You should provide more than one ancestry, however, when answering this question you should consider and mark the ancestries with which you most closely identify.
What if I am of South Sea Islander descent?	If you are a descendant of the South Sea Islanders brought to Australia as indentured labour around the turn of the twentieth century, you should write in 'AUSTRALIAN SOUTH SEA ISLANDER' in the 'Other - please specify' boxes
Why ask questions about my ancestry and the birthplace of my parents?	Questions on ancestry and the birthplace of parents have been included to further the understanding of the origins of Australians. Together with other census information, this information will provide a comprehensive picture of the ethnic background of Australians. This will assist in developing appropriate policies and services that reflect the needs of first and second generation Australians.

3.3 Ethnic Enumeration Strategy

The principle aim of the Ethnic Enumeration Strategy (EES) developed by the ABS for the 2001 Census was to gain support for, and facilitate and encourage participation in the Census, from the many community groups and nationalities that exist throughout Australia, in order to effectively count Australia's ethnic population. This involved:

- identifying ethnic groups, especially those likely to be missed or undercounted;
- raising awareness and encouraging cooperation by explaining the purpose of the Census to ethnic communities; and
- providing appropriate assistance to those people who needed it, particularly people from ethnic groups that had been undercounted in previous censuses and people who were likely to have difficulty understanding or needed assistance in completing a census form.

Identified groups were specifically informed about the existence and value of the Ancestry question in the 2001 Census, through open meetings, and via a range of promotional and information flyers distributed to ethnic communities. A language hotline was established to provide Census assistance in 20 major languages, running for the duration of the delivery/collection period.

4. PROCESSING ISSUES

4.1 Description of Coding Procedures

2001 Census forms were processed at the ABS's Data Processing Centre (DPC) in Sydney. After receipt, a scanned image of each form was taken, which was used for all further processing.

Ancestry responses were almost exclusively coded as a result of one of three successive processing procedures: Data Capture (DC), Automatic Coding (AC) and Computer Assisted Coding (CAC). The remaining 2% were coded through manual intervention by Validation staff.

Three quarters of all Ancestry responses were coded from the mark boxes. This does not mean that 75% of all Ancestry responses were made within those boxes. The fact that only the first two Ancestry responses were coded (see *5.4 Lost Ancestries*) biases the representation of mark box responses overall.

The coding breakdown was as follows:

TABLE 2: METHOD OF ANCESTRY CODING, 2001 CENSUS

Ancestry Coded Via:	*%*
Data Capture (DC)	75.1
Automatic Coding (AC)	14.8
Computer-Assisted Coding (CAC)	8.1
Other	2.0

4.1.1 Data Capture (DC)

Data capture is the process of scanning the forms into the image and text files that are used for all subsequent processes. At this stage, mark box responses are captured and coded, and text responses are translated into machine readable symbols (through a process that assigns percentages of surety for each individual character) which are examined for their fitness for automatic coding (AC). Where the degree of tolerance was so low that automatic coding was not possible, the field was sent to CAC.

4.1.2 Automatic Coding (AC)

Automatic coding is the process of computer matching the captured text responses to entries on an index for that topic. If no match is made during AC, the response is sent to an operator for computer assisted coding (CAC).

In this second stage of processing, the Automatic Coder attempted to match the textualised ICR version of a response to an entry in the Coding Index. A table of tolerances was created to provide a framework for operation. Using its own in-house developed system, the ABS

was one of the first international statistical agencies to utilise such technology to process Census forms.

The expectation was that around 80% of First Release Processing (which included the simplified response topics like Ancestry) could be coded automatically by the DC or AC processes: in fact a coding match rate of 89.9% for Ancestry was achieved.

While AC significantly reduced processing cost and time, any errors were inevitably systematic (see 'Fiji Indian' in *Section 4.4.4 Discrepancies requiring recoding).* This meant that significant emphasis had to be placed on the subsequent Quality Management (QM) process, to identify systematic errors and adjust tolerances where required. The aim was to reduce the error rate to 1% - considered the level of human error. This was achieved for AC.

4.1.3 Computer-Assisted Coding (CAC)

Computer assisted coding is the process of using procedures and rules to allow a coder to match the image of the text responses to entries on an index for that topic. If no match can be made, the response may be 'dump' coded to a less specific index entry, or to Inadequately Described. The operators also confirm if there is no response to the question for some fields.

Where AC could not definitively decipher a written response, or match it to an entry in the Coding Index, the response was assigned to manual coding. The coder would search the Ancestry Coding Index to select the appropriate match. Responses that couldn't be found on the Index were referred to staff in the Classification Section, who advised on the appropriate code for Index updating (see *Section 4.2 Index Issues*).

4.2 Index Issues

All coding of responses is done by matching to index entries that map to a standard output classification for the topic. Indexes are constantly updated during the processing phase, in response to the types of answers respondents have provided. All additions to the index must be mapped to a category in the standard output classification and are done so with the assistance and approval of the ABS' classifications experts. Index updates are requested by the coders to allow them to better code frequently occurring responses, and by the teams looking at the data throughout processing, such as in response to discrepancy reports.

4.2.1 New and Revised Classifications

4.2.1.1 Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG)

Responses to the 2001 Ancestry question are classified using the Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG). ASCCEG is a classification of cultural and ethnic groups based on the geographic area in which a group originated or developed and the similarity of cultural and ethnic groups in terms of social and cultural characteristics. The Classification aims to classify all claims of association.

Coding rules for the ASCCEG are summarised as follows:

1) Exact matches with ASCCEG Index entries are assigned the code;
2) Spelling differences, abbreviations or idiosyncratic terms in a partial match are assigned the code;
3) Partial matches with qualifying or extraneous words are given the code;
4) If index entry is not matched (as above) or there is reference to a separately identified cultural or ethnic group in the Classification, a Not Elsewhere Classified (n.e.c.) category code or Supplementary Code is assigned. Responses not precise enough to be coded to any category should be coded to 'Inadequately Described.'

ASCCEG was used for the first time in the 2001 Census and cannot be exactly equated with the structure used in the 1986 Census due to geopolitical changes. The ABS publication *Australian Standard Classification of Cultural and Ethnic Groups* (Cat. No.1249.0) released in October 2000 is available on the ABS Website at www.abs.gov.au. For more detail on this classification, see the ASCCEG publication.

An examination of a sample of 2001 coding results revealed that classification of Ancestry responses was straightforward for the overwhelming majority of cases. However, occurrences of possible inconsistencies, through to definitive errors in adjudication, were also noted, and the results fed back to the classification's authors for appraisal.

4.2.1.2 Standard Australian Classification of Countries (SACC)

Ancestry data is often cross-classified with other variables, such as Birthplace.

In the 1996 Census, Birthplace data was classified using the Australian Standard Classification of Countries for Social Statistics (ASCCSS). For 2001, this was replaced by the Standard Australian Classification of Countries (SACC). Although there is no electronic concordance that links ASCCSS and SACC, in most cases it is possible to recompile 2001 Birthplace to the previous classification at a country level.

See the ABS publication, *Standard Australian Classification of Countries (SACC), Rev 2.01* (Cat. no. 1269.0), released in December 1999, which is available from the ABS Website (www.abs.gov.au).

4.2.2 Coding of Dual Ancestry

Irish Australian and Italian Australian are examples of dual responses that were coded to independent elements in the ASCCEG Index.

In its Coding Procedures outline on page 15, the ASCCEG states:

'Many people do not identify with a single cultural or ethnic group only, and will give multiple responses to a question on ancestry, ethnicity or cultural identity. Often a response indicates an identification with a country in a national or cultural sense and also acknowledges continuing ties with other ethnic or cultural groups. Such responses include:

Irish Australian, Italian Australian etc. These responses should be assigned codes for both categories they relate to.'

This dual coding philosophy was not followed in some cases. 'Austro Hungarian' and 'Franco Mauritian', for example, were coded to 'Inadequately Described', and 'Mauritian', respectively. Both determinations are inappropriate, as sufficient information was supplied to identify the ancestries of importance to the respondents. In the former case, Classification Adjudication had stated that Austro-Hungarian was not a sufficient description for coding and in the latter, that what the respondents wanted to state was just 'Mauritian'.

A recommendation to improve the coding of dual ancestry in future censuses has been raised.

4.2.3 Inadequately Described, and Not Further Defined, categories

The Ancestry classification used in 1986 had 94 categories, plus 'Inadequately Described', 'Mixed', and 'Other' (the balance of all other identifiable Ancestries beyond the 94 coding categories). For 2001, an attempt was made to code each response to one of the 189 Cultural and Ethnic group codes in the ASCCEG. Examples include: Aussie to Australian; Myanmar to Burmese; Muscovite to Russian; and Fiji Indian to Indian.

Where coding to a specific Cultural and Ethnic Group code was not possible, the response was coded to one of the Supplementary codes: for example, 0901 'Eurasian, so described'; 2300 'Western European, not further defined'; or 0000 'Inadequately Described'.

It is recognised that the assigning of too many responses to 'Inadequately Described' and other nondescript codes weakens the quality and breadth of Census data as a whole.

4.2.3.1 Inadequately Described

Where a response contains insufficient information to be coded to any level of the classification, it is labeled 'Inadequately Described'. In 1986, a further two categories were employed to represent other recognisable responses that did not fit into the classification in use at that time.

Results for 1986 and 2001 are shown below:

TABLE 3: INADEQUATELY DESCRIBED AND OTHER UNDEFINED RESPONSES TO ANCESTRY, 1986 & 2001 CENSUSES

	1986		*2001*	
Response Type	*Number*	*% of Total Response Population*	*Number*	*% of Total Response Population*
Inadequately Described (a)	14,400	0.1	69,829	0.3
'Mixed'	21,500	0.1	-	-
'Other Ancestry'	116,500	0.7	-	-
Total not defined	**152,400**	**0.9**	**69,829**	**0.3**

(a) Scope changed between Censuses - see paragraph below

Table 3 demonstrates that the overall level of unidentified responses has decreased markedly, though this is due to coding to a larger range of Ancestries in 2001.

The increase in the number of Ancestry responses being coded to Inadequately Described in 2001 can be partially explained as a change in procedure. In 2001, no formal coding query assistance was made available to coders: if neither the coder nor supervisor could ascertain an appropriate code from the Coding Index, the response was dumped to 0000.

While the number of responses coded to Inadequately Described in 2001 rose to nearly five times the level of 1986, the overall rate was still low, equivalent to only one in every 308 responses. With respect to all Ancestry responses recorded, the 2001 coding process overall was a significant improvement on 1986.

4.2.3.2 Not Further Defined (nfd**)**

Not Further Defined codes are part of the ASCCEG Supplementary Code range and are designed to capture generalised responses that cannot be coded to a lower level. They exist at two and one levels above the Cultural and Ethnic Group classification, at the Broad (e.g. 'North African and Middle Eastern') and the Narrow (e.g. 'Arab') level. Information coded to 'not further defined' also impacts negatively on the quality of census data.

TABLE 4: 'NOT FURTHER DEFINED' BROAD AND NARROW GROUPS, 2001 CENSUS

Ancestry Response	*Number of responses (a)*	*% of grouping*
1000 Oceanian, nfd	8,879	0.1
1100 Australian Peoples, nfd	0	0.0
1200 New Zealand Peoples, nfd	16	0.0
1300 Melanesian and Papuan, nfd	429	3.8
1400 Micronesian, nfd	80	8.7
1500 Polynesian, nfd	1,828	2.5
2000 North-West European, nfd	3,794	0.0
2100 British, nfd	11,760	0.2
2300 Western European, nfd	67	0.0
2400 Northern European, nfd	1,992	2.0
3000 Southern and Eastern European, nfd	2,127	0.1
3100 Southern European, nfd	30	0.0
3200 South Eastern European, nfd	9,799	1.3
3300 Eastern European, nfd	8,227	2.2
4000 North African and Middle Eastern, nfd	1,138	0.3
4100 Arab, nfd	15,840	6.6
4900 Other North African and Middle Eastern, nfd	0	0.0
5000 South-East Asian, nfd	44	0.0
5100 Mainland South-East Asian, nfd	0	0.0
5200 Maritime South-East Asian, nfd	0	0.0
6000 North-East Asian, nfd	3	0.0
6100 Chinese Asian, nfd	3	0.0
6900 Other North-East Asian, nfd	5	0.0
7000 Southern and Central Asian, nfd	6	0.0
7100 Southern Asian, nfd	3,082	1.1
7200 Central Asian, nfd	22	0.1
8000 People of the Americas, nfd	488	0.3
8100 North American, nfd	9	0.0
8200 South American, nfd	7,903	14.1

8300 Central American, nfd	397	4.0
8400 Caribbean, nfd	1,735	39.5
9000 Sub-Saharan African, nfd	7,570	7.2
9100 Central and West African, nfd	143	2.8
9200 Southern and East African, nfd	298	0.3
Total 'nfd' responses recorded	87,714	

(a) excluding overseas visitors

By comparison, while the concept 'nfd' was not used in 1986 for Ancestry, a small number of categories were used to classify 460,979 generalised responses that contained some indicative information. In the 1986 Census, 337,879 ancestries were classified as 'British, so described', a further 98,139 ancestries were coded as 'Other Brit incl Anglo Saxon', while 24,943 ancestries were coded to 'Arab'.

4.3 Edits

The ABS Census program has a minimalist editing approach, with most data output as reported on census forms. However, editing is the systematic way of altering data to ensure that it is :

- more complete. For example, if the basic demographic variables of age, sex or usual residence are not stated, they are imputed based on known distributions;
- socially consistent to some extent. For example, age edits do not allow five year olds to be attending high school; and
- consistent with ABS classifications used in other ABS collections. Census labour force status is derived using the same derivation used in the Labour Force Survey, to allow clients to more accurately compare data.

4.3.1 Special Ancestry coding rules

For Ancestry processing, programs and coders were instructed to only code the first two Ancestries on a Census form. *Section 5.4 Lost Ancestries* examines the impact of this rule on Census output.

To remove any duplication where two identical responses were captured (where an Ancestry had been selected in the mark box range and had also been written in the 'Other - please specify' Write-in Response Box section), the second response was set to 'Not Applicable' by an edit.

4.4 Quality Management (QM)

A Quality Management (QM) system was established to identify systematic discrepancies in processing, provide feedback to coders on discrepancies, and produce and analyse discrepancy rates by topics.

4.4.1 The QM Process

Quality Management processing takes a sample of each coder's work, plus samples of codes resulting from data capture and automatic coding, for duplicate coding by a second coder When the original code and second code differ, both outcomes are written to a mismatch file; these mismatches are then recoded for a third time, by an adjudicator, who determines which is the correct code. When the adjudicator determines a code that differs from the original and/or second coder, a discrepancy is recorded for that source; in some cases the adjudicator may determine both are incorrect, and both will have a discrepancy recorded. A report of these discrepancies is fed back to the relevant coder, or process, so that retraining can be done, or systems updates can be made.

4.4.2 Discrepancy Rates

In the majority of cases, the data is not corrected as a result of this sampling: the aim is to improve the coder or process so that such errors do not reoccur. Discrepancy Rates therefore show error rates that are very close to those existing in the final data. However, in extreme cases the production data is recoded - as with the initial coding to Sikh, where there was a systemic problem of a serious nature, and also with 'big ticks' for Chinese/Australian Ancestry (see *Section 4.4.4 Discrepancies requiring recoding*). The discrepancies are also aggregated into the Management Information System (MIS) reports which provide data on the types and frequencies of coding errors over time.

The QM system in place during processing allowed the detection of discrepancies and the calculation of a crude discrepancy rate. This crude discrepancy rate differs from a true discrepancy rate for the following reasons:

- a higher proportion of 'poor' coders' work was included in the quality monitoring sample;
- the Quality Management check coders could make the same mistake as the original coder and therefore an error would not be detected;
- there is not always an absolutely correct code for every response; and
- discrepancies were recorded for any difference in coding between the Quality Management coder and the original coder.

The DPC routinely reviewed between 10% and 50% of automatic and manual coding. This practice was ongoing, though, particularly with a 'human' coder, the percentage chosen for review varied depending on their performance. In this way a measure of quality could be made, and extra training or ongoing support provided if a staff member was having continuing problems. Automatic processes were also continuously monitored.

There would invariably have been errors that coders or systems would have made that were repeated by the QM coders - therefore ensuring that further review of Adjudication never occurred. Such occurrences, however, would have been small - no doubt less than the confirmed Discrepancy Rate. Balancing out this aspect was the greater scrutiny of coders experiencing difficulty.

4.4.3 Discrepancy Rates in final data

Generally, the Discrepancy Rates outlined below can be presumed to be close to the error rate with the finally released data.

Nevertheless, there were also system 'fixes' and retrospective recodes (as with Sikhs) for big ticks that covered Chinese as well as the intended Australian (see *Section 4.4.4,* below). Discrepancy Rates were not recalculated for these.

The final Discrepancy Rates for the coding of first and second Ancestry responses (referred to as ANC1 and ANC2) were:

TABLE 5: OVERALL DISCREPANCY RATES FOR ANC1 & ANC2, 2001 CENSUS

Variable	*Discrepancy Rate*
ANC1	1.4%
ANC2	0.7%

These results were close to the 1% expected error rate had all records been coded manually.

When broken down to each coding process, the figures are:

TABLE 6: DISCREPANCY RATES FOR ANC1 & ANC2 BY CODING PROCESS, 2001 CENSUS

Variable	*Data Capture*	*Automatic Coding*	*Computer Assisted Coding*
ANC1	1.5%	0.7%	1.4%
ANC2	0.5%	0.6%	1.9%

The most significant difference was in the DC process, where ANC1 had more than three times the error rate of ANC2. This variation can be explained by the large 'Not Applicable' component in ANC2 coding - reducing the likelihood of error. The Discrepancy Rates for ANC1 and ANC2 were primarily due to 'big ticks' and 'cross-outs' (see *Section 4.4.4 Discrepancies requiring recoding* below).

Discrepancy Rates recorded for the 8.1% of Ancestries coded by CAC reflect the variances that can occur with increased manual intervention, and the higher proportion of write-in responses coded by the CAC process.

Overall, for all coding processes, the Discrepancy Rate figures averaged around 1% with a potential total of around 378,000 errors in Ancestry coding. An estimated 98% of these remain in the final data**.**

4.4.4 Discrepancies requiring recoding

The following examples show how discrepancy data are used to monitor data quality and determine where reprocessing was required.

(a) A self-described Fiji Indian was found to have been automatically coded to Sikh. This had occurred after the four letters of Fiji were mistakenly 'read' by AC as Sikh. For subsequent occurrences of Fiji Indian, the system continued to automatically assign the incorrect Sikh Ancestry code.

Upon discovery, the data processed to that stage were reviewed, the Coding Index was updated so that this response was no longer able to be coded by AC, and affected records were amended to the correct code (Indian).

Similar problems occurred with South Africa/Sudan Africa, and N. Ireland/Netherlands, for birthplace.

(b) Discrepancy rates, and coder feedback, identified some records for which a big tick (rather than the appropriate dash) in a mark box triggered an unintentional count for the Ancestry listed above it.

FIGURE 3: EXAMPLE OF 'BIG TICK'

DPC staff investigated a sample of 83,644 form images. They found that in that number, there were 102 instances where Chinese and Australian had both been coded.

In 30 of these cases, a big tick for Australian had passed through the coding area for Chinese, leading to an additional Ancestry being coded. In a further eight others, Chinese had been crossed-out and Australian marked, but both had been counted.

TABLE 7: DC CODING ERRORS: CHINESE/AUSTRALIAN, 2001 CENSUS

Option	*No. Dual Coded Chinese/Aust in Error*	*% of Sample Group (83,644)*	*Extrapolated to Australia*
Big Ticks	30	0.04	6,566
'Chinese' crossed-out	8	0.01	1,751

Overall, it was estimated that the number claiming Chinese Ancestry could be inflated by around 1.5%. It was decided to review all coded Chinese/Australian combinations. As a result, over 5,000 records were recoded - though the error is still reflected in the Discrepancy Rate.

The Discrepancy Rate indicates that there were around 206,000 DC errors for ANC1, with all but the recodes for Chinese/Australian likely to be in final data.

(c) Cross-outs, like big ticks, also affected Ancestry coding, and were revealed in ANC1 and ANC2 Discrepancy Rates.

FIGURE 4: EXAMPLE OF 'CROSS-OUT'

Part way through the processing phase, a system edit was created to ensure that English-Australian combinations were reviewed (during the Repair process) to see if English had been crossed out. Where this was the case, English was ignored. This change would have been too late to stop some errors of this type being reflected in the Discrepancy Rates.

5 SAMPLE DATA

5.1 Data Quality Investigation Sample

A 2% statistically derived sample of Collection Districts (CDs) (approximately 760) from each State and Territory in Australia, representing a range of urban and rural CDs; and two smaller samples, focused on Indigenous, and Homeless populations, were identified for 2001. Using these samples, Data Quality Investigation tasks (DQIs) were carried out at the 2001 DPC, directly related to the areas for which in-depth investigations were planned. The resulting data quality information is made available to clients in Census Papers and other related publications, and through analysis provided via the Census query service.

The main DQI sample used for additional analysis comprised:

TABLE 8: THE DQI SAMPLE (a), BY STATE, 2001

State	*Persons in DQI Sample*
New South Wales	122,755
Victoria	90,523
Queensland	68,891
South Australia	27,821
Western Australia	37,372
Tasmania	8,570
Northern Territory	3,372
Australian Capital Territory	7,051
Other Territories	312
Total	**366,667**

(a) enumerated on Household or Personal Forms only. Excludes Overseas Visitors, and SCRs (non contact and administrative record data).

This sub-population was the base used to extrapolate to Australia-wide comparisons in the following sections.

5.2 DQI for Ancestry

Using the sample CDs for the Ancestry topic, the DQI Team collated the number of Ancestries reported by each person, up to six, and the details of Ancestry, Birthplace, Birthplace of Parents and Language for persons who reported more than two Ancestries. The following analyses were then undertaken:

- an investigation of Multiple Marking - where a respondent identified more than one Ancestry;
- the most common Ancestries lost under the Two Ancestries rule; and
- the strength of ethnic identification as measured by Birthplace for those excluded under the Two Ancestries rule.

5.3 The Ancestry Population

In determining the 'Ancestry Population' - a meaningful subset of the number of people counted in Australia on Census Night to be used as a benchmark for Ancestry data quality analysis - a total of 666,253 non-contributing records were excluded. These records are Overseas Visitors (203,101), and System Created Records (SCRs), comprising non-contact (403,729) and admin/other (59,423).

TABLE 9: KEY ANCESTRY-RELATED FIGURES, 2001 CENSUS

Component Details	*Count*
Total Census population	18,972,350
Records that could not respond to Ancestry (SCRs, Overseas Visitors)	666,253
Potential respondents to Ancestry (the 'Ancestry Population')	**18,306,097**

The DQI Multiplier Factor used to extrapolate from the DQI Sample total, to the Ancestry Population, is:

18,306,097 divided by 366,667 = 49.93

5.4 Lost Ancestries (The impact of coding only the first two Ancestries)

The issue that had the most impact on data quality for Ancestry, was the decision to code the first two Ancestry responses only, for each person. This decision was taken to limit processing costs and time and, on the surface, seemed to be logical and consistent with 1986 practice. It was also consistent with the ASCCEG, which recommended (p16):

> 'It is suggested that a minimum of two cultural and ethnic groups be coded if a multiple response is given. This will improve the accuracy and usefulness of the data.'

However, no mention of this limitation was printed on the Census Form or Guide. Therefore, respondents completed the question in good faith, believing that every identification made would be retained and counted. A stated limit could have imposed some prioritisng of responses.

Respondents may have been encouraged into multiple response by the presence of the mark boxes. By the time they had reached the Write-in Boxes, some respondents, unknowingly, were already over the undisclosed limit of two Ancestries.

The following table outlines the frequency of multiple marking for the Ancestry question in the DQI Sample:

TABLE 10: NUMBER AND FREQUENCY OF PEOPLE REPORTING ANCESTRY, based on DQI Sample

Ancestries per Person (A)	*Frequency (B)*	*Percentage*	*Cumulative %*	*Total No. of Ancestries (AxB)*
0	16,920	4.6	4.6	0
1	270,050	73.7	78.3	270,050
2	53,536	14.6	92.9	107,072
3	17,728	4.8	97.7	53,184
4	6,067	1.7	99.3	24,268
5	1,777	0.5	99.8	8,885
6	463	0.1	100.0	2,778
7	92	0.0	100.0	644
8	24	0.0	100.0	192
9	3	0.0	100.0	27
10	6	0.0	100.0	60
12	1	0.0	100.0	12
Total	**366,667**	**100.0**		**467,172**

The sample indicates that 92.9% of people reported between 0 and 2 ancestries, and 7.1% of the sample population had some of their reported ancestries excluded from further processing under the 'code two Ancestries only' rule.

There were 467,172 Ancestries in the DQI Sample, making an average of 1.27 Ancestries per person (when non-responses were included) and 1.34 Ancestries per person for the 349,747 DQI Sample persons who responded.

Using this outcome from the DQI Sample produces the following results for the potential number of respondents in the Census who stated more than two ancestries:

TABLE 11: PERSONS STATING MORE THAN TWO ANCESTRIES (a), based on DQI Sample

Persons stating 3+ Ancestries (DQI Sample)	*Estimated Number of Persons stating 3+ Ancestries (extrapolated to Ancestry Population using DQI Multiplier Factor)*
26,161	1,306,219

(a) Excludes Overseas Visitors, but includes non-responses. If non-responses are excluded, the figure would be 7.5%, or 1,306,742 respondents.

A total of 37,728 DQI Sample responses to the Ancestry question were not captured (8.1% of all Ancestry responses), representing nearly 1.9 million lost ancestries Australia-wide.

TABLE 12: NUMBER OF ANCESTRIES LOST, based on DQI Sample

Number of Ancestries Lost (DQI Sample)	*Estimated Number of Ancestries Lost (using 21,513,178 coded)*
37,728	1,890,044

An analysis of ancestries in the 2001 Census Dress Rehearsal (DR) revealed that 52.8% of all lost Ancestries had been write-in responses. Extrapolated to the estimated number of Ancestries lost in the 2001 Census, it could be surmised that 997,943 of the lost responses were written in, and not from the list of mark-box options on the form.

It should be noted that the DQI Team only coded up to six Ancestries per person. The sample uncovered 126 persons who stated more than six - and even up to 12 - Ancestries. Extrapolated nationally (using the DQI Sample multiplier of 49.93), they would represent 6,291 persons, and 8,916 lost ancestries.

Though Ancestries 7 to 12 are included in the figures in the above tables, none of the DQI tables or extrapolated counts in this paper that involve a breakdown to Ancestry level (including *Appendix B: The Impact of Lost Ancestries*) allow for these extra lost Ancestries.

With around one in every 12 Ancestries lost, it is important to analyse which Ancestries were the most affected:

TABLE 13: FREQUENCY OF LOST ANCESTRIES: TOP 10 (a), based on DQI sample

Ancestry Lost (DQI Sample)	*Frequency Lost in DQI Sample*	*Estimated Frequency Lost: Aust (DQI x 49.93)*
Australian	12,137	606,000
Scottish	6,677	333,383
German	3,524	175,953
French	1,451	72,448
Welsh	1,309	65,358
Dutch	1,045	52,177
Italian	960	47,933
New Zealand	762	38,047
Spanish	679	33,902
Polish	644	32,155

(a) This does not include any Ancestries past the sixth stated for a person (which were not coded in the DQI Sample), nor those coded into Narrow groups (the 00s) in the ASCCEG Classification.

The figures in the above table show that Australian was the most frequently lost Ancestry - making up nearly one third of the estimated 1.9 million Ancestries lost. While this is a significant percentage of all Ancestries lost and the extrapolated total of over 600,000 Australia-wide is large, the actual percentage of Australian lost, at 8.2%, is relatively small.

Further analysis of lost Ancestries has estimated that of those who lost Australian Ancestry, 91.8% had also marked English (and 75.4% Irish). As so many of those losing Australian also selected English and Irish, it is reasonable to presume that the mark box options encouraged their selection.

A much more useful measure of data quality is the percentage of each Ancestry lost. Based on DQI Sample analysis, 31 Ancestries lost over 25% of their count. French was most affected, with nearly half of its write-in responses not captured. In contrast, while significant in count, Australian (8.2%) ranked only 91st (in percentage terms) of all Cultural and Ethnic Groups affected.

TABLE 14: PERCENTAGE OF ANCESTRY LOST: TOP 10 (a), based on DQI Sample

Ancestry Lost (DQI Sample)	*Frequency Lost in DQI Sample*	*Frequency Lost: Aust*	*Lost Aust (%)*
French	1,451	72,448	47.8
Swedish	429	21,420	46.7
Danish	633	31,606	45.0
Welsh	1,309	65,358	43.7
Norwegian	246	12,283	41.5
Scottish	6,677	333,383	38.1
American	442	22,069	33.2
Spanish	679	33,902	31.0
Swiss	199	9,936	30.9
Canadian	158	7,889	28.3

(a) Only shows Ancestries for which the Census count was 10,000 persons or more

A full listing of all Cultural and Ethnic Groups and how they have been affected by Ancestry loss appears in *Appendix B: The Impact of Lost Ancestries*.

Other Census data can be used to cross-classify with those respondents stating three or more Ancestries, to gain some measure of association with lost Ancestries. The percentage of those stating three or more (3+) Ancestries who were born in Australia, and/or had a father or mother born in Australia, assists in this. Comparing Birthplace of an Individual (BPLP), Birthplace of Male Parent (BPMP) and Birthplace of Female Parent (BPFP) presents the following cross-classification:

TABLE 15: THREE OR MORE (3+) ANCESTRIES & BIRTHPLACE (a), based on DQI Sample

% of 3+ with Birthplace in Australia	*% of 3+ with Father born in Australia*	*% of 3+ with Mother born in Australia*	*% of 3+ with both parents born in Australia*
89.6%	75.8%	78.6%	66.1%

(a) Excludes Overseas Visitors

The figures above clearly show that most of those who stated three or more Ancestries were both Australian born and had at least one parent born here.

While two-thirds of those with lost Ancestries were at least third generation Australians, only 32% of Ancestries lost were Australian. This shows that there is a smaller degree of Ancestral Distance - less than 2 - for around one third of those who lost an Ancestry other than Australian.

Respondents claiming lost Ancestries were predominantly born in Australia, as the next table shows:

TABLE 16: TOP 10 BIRTHPLACES OF ANCESTRY LOSERS, based on DQI Sample

Birthplace of 3+	*Frequency in 3+ group*	*Estimated Frequency of 3+ (extrapolated to Australia-wide)*
Australia	23,447	1,170,709
New Zealand	645	32,205
England	390	19,473
USA	193	9,636
South Africa	99	4,943
Canada	70	3,495
Scotland	52	2,596
Malaysia	48	2,397
Papua New Guinea	48	2,397
Philippines	42	2,097

It is also interesting to note that all the top six birthplaces in the table above are countries which have accepted significant numbers of immigrants and refugees in the post World War II period.

The result of coding only the first two Ancestries was that over 8% (close to two million) of all stated Ancestries were never coded - though the percentage never included the first two Ancestries in the mark box listing (when selected), English and Irish, which by virtue of their positioning were always counted. The impact on a large number of mark box, and write-in, Ancestry counts, was therefore severe.

A full listing of all 189 ASCCEG Cultural and Ethnic Groups, their Census Ancestry count, revised Ancestry estimate and percentage of Ancestries lost, as well as a frequency ranking of both the Census count and the revised Ancestry estimate, can be found in *Appendix B: The Impact of Lost Ancestries*.

6. FINAL DATA

6.1 Key Ancestry-related Figures

The following benchmarks from the 2001 Census give some perspective to the analysis in this chapter.

TABLE 17: KEY ANCESTRY-RELATED FIGURES, 2001 CENSUS

Component Details	*Count*	*Calculation*
Total number of persons counted in Australia on Census night (incl Overseas Visitors).	**18,972,350**	
All valid responses to Ancestry recorded	21,513,178	
ANC1s (First Ancestry Responses)	17,469,527	
Multiple responses	4,043,651	21,513,178 - 17,469,527
Multi-response % of Ancestry Population	22.1%	4,043,651 / 18,306,097
Multi-response % of Respondents to Ancestry (ANC1)	23.2%	4,043,651 / 17,469,527
Non-response (Not Stated)	836,570	
Non-response Rate	4.6%	836,570 / 18,306,097

6.2 Non-response

6.2.1 Non-response Rates

Non-response to Ancestry in 2001 was reduced by a third when compared to 1986. In 2001, approximately 836,600 people did not answer the Ancestry question when completing their 2001 form, down from an estimated 1,063,400 in 1986. The addition of mark box options that included Australian, may well have been the most significant factor.

TABLE 18: NON-RESPONSE TO ANCESTRY & RELATED QUESTIONS, 1986 & 2001 CENSUSES

Census Question	*1986 Non-response Rate (%) (a)*	*2001 Non-response Rate (%) (b)*
Ancestry (ANCP)	6.8	4.6
Birthplace of Individual (BPLP)	1.6	3.2
Birthplace of Male Parent (BPMP)	3.1	2.1
Birthplace of Female Parent (BPFP)	2.8	3.3
Language Spoken at Home (LANP)	1.8	2.4
Religion (RELP)	11.9	7.5

(a) excludes Overseas Visitors.
(b) excludes Overseas Visitors, non-contact and admin/other records (see Section 5.3)

The non-response rate in 2001 decreased across all states and territories. The Northern Territory and Victoria registered the greatest reductions in non-response (down 3.6, and 3.2 percentage points, respectively). The ACT again recorded the lowest non-response rate.

TABLE 19: NON-RESPONSE TO ANCESTRY BY STATE/TERRITORIES & AUSTRALIA, 1986 & 2001 CENSUSES

%

Non-response by State/Territory	*NSW*	*VIC*	*QLD*	*SA*	*WA*	*TAS*	*NT*	*ACT*	*AUST*
Ancestry 1986	6.6	8.0	6.8	5.7	5.5	7.3	8.0	5.0	6.8
Ancestry 2001	4.5	4.8	4.9	4.1	4.0	5.6	4.4	3.3	4.6 (a)

(a) Includes Other Territories (5.2%)

6.2.2 Characteristics of Non-respondents

While everyone was expected to complete the Ancestry question, there was a general perception that the Ancestry question was designed for those with non-Australian Ancestries. This resulted in a higher non-response rate amongst the Australian born:

TABLE 20: NON-RESPONSE TO ANCESTRY COMPARED TO TOTAL POPULATION, BY BIRTHPLACE, 1986 & 2001 CENSUSES

	1986		*2001*	
Birthplace	*Non-response to Ancestry (%)*	*% of Total Population*	*Non-response to Ancestry (%)*	*% of Total Population*
Australia	7.0	77.6	4.2	74.4
Overseas	1.3	20.8	2.8	22.4
Not Stated	71.2	1.6	26.8	3.2

As can be seen from the table above, the Overseas-born rate of non-response has more than doubled, while those born in Australia have nearly half the non-response rate of their 1986 counterparts. Non-response for those not stating Birthplace, decreased substantially.

The following table shows the top 20 non-response rates to Ancestry for those who stated their Birthplace:

TABLE 21: TOP 20 NON-RESPONSE RATES TO ANCESTRY BY BIRTHPLACE (a), 1986 & 2001 CENSUSES

	1986			*2001*		
	Birthplace (BPLP)	*Freq (BPLP)*	*Non-response Rate (%)*	*Birthplace (BPLP)*	*Freq (BPLP)*	*Non-response Rate (%)*
1	Norfolk Island	719	39.2	Kyrgyz Republic	102	11.8
2	Australia	12,110,456	7.0	Samoa	13,206	8.3
3	Tonga	4,474	3.4	Somalia	3,711	7.7
4	Cook Islands	1,456	3.3	Tonga	7,656	6.7
5	PNG	21,352	3.2	Ethiopia	3,540	6.5
6	Argentina	9,195	3.1	Paraguay	312	6.4
7	Western Samoa	2,983	3.0	Eritrea	1,599	6.0
8	New Caledonia	1,180	2.9	Seychelles	2,447	5.9
9	El Salvador	2,103	2.9	Norfolk Island	199	5.5
10	Brazilian	2,006	2.8	Tunisia	417	5.5
11	Albania	1,130	2.7	Cook Islands	4,733	5.5
12	Israel	185	2.6	Niue	494	5.5
13	Lithuania	5,346	2.4	Costa Rica	297	5.4
14	Mexico	678	2.4	Moldova	477	5.2
15	Mauritius	13,087	2.4	American Samoa	153	5.2
16	Portugal	14,912	2.3	Albania	1,440	4.9
17	Romania	8,117	2.3	El Salvador	9,689	4.7
18	Chile	18,740	2.3	Malta	46,971	4.7
19	Nauru	536	2.2	Tokelau	262	4.6
20	South Africa	37,061	2.2	Nicaragua	699	4.6

(a) Only accounts for countries with Birthplace of Individual frequency of 100 or more. Others in 2001, such as Mauritania, Marshall Islands, Gabon, St Kitts & Nevis, Cape Verde and Turkmenistan all had higher non-response rates but a frequency of less than 100.

Australian-born non-response in 2001 was 4.2% (making Australia the 52nd highest non-response by Birthplace). Pacific island nations (excluding Norfolk Island), made up six of the top 20, though New Zealand placed 73rd with only 3.6%.

A similar improvement can be seen for those of Australian parentage:

TABLE 22: NON-RESPONSE TO ANCESTRY BY BIRTHPLACE OF PARENTS, 1986 & 2001 CENSUSES

	1986		*2001*	
Parents' Birthplace	*Non-response to Ancestry (%)*	*% of Total Population*	*Non-response to Ancestry (%)*	*% of Total Population*
Both Australian-born	7.0	58.5	4.1	55.1
Father Australian-born (a)	5.6	4.0	5.4	5.7
Mother Australian-born (a)	5.9	7.1	4.5	7.3
Both parents born Overseas (b)	2.0	28.0	3.0	30.5
Both parents Birthplace Not Stated	63.8	2.4	52.3	1.5
Total		**100.0**		**100.0**

(a) Other parent born Overseas or Not Stated (b) Or one parent born overseas and other not stated

Non-response for all groups (except those with both parents born overseas) dropped for 2001. The fact that those with both parents born in Australia, fell the most (from 7.0% to 4.1%) and that when both were overseas-born, it rose, strongly suggests that the inclusion of Australian Ancestry as an option on the Census Form triggered these non-response changes.

The reduction in non-response for those of Australian Ancestry is reflected in the lower non-response for those with English Only, as a language, as shown in the next table:

TABLE 23: NON-RESPONSE TO ANCESTRY BY LANGUAGE SPOKEN AT HOME, 1986 & 2001 CENSUSES

	1986		*2001*	
Language Spoken At Home	*Non-response to Ancestry (%)*	*% of Total Population*	*Non-response to Ancestry (%)*	*% of Total Population*
English only	6.1	84.1	3.9	82.0
Other language	2.6	14.0	3.2	15.6
Not Stated	74.4	2.0	36.6	2.4
Total		**100**		**100**

Consistent with results from Table 22, those speaking an Other Language at home, were the only group to rise in non-response. English Only, containing most of those with Australian Ancestry, fell from 6.1% non-response to 3.9%.

There is little difference between the sexes when answering the Ancestry question, although males were slightly more likely to answer than females, a change from 1986.

TABLE 24: NON-RESPONSE TO ANCESTRY, BY SEX, 1986 & 2001 CENSUSES

	1986		*2001*	
Response Type	*Male (%)*	*Female (%)*	*Male (%)*	*Female (%)*
Non-response	6.9	6.8	4.5	4.6

The general drop in non-response by sex was naturally expected with the drop in non-response overall.

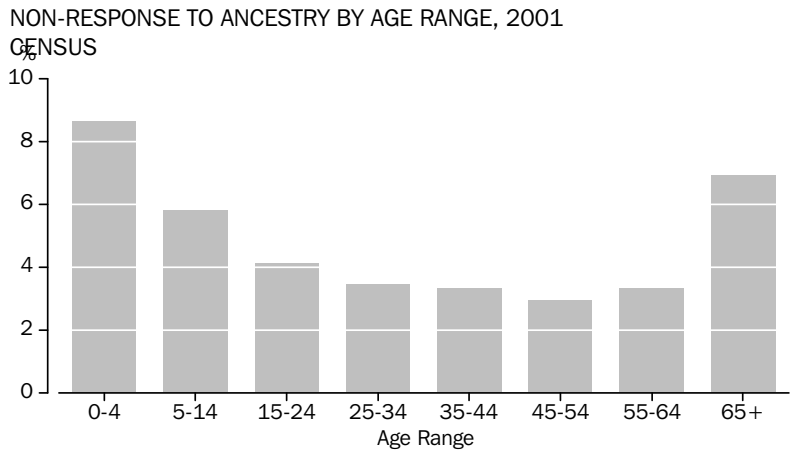
An examination of non-response to Ancestry by age shows the following distribution:

TABLE 25: NON-RESPONSE TO ANCESTRY BY AGE, 2001 CENSUS

Age Range	*Total All*	*Non-response*	*Non-response %*
0 - 4	1,213,588	105,169	8.7
5 - 14	2,606,743	151,435	5.8
15 - 24	2,497,398	103,242	4.1
25 - 34	2,644,604	91,815	3.5
35 - 44	2,794,036	92,909	3.3
45 - 54	2,509,109	74,092	3.0
55 - 64	1,721,372	57,467	3.3
65 & Over	2,319,247	160,441	6.9
Total	**18,306,097**	**836,570**	**4.6**

The difference in non-response rate between age ranges is more clearly shown in the diagram below:

FIGURE 5: NON-RESPONSE TO ANCESTRY BY AGE RANGE, 2001 CENSUS



Non response for many Census questions is typically higher for infant children than for adults; testing has identified that some parents consider many Census questions, or the Census itself, irrelevant to their infants.

It is possible that the larger non-response for persons aged 65 and over is due to a lack of carer knowledge or any relevant administrative records.

6.3 Multiple Response

Overall, 22.1% of the total Ancestry population indicated multiple Ancestry. When non-respondents (836,570) are excluded, of those who responded, 23.1% indicated more than one Ancestry.

The increase on the 1986 multiple ancestry rate (12.6%) can in part be attributed to the seven mark box format that encouraged a range of possibilities (see *Section 6.2.4 Ancestry Multiple Responses and the List Effect)*.

6.3.1 Multiple response by State, Age, and Sex

All States and Territories showed a sharp rise in their percentage of multiple response. The biggest increase was from Tasmania, though it came from the lowest base. At the other end, the ACT recorded the most ‘mixed’ population (29.7%).

TABLE 26: MULTIPLE RESPONSE RATE (a) TO ANCESTRY, BY STATES/TERRITORIES & AUSTRALIA, 1986 & 2001 CENSUSES

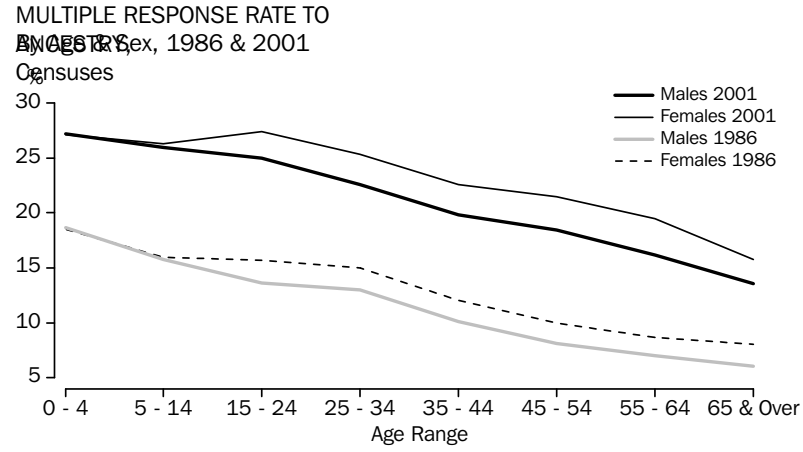
State/Territory (b)	*1986 (%)*	*2001 (%)*
NSW	12.1	20.8
VIC	12.0	20.7
QLD	14.3	24.9
SA	12.4	22.6
WA	12.9	23.3
TAS	9.8	20.2
NT	13.3	20.9
ACT	17.4	29.7
Australia	**12.6**	**22.1**
Australia (No.)	**1,960,400**	**4,043,227**

(a) Persons responding with more than one Ancestry as a percentage of total persons in each category
(b) Excludes Other Territories

As Figure 6 shows, the multiple response rate in 2001 across all age ranges was significantly and consistently higher than in 1986 - by around 10 percentage points. Note that in both Censuses, only the first two Ancestries were coded. Two possible reasons for this increase include:

- the occurrence of multiple ancestries has increased in the past 15 years, or
- people were less inclined to report multiple ancestries in 1986, but more inclined to do so in 2001.

FIGURE 6: MULTIPLE RESPONSE RATE TO ANCESTRY, BY AGE, SEX, 1986 & 2001 CENSUSES



There is a marked similarity between the multiple response rates in 1986 and the same points 15 years later. This generally validates both 1986 and 2001 responses at this broad level.

Of note also in Figure 6 are the consistently higher multiple response rates for females, and the decreasing tendency to report multiple responses as age increases.

6.3.2 Multiple Response by Birthplace, and Birthplace of Parents

Making exact comparisons of Birthplace data from 1986 and 2001 is not possible at all levels due to political and national boundary, as well as classification, changes. However, most

elements are directly comparable within the 2001 classification structure. Refer to Appendix C for full details.

Traditional immigrant nations like the United States (40.3%), Canada (37.5%) and to a lesser extent Argentina (29.0%) and New Zealand (28.5%) displayed the highest multiple response rates in 2001. The lowest multiple ancestry rates were recorded for those born in Greece (1.8%), Italy (1.7%), China (1.5%) and South Korea (1.1%).

Almost 82% of countries for which comparative data is available showed an increase in multiple response. However, reductions were recorded for a number of birthplace groups; such as Cyprus (14.3% to 6.9%), India (21% to 9.8%) and Bangladesh (11.5% to 3.5%).

The pairing of an immigrant parent with a locally-born one is the most likely parental combination to produce Multiple Ancestry Response:

TABLE 27: MULTIPLE RESPONSE TO ANCESTRY, BY BIRTHPLACE OF PARENTS, 1986 & 2001 CENSUSES

	1986		*2001*	
Parents' Birthplace	*Number*	*% (a)*	*Number*	*% (a)*
Both Australian-born	1,072,300	11.8	2,297,723	22.8
Father Australian-born (b)	186,300	29.6	441,001	42.5
Mother Australian-born (b)	335,000	30.1	625,130	46.9
Both parents born Overseas (c)	357,100	8.2	646,169	11.8
Both parents Birthplace Not Stated	9,500	2.5	18,888	7.01
Total	**1,960,200**	**12.6**	**4,043,651**	**22.1**

(a) Persons in each category responding with more than one Ancestry as a percentage of all persons in that Parents' Birthplace category. (b) Other parent born Overseas or Not Stated. (c) Or one parent born Overseas and other Not Stated.

6.3.3 Propensity to report multiple Ancestries

Table 28 presents a subset of the Ancestries most often combined with at least one other Ancestry, as well as identifying those Ancestries least found in Ancestry combinations.

TABLE 28: PERSONS RESPONDING: PERCENT GIVING MULTIPLE RESPONSE (AND FREQUENCY), BY ANCESTRY, TOP & BOTTOM 30 (a), 2001 CENSUS

	Top 30			*Bottom 30*		
	Ancestry	*Multiple Response Frequency*	*% Multiple Response 2001 (b)*	*Ancestry*	*Multiple Response Frequency*	*% Multiple Response 2001 (b)*
1	Irish	1,456,032	75.8	Jordanian	522	19.4
2	Native North American Indian	1,293	69.7	Armenian	2,659	18.1
3	German	506,990	68.3	Coptic	605	18.1
4	French	51,094	64.6	Sinhalese	10,333	17.6
5	Jamaican	764	65.4	Sudanese	667	17.6
6	Swedish	15,773	64.6	Indian	26,826	17.1
7	Canadian	12,868	64.3	Pakistani	2,150	17.0
8	Norwegian	10,842	62.7	Iraqi	1,893	16.9
9	Welsh	52,800	62.7	Australian Aboriginal	14,763	15.6
10	Danish	23,987	62.1	Chinese	82,242	14.8
11	American	26,192	59.2	Khmer	3,129	14.7
12	Scottish	308,466	57.1	Lao	1,351	13.4
13	French Canadian	688	55.2	Iranian	2,270	12.1
14	Mexican	873	53.4	Lebanese	19,476	12.0
15	Niuean	687	52.8	Taiwanese	498	11.3
16	Swiss	11,610	52.4	Kurdish	484	10.8
17	New Zealander	63,924	51.8	Salvadoran	710	10.7
18	Zimbabwean	1,469	50.7	Turkish	5,845	10.7
19	African American	608	50.5	Macedonian	8,406	10.3
20	Austrian	19,147	50.2	Nepalese	268	9.1
21	Papua New Guinean	4,684	49.6	Assyrian/ Chaldean	1,689	9.1
22	Malay	8,793	48.1	Ethiopian	266	8.7
23	Spanish	35,157	46.7	Bosnian	1,514	8.4
24	Estonian	3,523	46.7	Afghan	872	7.0
25	Argentinian	2,922	45.1	Eritrean	139	6.8
26	Maori	32,563	44.6	Vietnamese	9,429	6.0
27	Afrikaner	731	44.4	Bengali	500	4.2
28	Lithuanian	5,458	44.3	Somali	215	4.3
29	Russian	25,955	43.1	Korean	1,410	3.2
30	Dutch	114,832	42.7	Hmong	51	2.8

(a) Includes only Ancestries with total counts of 1,000 or more. (b) Persons in each Ancestry category responding with more than one Ancestry as a percentage of all persons in that Ancestry category.

Irish is the most common in having at least one other Ancestry (75.8%) associated with it; English, at 41.8%, ranked 31st. Australian (at 24.3%) was below the average of 37.6%.

The high Irish multiple rate may have been influenced by its prominence (listed second) in the mark box listing. English was mentioned as an example in the 1986 question, so its first placing in the options now would have given little or no extra advantage. The lower English Ancestry count for 2001 confirms this.

While it may be presumed from these figures that much of the growth in Irish since 1986 is due to Australians claiming an Irish past, the figures below don't support this:

TABLE 29: PERSONS RESPONDING: PERCENT INCLUDING AUSTRALIAN, BY ANCESTRY (a), TOP & BOTTOM 30, 2001 CENSUS

	Top 30				*Bottom 30*			
	Ancestry	*Aust ANC1 (Freq)*	*Aust ANC2 (Freq)*	*% Aust Ancestry 2001 (b)*	*Ancestry*	*Aust ANC1 (Freq)*	*Aust ANC2 (Freq)*	*% Aust Ancestry 2001 (b)*
1	American	11,243	257	26.0	Chinese	842	16,581	3.1
2	Canadian	4,935	126	25.3	Syrian	269	27	2.9
3	Papua New Guinean	1,972	113	22.1	Nepalese	79	4	2.8
4	African American	247	18	22.0	Palestinian	186	10	2.8
5	New Zealander	26,368	365	21.7	Sikh	29	0	2.6
6	Dutch	42,503	1,566	16.4	Turkish	1,167	109	2.3
7	Aust. South Sea Islander	535	3	15.6	Armenian	306	32	2.3
8	Swedish	3,532	87	14.8	Coptic	77	0	2.3
9	Norwegian	2,485	63	14.7	Afghan	249	36	2.3
10	Scottish	76,817	1,766	14.6	Timorese	115	8	2.2
11	Jamaican	164	3	14.3	Macedonian	1,548	113	2.0
12	English	5,059	895,618	14.2	Iranian	327	32	1.9
13	Estonian	1,012	15	13.6	Ethiopian	54	3	1.9
14	Welsh	11,166	289	13.6	Sudanese	63	3	1.7
15	Danish	5,033	129	13.4	Salvadorian	100	3	1.6
16	Finnish	2,246	92	12.9	Lao	149	5	1.5
17	Native North American Indian	233	3	12.7	Bengali	133	7	1.5
18	Swiss	2,690	118	12.7	Punjabi	27	5	1.4
19	Latvian	2,306	65	12.5	Bosnian	224	29	1.4
20	Austrian	4,230	114	11.4	Taiwanese	57	3	1.4
21	Mexican	173	10	11.2	Khmer	257	28	1.3
22	German	1,990	80,858	11.2	Eritrean	22	5	1.3
23	French Canadian	138	0	11.1	Somali	56	7	1.3
24	Kenyan	118	4	10.9	Iraqi	124	6	1.2
25	Maltese	13,593	628	10.4	Kurdish	44	8	1.2
26	Malay	1,799	86	10.3	Assyrian/ Chaldean	182	14	1.1
27	Zimbabwean	287	11	10.3	Korean	419	20	1.0
28	Maori	7,037	293	10.1	Vietnamese	1,442	70	1.0
29	Torres Strait Islander	947	18	9.9	Tamil	54	0	0.7
30	Irish	2,651	186,157	9.8	Hmong	9	0	0.5

(a) Persons in each Ancestry category responding with more than one Ancestry as a percentage of all persons in that Ancestry category. (b) Persons in each Ancestry category also stating Australian Ancestry, as a percentage of all persons in that Ancestry category.

Irish, while the leading Ancestry for multiple response, was only 30th with 9.8% (188,808 people) for Australian Ancestry. When Irish as an Ancestry (1,919,727) is cross-tabulated with other key options, the following table results:

TABLE 30: KEY ANCESTRIES WITH IRISH, 2001 CENSUS

Number of Ancestries	*Frequency*	*% of Irish*
English + Irish	1,024,279	53.4
Australian + Irish	188,808	9.8
Scottish + Irish	56,775	3.0
Welsh + Irish	7,002	0.4

As noted in *Section 5.4 Lost Ancestries*, of those who lost Australian Ancestry due to the limitation of capturing only two Ancestries, 75.4% had also marked Irish (and 91.8% English). This indicates that the 9.8% 'Australian + Irish' combination shown above is a severe understatement when lost Ancestries are considered. Of the estimated 606,000 persons having lost Australian as an Ancestry (see *Appendix B: The Impact of Lost Ancestries*), the likely number of these with an Irish connection could have been around 460,000. This would indicate that the true 'Australian + Irish' component would be considerably greater.

6.3.4 Ancestry multiple responses and the List Effect

Where a question offers a list of mark box options for responses, there can be a possible bias in self-coded responses, known as a 'list effect'. To be subject to 'list effect', a question would have to offer a series of mark boxes or examples, while the question subject would ideally not have a single, definitive answer (such as Age), but have multiple possible answers, that could be influenced by 'mark box' presence - such as Ancestry.

For the Census, the benefits of using a list of options include: the ease and cost of processing without manual intervention (saving on overall processing time and money); and for respondents, an indication of the type of response required through the proffering of the most commonly selected options. However, the main drawback is the impact on data quality: a mark box option is an easier way of completing the question than a Write-in Response.

The impact of this format of question design could produce one or more of the following results:

- an increase in response to the top option on the list;
- people may choose a category from the list of response options in preference to one not on the list;
- the response options listed encourage responses different from those which would have been provided without them; or
- the options listed influence respondents to answer in a different way, generally in a following write-in section if applicable.

During the form design and testing phase of the Census program, questions are assessed for any such impact before being approved for use in the final format. For more information, refer to Information Paper *2001 Census of Population and Housing: Nature and Content* (2008.0).

An examination of multiple responses for 1986 and 2001, focusing on those Ancestries listed on the 2001 form reveals the following results:

TABLE 31: SIGNIFICANT ANCESTRIES, PERCENT OF ALL (a), 1986 & 2001 CENSUSES

	1986		*2001*		*% Change*
Ancestry	*Number*	*% of all Ancestries*	*Number*	*% of all Ancestries*	*(In Number) 1986 - 2001*
Mark box options:					
English	6,587,834	40.1	6,358,880	29.6	-3.5
Irish	899,809	5.5	1,919,727	8.9	113.3
Italian	618,910	3.8	800,256	3.7	29.3
German	507,971	3.1	742,212	3.5	46.1
Greek	335,667	2.0	375,703	1.8	11.9
Chinese	197,839	1.2	556,554	2.6	181.3
Australian	3,400,245	20.7	6,739,594	31.3	98.2
The Balance:	3,885,417	23.7	4,020,252	18.7	3.5
Total	**16,433,692**	**100**	**21,513,178**	**100.0**	

(a) 1986 figures exclude Not Stateds and those Usually Resident Overseas, while 2001 figures exclude Not Stateds, Substitute Forms and Overseas Visitors.

The most significant points from the table are: the dramatic change from English predominance (down from 40.1% to 29.6%) of Ancestries stated, towards Australian (up from 20.7% to 31.3%); and the substantial increases in the number of people reporting Chinese, and Irish Ancestries (up 181.3%, and 113.3%, respectively).

Quantifying the contribution of mark boxes to Ancestry count increase is very difficult, due to the factors involved. Using Irish as an example, its placement in the second position of the mark box range undoubtedly helped - there were no such suggestive options in 1986 - though this alone cannot explain the rise (where English fell).

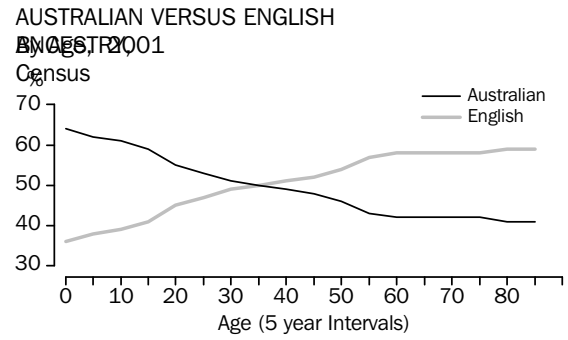
6.3.4.1 English versus Australian Ancestry

Despite any benefit from the 'list effect', English Ancestry, for the first time in Australia's history is not the dominant response. Australians, if they see themselves as anything ancestrally, now are more likely to claim to be Australian than English.

A major factor affecting the Australian count was its inclusion in the mark box selection range. In 1986, most of those who did not respond to the Ancestry question were of Australian birth. The listing of 'Australian' as an acceptable option in 2001 may have reduced confusion (or negative feelings) amongst respondents.

Analysis by age of respondent shows a clear pattern in Australian versus English Ancestry. For persons over 50, around 42% of those choosing Australian or English Ancestry chose Australian. In the 35-39 year age group it was evenly divided, but the Australian proportion increased progressively as age reduced. For the recently-born, it was 64% (see Figure 7).

FIGURE 7: AUSTRALIAN AND ENGLISH ANCESTRY, BY AGE, 2001 CENSUS



6.4. Australian Ancestry

6.4.1 Australian Ancestry by State and Birthplace

The state breakdown for those stating Australian Ancestry is as follows:

TABLE 32: PERSONS RESPONDING 'AUSTRALIAN' TO ANCESTRY QUESTION (a), STATES, TERRITORIES & AUSTRALIA, 1986 and 2001 CENSUS (b)
(%)

	1986		*2001*	
State	*1st Response*	*2nd Response*	*1st Response*	*2nd Response*
NSW	20.5	1.5	31.2	31.6
VIC	19.1	1.4	28.9	30.2
QLD	22.2	1.6	34.9	29.5
SA	20.7	1.7	30.6	35.3
WA	18.5	1.8	28.1	34.0
TAS	21.1	1.3	42.3	40.5
NT	16.3	1.4	30.2	30.0
ACT	22.4	2.1	31.1	32.9
Other Territories	-	-	24.6	20.0
Australia	**21.8**	**1.7**	**31.3**	**31.6**
Australia (No.)	**3,159,717**	**240,528**	**5,462,014**	**1,277,580**

(a) Persons responding with 'Australian' Ancestry as a percentage of the population in each State.
(b) 1986 figures exclude Not Stated, and Those Usually Resident Overseas. 2001 figures exclude Not Stated, and Overseas Visitors.

In 2001, all States and Territories recorded increases in 'Australian' as a 1st Response, though the 2nd Response increase was dramatic. This clearly demonstrates both the increased propensity to record more than one response, as well as an increased awareness of 'Australian' as an acceptable response in 2001. 1st Responses of 'Australian' doubled for Tasmanians, who recorded the highest frequency of Australian for both 1st and 2nd Responses.

Australian Ancestry responses by Birthplace provides the following percentages:

TABLE 33: PERCENT OF POPULATION STATING AUSTRALIAN ANCESTRY AS A RESPONSE: BY BIRTHPLACE & BIRTHPLACE OF PARENTS, 1986 & 2001 CENSUSES

	1986		*2001*	
Birthplace	*Australian as 1st or only Ancestry response*	*Australian as 2nd response*	*Australian as 1st or only Ancestry response*	*Australian as 2nd response*
Of Individual:				
Australia	25.8	1.9	38.5	8.9
Overseas	0.9	0.3	1.5	0.8
Not Stated	1.9	0.2	26.8	4.2
Of Parents:				
Both Australian-born	29.4	0.9	46.3	7.8
Father Australian-born (a)	25.5	6.5	30.9	18.1
Mother Australian-born (a)	20.9	9.4	25.6	20.0
Both parents born Overseas (b)	1.3	0.2	1.6	0.5
Both parents Birthplace Not Stated	3.9	0.2	15.9	2.0

(a) Other parent born Overseas or Not Stated. (b) Or one parent born Overseas and other Not Stated.

Every perspective of Birthplace has featured an increase in Australian response, relative to its 1986 proportion. This was to be expected, with the appearance of Australian in the mark box listing and its significantly increased count.

The low proportion of Overseas-born individuals, or those with overseas-born parents, recording Australian, generally validates the data. See *Section 6.4.2 Aspirational Australian Ancestry* for further examination of this topic.

The fact that the total estimated real count (2001 Census plus DQI Sample results, see *Appendix B: The Impact of Lost Ancestries*) for those acknowledging Australian Ancestry, was 7,345,415 and that this is 59.0% of the 12,457,288 individuals with at least one parent born in Australia, indicates that there is still potential for further increase in the Australian Ancestry component in future censuses.

6.4.2 Aspirational Australian Ancestry

Feedback from Migrant Resource Centres during the 2001 Census Dress Rehearsal (conducted in 2000) indicated that recently arrived refugees or migrants, the prime group that the Ancestry question aims to identify, were sometimes claiming Australian Ancestry as a statement of their desire to be seen from hereon as just 'Australians'.

If consistently adopted, this aspirational interpretation of Ancestry would have serious implications for data quality.

An interrogation of Census data, cross-classifying 'Australian' Ancestry with a person's Year of Arrival minus those who had a father or mother born in Australia, revealed that the numbers were fairly small (table 34). As a proportion of the Australian Ancestry population, 33 in every 10,000 is not statistically significant.

TABLE 34: INDICATING ASPIRATIONAL ANCESTRY, 2001 CENSUS

Australian Ancestry (A)	*'A' but With Year of Arrival (B)*	*'B' but with Mother or Father Born in Australia (C)*	*Minimum Aspirational (D: B - C)*	*Minimum Aspirational % of Australian Ancestry (D/A)*
6,739,594	85,616	63,128	22,488	0.33%

The true Maximum Aspirational count can be calculated by finding those who claimed Australian Ancestry but had neither parent born in this country:

TABLE 35: MAXIMUM ASPIRATIONAL ANCESTRY, 2001 CENSUS

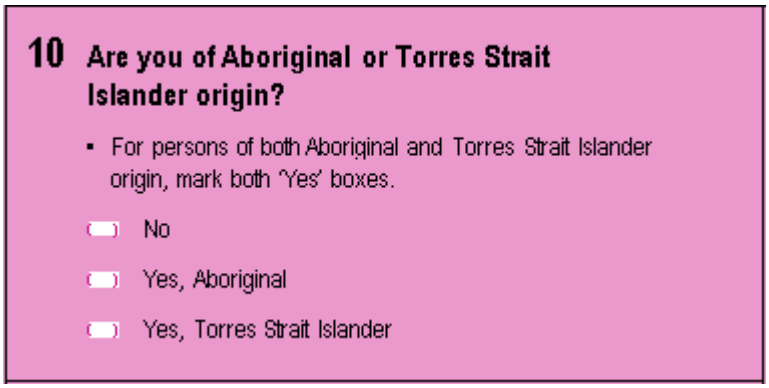
Australian Ancestry (A)	*Non-Aspirational: 'A' but with Mother or Father Born in Australia (B)*	*Maximum Aspirational (C: A-B)*	*Maximum Aspirational % of Australian Ancestry (C/A)*
6,739,594	6,570,736	168,858	2.51%

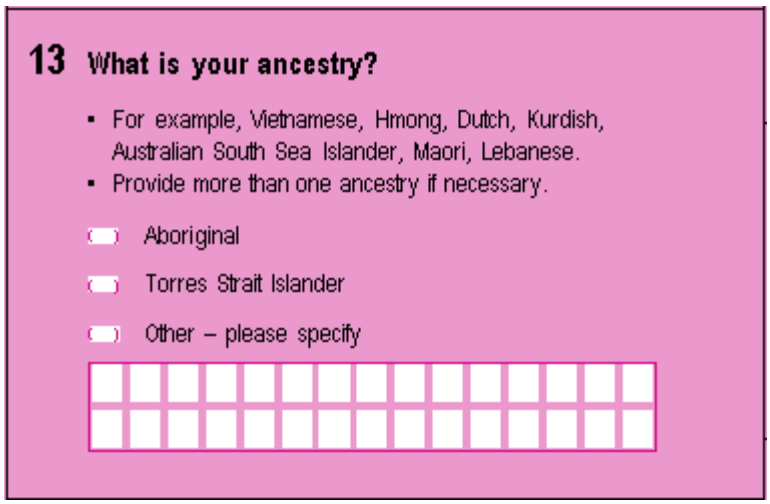
When viewed from the perspective that two and a half in every hundred claimed Australian Ancestry without Australian-born parentage, the Aspirational Australian count can be said to have had very limited impact.

6.5 Ancestry and Special Indigenous Personal Forms

Most people who identified as being of Aboriginal or Torres Strait Islander descent were enumerated on mainstream Household, or Personal Forms. However, if they lived in an identified Indigenous area (usually in remote areas), their responses were recorded by an interviewer on the Special Indigenous Personal Form (SIPF). This Form had a question on Origin (Q10) and then another on Ancestry (Q13).

FIGURE 9: ORIGIN AND ANCESTRY QUESTIONS ON SIPF, 2001 CENSUS





On the mainstream Household and Personal Forms, the Origin question immediately preceded the Ancestry question; the list box options were also different (compare Figure 1 and Figure 9 for details).

Feedback from interviewers indicated that many SIPF respondents thought they were being asked the same question twice.

TABLE 36: RESPONSES TO ORIGIN & ANCESTRY BY FORM TYPE, 2001 CENSUS

		Ancestry					
		Aboriginal		*Torres Strait Islander*		*Australian*	
Identification	*Origin Count*	*Number*	*% of origin*	*Number*	*% of origin*	*Number*	*% of origin*
SIPF:							
Aboriginal	68,087	67,581	*99.3*	162	*0.2*	75	*0.1*
Torres Strait Islander	4,176	88	*2.1*	4,094	*98.0*	10	*0.2*
Both Aboriginal & Torres Strait Islander	1,366	1,272	*93.1*	1,232	*90.2*	3	*0.2*
Household and Personal Forms:							
Aboriginal	293,282	21,225	*7.2*	44	*0.0*	196,561	*67.0*
Torres Strait Islander	21,758	99	*0.5*	3,167	*14.6*	10,777	*49.5*
Both Aboriginal & Torres Strait Islander	16,149	1,218	*7.5*	679	*4.2*	9,565	*59.2*

On the SIP Forms, there was a very high correlation between what was stated for Origin, and Ancestry; extremely few respondents stated that they were of Australian Ancestry.

For Household and Personal Forms, the fact that 67% of those nominating as having Aboriginal origin stated Australian as an Ancestry, and only 7% claimed Aboriginal ancestry, highlights a lack of consistency in results for Indigenous persons overall and contrasts with those of Indigenous origin based in Indigenous communities.

It should be acknowledged that the strength of identification would be expected to be far stronger in discrete Indigenous communities, than in the mainstream population.

Nevertheless, form design and question sequencing, as well as method of enumeration are all contributing factors to the Indigenous count for Ancestry being significantly less than could have been expected.

6.6 Correlation of Non-Australian Ancestry with other census variables

There are three Census variables with which it would be expected that there would be a relatively high correlation to confirm the validity of non-Australian Ancestry data. They are Birthplace of individual, Birthplace of parents, and Language spoken at home.

TABLE 37: CORRELATION WITH NON-AUSTRALIAN ANCESTRY FOR PERSONS BORN OVERSEAS, OR WITH AT LEAST ONE PARENT BORN OVERSEAS, OR LANGUAGE SPOKEN AT HOME OTHER THAN ENGLISH, 2001 CENSUS

Variable	*Total Number*	*Percentage with Non-Australian Ancestry*
Individual born overseas	4,044,349	97.7
At least one parent born overseas	7,656,877	91.5
Language other than English spoken at home	2,852,227	98.3

The above table validates the non-Australian Ancestry results from the 2001 Census. Speaking a language other than English at home at 98.3%, is the most reliable indicator of non-Australian Ancestry.

6.6.1 Correlation for specific Ancestries

Three of the largest Ancestries recorded in the 2001 Census, where English is not the traditional or official language of that Ancestral group, show quite different levels of correlation between Ancestry, Language, and Birthplace:

TABLE 38: LARGE NON-ENGLISH-SPEAKING GROUPS: COMPARISON OF PERSONS IN EACH GROUP BASED ON COMMON ANCESTRY, LANGUAGE & BIRTHPLACE OF INDIVIDUAL, 1986 & 2001 CENSUSES

	1986			*2001*		
Ancestry	*Frequency (a)*	*Language (b)*	*Own Birthplace (c)*	*Frequency (a)*	*Language (b)*	*Own Birthplace (c)*
Italian	620,227	61.5%	40.8%	800,256	39.9%	26.1%
German	510,402	14.3%	16.8%	742,212	7.5%	11.5%
Greek	336,782	78.0%	39.3%	375,703	65.3%	29.1%

(a) Total frequency (ANC1+ANC2). (b) Language Spoken at Home is same as Ancestry (c) Birthplace of Individual equates with Ancestry.

Table 38 shows that in August 2001 Greek was spoken at home by 65.3% of those claiming Greek Ancestry - a far higher percentage than the relative figures for Italian (39.9%) and

German (only 7.5%). Between 1986 and 2001, the proportions changed significantly: this is most pronounced for those of German Ancestry (where Language nearly halved), although Italian and Greek also recorded decreases.

The 2001 German Ancestral count (742,212 people), while seven percent lower than that for Italian Ancestry, was 56% lower in terms of Own Birthplace, and 81% lower in terms of Language Spoken at Home. Conversely, while the 2001 Greek Ancestral count was 53% lower than that for Italian Ancestry, and 49% lower than German, it had higher correlation rates than either Italian or German, for Own Birthplace, and Language Spoken at Home.

7 CONCLUSIONS

1. The error rate associated with the Data Capture (DC) process reflects little improvement on manual coding.

While the error rate was only 1.5% for ANC1 (see Table 6), this still indicated that respondent cross-outs or big ticks led to over 200,000 coding mistakes.

It was difficult to ascertain the net picture of system modifications on data quality and Discrepancy Rates, as reporting of incidences and subsequent actions taken were fragmented and not cumulated to form a clear, final image.

2. The coding of only two Ancestries, combined with the introduction of mark box options, had the greatest negative impact on data quality

An estimated total of 8.1% (1,890,044) of all Ancestries were lost, from about 7.1% (1,306,200) of respondents (*Section 5.4*). While this may not seem statistically significant at the Australia-wide level, potentially all Ancestry data was undercoded, from Australian (606,000 lost), down.

In terms of percentage lost from the formal count (the most appropriate indicator of data quality), Australian was only the 91st worst affected Ancestry. Others, such as French, Swedish, Danish and Welsh - all of which had to be write-in responses - lost close to half their number, rendering the Census count for these groups misleading (see *Appendix B: The Impact of Lost Ancestries*).

The fact that many lost Ancestries were to those born in this country and having at least one parent born here, indicates a degree of Ancestral Distance. This should not be used to excuse the loss of Ancestries, as the mere identification with an Ancestry indicates an attachment, beyond the immigrant generation.

Due to their placement in the mark box listing, both English and Irish maintained close to their maximum Ancestry response. These Ancestries, unless ignored in the mark boxes and then written in below as a third or later Ancestry, were never ‘lost’.

The fact that only two Ancestries were to be coded was not conveyed to respondents, either on the Census Form or in the accompanying Census Guide. This meant that respondents could not prioritise their Ancestries, though some may have attempted to do so in using write-in boxes to record mark box Ancestries. Many of the estimated 997,900 lost write-in responses were due to respondents first being ‘tempted’ into choosing at least two of the mark box options.

3. Non-response was reduced considerably from 1986.

Nearly 21 out of every 22 persons responded to the Ancestry question. The decrease in the non-response rate to 4.6% (from 6.8% in 1986), was largely driven by the increase in response from those with both parents Australian-born (where non-response fell from 7.0% to 4.1%), indicating that the inclusion of Australian as an Ancestry mark box option, increased response.

4. Even excluding those Ancestries lost in 2001, over five million more Ancestries were coded than in 1986.

Factors such as lower non-response rate, increased immigration and population diversity, intermarriage, population growth, social acceptance and the presence of mark boxes, have all contributed to the increase in multiple response from 12.6% in 1986 to 22.1% in the 2001 Census.

5. The concern that respondents may want to claim Australian Ancestry, when this was not justified through birthplace of parent or earlier ancestor, was proved to be largely unfounded.

Only 2.5% of the population claimed Australian Ancestry when neither of their parents were born in Australia. Such a low maximum Aspirational Australian component is heartening for data quality generally, as Australian was the Ancestry at most risk of fabrication or misinterpretation.

It is this latter perspective that leads to the assessment that overwhelmingly, respondents understood what was meant by Ancestry, even though acknowledgment of an Ancestry would not have been just a function of Ancestral Strength.

6. Overall, Indigenous counts for Ancestry were significantly less than totals for the Indigenous Origin question.

While the correlation between Indigenous identification (Origin) and Ancestry on the SIPF was high, on the Household and Personal Forms, where the bulk of respondents claiming Indigenous status were enumerated, the opposite was the case. Here, a majority claimed Australian Ancestry, leading to some doubt over its interpretation (see *Section 6.5 Ancestry and Special Indigenous Personal Forms).*

Given the existence of, and importance placed on, the Indigenous Origin question, Ancestry does not appear to be an appropriate or reliable source for Indigenous counts in 2001.

8 RECOMMENDATIONS

1. *That a minimum of at least four separately-stated Ancestries be coded*.

This step is needed and is logical. This allows for the increasing number of grandchildren of post World War II migrants to count themselves as being of Australian Ancestry (through their parents), and also incorporate possibly varying Ancestry via three of their four grandparents.

Quite clearly, a respondent should not have to choose between their parental Ancestries for a further, single option beyond Australian - as would have to be done under the current two coded Ancestries only policy.

Based on 2001 Census DQI Sample figures, the coding of four Ancestries would have covered 99.3% of respondents. This would make data far more accurate than in the 2001 Census, when only about 93% of respondents had all of their Ancestries coded.

2. *That the maximum number of Ancestries to be coded must be stated on the Census Form, and in the Census Guide.*

3. *That 'Australian' be the only listed Ancestry option (above 'Other, please specify') in the mark box sequence*.

The limiting of mark box options is recommended given the problems encountered in correctly coding respondent markings that have included cross-outs and big ticks. Any prospect of bias towards all but Australian would therefore be removed.

Australian was the most nominated Ancestry in 2001, despite being only seventh in the mark box listing. This fact and the inevitable 'Australianisation' of the grandchildren of immigrants would guarantee it will be the most nominated Ancestry, irrespective of future placement. The recommended change in format will not alter its frequency ranking.

While it is logical that there be a growth in Ancestries reported over time, the removal of multiple options from the mark box listing is likely to more than compensate - leading to an anticipated lower total of reported Ancestries.

Removal of the mark-box next to 'Other - please specify' should be considered. By omitting it, the specific problem of big ticks in the Ancestry question will be removed.

The changes recommended should result in reliable results that reflect significantly improved data quality.

4. *That a supplementary listing of Dual Ancestries be created and then utilised in coding*.

While Dual Ancestries such as Irish-Australian and Italo-Australian have been acknowledged in the ASCCEG introduction, no formal listing exists. A supplementary listing should be created. This would guide coders and help produce a more accurate count.

5. That the DPC provide a query process for Ancestry, as well as ensuring there is Ancestry expertise on site to resolve issues.

6. That each parent's specific Country of Birth be asked for in the 2006 Census.

The extra costs and work would be minimal as the coding and classification systems already exist. This extra information would add considerably to researchers' ability to definitively step beyond the 'Overseas' birthplace tag for parents.

7. That DPC management systems should be able to provide an easily accessible and succinct summary of the cumulated effect of all modifications to processing systems and records, as well as their impact on Discrepancy Rates.

It should then be possible to obtain more detail, if required.

9 OTHER INFORMATION AVAILABLE

Australian Census Analytical Program

Ethnic Diversity, Ethnic Intermixture and the Development of 'Australian Ancestry'

This project proposes a comprehensive analysis of the census Ancestry data to examine a number of issues relating to ethnic diversity, ethnic intermixture and the development of the concept of 'Australian Ancestry'.

The project will use data from the 1986 and 2001 Census, as well as emigration statistics for the years 1986 - 2001 from the Department of Immigration and Multicultural and Indigenous Affairs.

The final report from the project is due in late 2003. Further details may be obtained from the Director, Census Products and Services, by phone (02) 6252 7007.

REFERENCES

Population Census Ethnicity Committee's report *The Measurement of Ethnicity in the Australian Census of Population and Housing* (Cat. No. 2172.0); 1984

Information Paper: *Census 86: Data Quality Ancestry* (Cat. No. 2603.0); 1990

Australian Standard Classification of Cultural and Ethnic Groups (Cat. No. 1249.0); 2000

Standard Australian Classification of Countries (SACC), Rev 2.01 (Cat. No. 1269.0); 1999

Information Paper: *2001 Census of Population and Housing: Nature and Content* (Cat. No. 2008.0)

Fact Sheet: *Effect of Census Processes on Non Response Rates and Person Counts*

Census Paper 02/03 - 2001 Form Design Testing Paper

GLOSSARY

AC - Automatic Coding. The matching of textual responses (as interpreted by ICR) to the Index, without manual intervention.

ANC1 - the first response coded to an Ancestry for a respondent. It is always the first in sequence (mark box before Write-in Response) on the respondent's form i.e. If English and Irish are marked, English will be ANC1 and Irish ANC2. No more than two responses were coded for any respondent.

Ancestor - *n*. any person from whom one's father or mother is descended, forefather.[1] Note that the common interpretation includes one's parent in Ancestry, even if it were for that single generation.

Ancestral - *a.* belonging to or inherited from ancestors. [1]

Ancestral Distance - the degree of removal from the most recent known example of a particular Ancestry e.g. an Ancestral Distance of 'one' to a person's parent, 'two' if a grandparent, etc.

Ancestry - *n*. (lineage of) ancestors**.** [1]

ASCCEG - *Australian Standard Classification of Cultural and Ethnic Groups* (ABS Cat. no. 1249.0). This is the standard classification used for coding Census Ancestry responses.

Aspirational Ancestry - where a person makes a selection based not on their background, but on their desire. Most commonly used by more recent refugees or migrants who desperately want to be seen as Australian (see *Section 6.4.2 Aspirational Australian Ancestry*).

ATSI - Aboriginal and Torres Strait Islander.

Classification - grouping arrangement, often a hierarchy such as the ASCCEG.

Census Guide - an explanatory booklet that provides advice and background information on how to complete a Census Form (see *Appendix A*). A Guide was distributed with each Form.

Census Inquiry Service (CIS) - a phone-based (13 number) facility set up to provide translation and other information services relating to the 2001 Census.

Data Capture (DC) - the process that ensures all marks on the Form (mark box or writing) are reproduced on an image. DC registers and codes mark box responses.

Discrepancy Rate - the rate at which Quality Management and subsequent Adjudication coding differed from that of an individual human or system coding. It is expressed as a percentage and is regarded as the error rate within final data.

DPC - Data Processing Centre for the 2001 Census. A centralised facility which was located in Ultimo, Sydney.

DQI - Data Quality Investigation. A DQI Team operated at the DPC, conducting additional coding exercises to uncover data quality issues.

Dress Rehearsal (DR) - generally the last in a regular series of tests of census field materials and procedures that occurs around a year before Census date. The 2001 DR was conducted on 27 June 2000 and involved a total of 40,097 dwellings in Melbourne and Mildura.

FRP - First Release Processing. Responses to questions that were processed within this first phase included those to Ancestry.

ICR - Intelligent Character Recognition. The system used to interpret handwritten responses in Write-in Boxes and convert them into machine-readable text suitable for AC.

Index - the listing of valid responses to a Census question or topic.

Mark boxes - invite the respondent to place a dash within at least one of a possible series of selection boxes on the Census Form. The ICR system then identified marked boxes during the Data Capture process.

Quality Management - (in this paper) the process of regular review of a percentage of a coding work, though also a term for broader DPC-wide ongoing reviews.

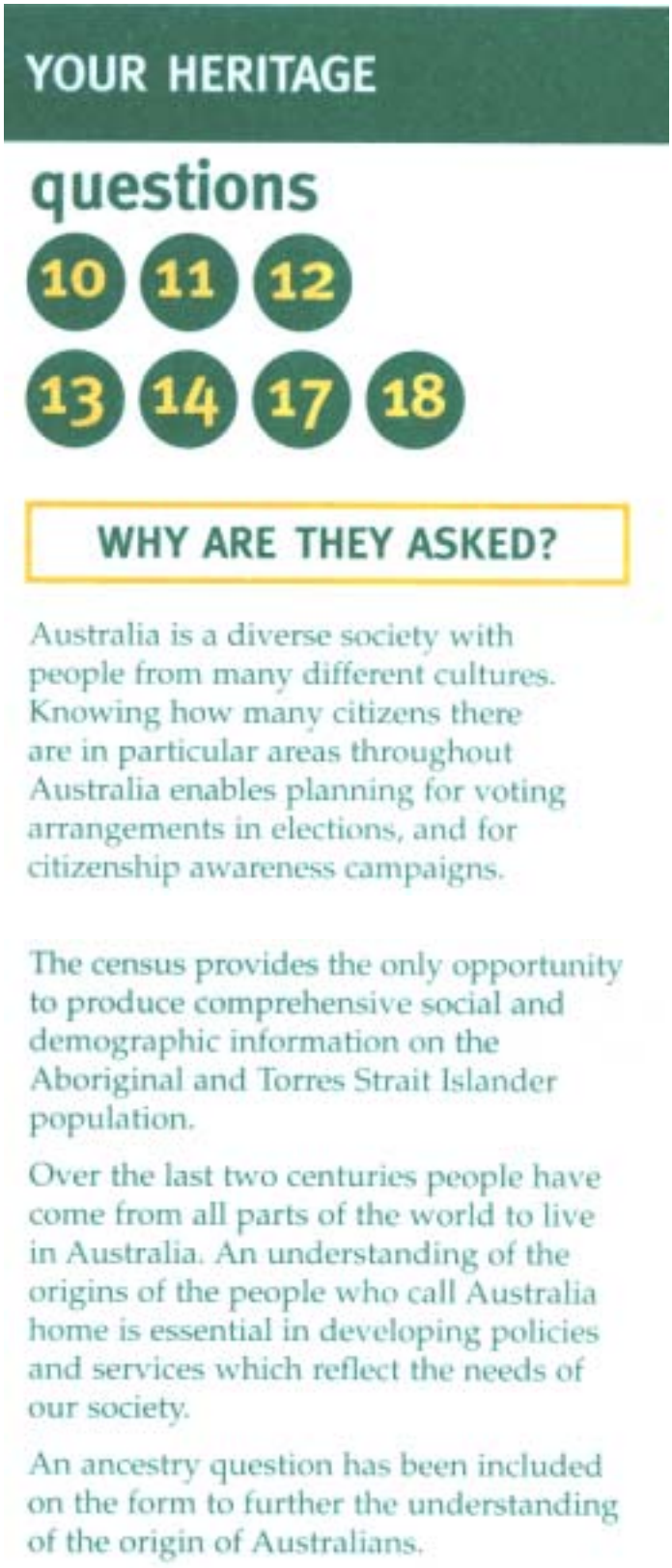
SIPF - Special Indigenous Personal Form. The standard form used in the enumeration of Indigenous communities. Information was collected via interview.

Validation - the checking of all Census variables for signs of any remaining or emerging system problems. This was undertaken by the DPC-based Validation Team, who included aspects of Ancestry in their work.

Write-in Response Boxes - a response box on the Census Form requiring a written response. It was generally coded using ICR (Intelligent Character Recognition) and then AC.

[1] The Australian Concise Oxford Dictionary, Oxford University Press, 1987

APPENDIX A - Ancestry-related information in the 2001 Census Household Guide



YOUR HERITAGE

questions

10 11 12

13 14 17 18

WHY ARE THEY ASKED?

Australia is a diverse society with people from many different cultures. Knowing how many citizens there are in particular areas throughout Australia enables planning for voting arrangements in elections, and for citizenship awareness campaigns.

The census provides the only opportunity to produce comprehensive social and demographic information on the Aboriginal and Torres Strait Islander population.

Over the last two centuries people have come from all parts of the world to live in Australia. An understanding of the origins of the people who call Australia home is essential in developing policies and services which reflect the needs of our society.

An ancestry question has been included on the form to further the understanding of the origin of Australians.

HOW TO ANSWER

Q18 When answering this question consider and mark the ancestries with which you most closely identify.

Count your ancestry back as far as three generations, if known. For example, consider your parents, grandparents and great grandparents.

If you are a descendant of South Sea Islanders brought to Australia as indentured labour around the turn of the twentieth century, please answer 'AUSTRALIAN SOUTH SEA ISLANDER'.

APPENDIX B - The Impact of Lost Ancestries, 2001 Census

IMPACT OF LOST ANCESTRIES, COUNTS & ESTIMATES,
ALL 189 ASCCEG CULTURAL & ETHNIC GROUPS, 2001 CENSUS

Ancestry	*2001 Census Ancestry Count*	*Lost (a) Ancestry Estimate (Freq. Aust.)*	*Revised Ancestry Estimate*	*% Lost*	*Frequency Ranking: Count \ Estimate*
Afghan	12,410	1,049	13,459	7.8	62 \ 63
African American	1,203	350	1,553	22.5	115 \ 115
Afrikaner	1,645	499	2,144	23.3	108 \ 105
Akan	124	0	124	0.0	170 \ 172
Albanian	10,459	399	10,858	3.7	67 \ 68
Algerian	696	100	796	12.5	132 \ 132
American	44,255	22,069	66,324	33.2	30 \ 28
Anglo-Burmese	822	50	872	5.7	125 \ 128
Anglo-Indian	12,327	1,098	13,425	8.2	63 \ 64
Angolan	115	0	115	0.0	172 \ 174
Arab, nec	623	0	623	0.0	134 \ 139
Argentinian	6,482	1,248	7,730	16.1	80 \ 77
Armenian	14,667	899	15,566	5.8	59 \ 60
Assyrian/Chaldean	18,667	100	18,767	0.5	49 \ 54
Australian	6,739,594	606,000	7,345,594	8.2	1 \ 1
Australian Aboriginal	94,950	9,936	104,886	9.5	19 \ 22
Australian South Sea Islander	3,442	250	3,692	6.8	91 \ 93
Austrian	38,112	9,986	48,098	20.7	33 \ 32
Basque	454	100	554	18.0	141 \ 143
Belarusan	1,378	499	1,877	26.6	111 \ 111
Bengali	9,549	100	9,649	1.0	71 \ 73
Berber	107	50	157	31.8	176 \ 169
Bolivian	473	0	473	0.0	138 \ 145
Bosnian	17,993	449	18,442	2.4	52 \ 57
Brazilian	3,763	799	4,562	17.5	89 \ 86
Breton	60	50	110	45.4	182 \ 177
British, nec	2,289	1,548	3,837	40.3	99 \ 89
Bulgarian	4,179	1,398	5,577	25.0	87 \ 83
Burgher	919	50	969	5.1	123 \ 126
Burmese	10,557	3,645	14,202	25.6	66 \ 62
Canadian	20,007	7,889	27,896	28.3	46 \ 44
Caribbean Islander, n.e.c.	722	100	822	12.1	131 \ 130
Catalan	109	0	109	0.0	175 \ 178
Central American, n.e.c.	815	0	815	0.0	126 \ 131
Central and West African, n.e.c.	1,692	399	2,091	19.1	107 \ 106
Central Asian, n.e.c.	1,356	0	1,356	0.0	112 \ 118
Chilean	21,579	1,148	22,727	5.0	43 \ 45
Chinese	556,554	19,722	576,276	3.4	6 \ 7
Chinese Asian, n.e.c.	500	0	500	0.0	137 \ 144
Colombian	3,475	0	3,475	0.0	90 \ 94
Cook Islander	8,154	649	8,803	7.4	73 \ 75
Coptic	3,344	50	3,394	1.5	92 \ 95
Croatian	105,747	7,490	113,237	6.6	17 \ 19
Cuban	414	150	564	26.5	145 \ 142
Czech	17,126	3,745	20,871	17.9	55 \ 51
Danish	38,637	31,606	70,243	45.0	32 \ 27
Dutch	268,754	52,177	320,931	16.2	9 \ 9

Ancestry (cont.)	*2001 Census Ancestry Count*	*Lost (a) Ancestry Estimate (Freq. Aust.)*	*Revised Ancestry Estimate*	*% Lost*	*Frequency Ranking: Count \ Estimate*
Eastern European, n.e.c.	431	150	581	25.8	143 \ 141
Ecuadorian	976	0	976	0.0	121 \ 125
Egyptian	27,001	3,245	30,246	10.7	39 \ 41
English	6,358,880	10,186	6,369,066	0.2	2 \ 2
Eritrean	2,029	0	2,029	0.0	103 \ 107
Estonian	7,543	1,947	9,490	20.5	76 \ 74
Ethiopian	3,054	0	3,054	0.0	93 \ 97
Fijian	16,620	2,147	18,767	11.4	56 \ 55
Filipino	129,821	5,293	135,114	3.9	15 \ 18
Finnish	18,106	4,394	22,500	19.5	51 \ 46
Flemish	460	399	859	46.4	139 \ 129
French	79,079	72,448	151,527	47.8	22 \ 16
French Canadian	1,246	399	1,645	24.3	114 \ 114
Fulani	25	0	25	0.0	184 \ 185
Georgian	332	0	332	0.0	151 \ 154
German	742,212	175,953	918,165	19.1	5 \ 4
Ghanaian	1,816	100	1,916	5.2	106 \ 108
Greek	375,703	15,279	390,982	3.9	8 \ 8
Gujarati	120	0	120	0.0	171 \ 173
Gurkha	15	0	15	0.0	185 \ 186
Guyanese	301	100	401	24.9	154 \ 148
Hispanic (North American)	2,606	0	2,606	0.0	97 \ 102
Hmong	1,836	0	1,836	0.0	105 \ 113
Hungarian	62,859	9,137	71,996	12.7	25 \ 26
I-Kiribati	358	50	408	12.2	148 \ 147
Icelandic	625	150	775	19.3	133 \ 133
Indian	156,628	15,029	171,657	8.7	11 \ 11
Indonesian	28,267	2,946	31,213	9.4	37 \ 39
Iranian	18,798	449	19,247	2.3	48 \ 53
Iraqi	11,190	0	11,190	0.0	65 \ 67
Irish	1,919,727	9,087	1,928,814	0.5	3 \ 3
Italian	800,256	47,933	848,189	5.6	4 \ 6
Jamaican	1,169	1,198	2,367	50.6	116 \ 103
Japanese	31,433	3,096	34,529	9.0	36 \ 37
Javanese	597	449	1,046	42.9	136 \ 122
Jewish	22,553	7,040	29,593	23.8	41 \ 42
Jordanian	2,687	100	2,787	3.6	96 \ 101
Kazakh	86	0	86	0.0	180 \ 181
Kenyan	1,118	300	1,418	21.1	119 \ 117
Khmer	21,361	100	21,461	0.5	44 \ 49
Korean	43,753	100	43,853	0.2	31 \ 34
Kurdish	4,494	0	4,494	0.0	85 \ 87
Kuwaiti	359	0	359	0.0	147 \ 151
Lao	10,086	50	10,136	0.5	69 \ 71
Latvian	18,938	3,196	22,134	14.4	47 \ 47
Lebanese	162,239	6,091	168,330	3.6	10 \ 12
Libyan	176	0	176	0.0	164 \ 166
Lithuanian	12,317	2,297	14,614	15.7	64 \ 61
Macedonian	81,898	1,548	83,446	1.9	21 \ 24
Madurese	14	0	14	0.0	186 \ 187
Mainland South-East Asian, n.e.c.	767	0	767	0.0	129 \ 136
Malawian	106	0	106	0.0	177 \ 179
Malay	18,294	2,896	21,190	13.7	50 \ 50

Ancestry (cont.)	*2001 Census Ancestry Count*	*Lost (a) Ancestry Estimate (Freq. Aust.)*	*Revised Ancestry Estimate*	*% Lost*	*Frequency Ranking: Count \ Estimate*
Malayali	91	50	141	35.4	179 \ 170
Maltese	136,754	15,528	152,282	10.2	14 \ 15
Maori	72,956	17,525	90,481	19.4	24 \ 23
Marathi	9	0	9	0.0	188 \ 188
Maritime South-East Asian, n.e.c.	1,387	499	1886	26.4	110 \ 109
Mauritian	17,886	2,596	20,482	12.7	53 \ 52
Melanesian and Papuan, n.e.c.	152	50	202	24.7	168 \ 163
Mexican	1,635	250	1,885	13.2	109 \ 110
Micronesian, n.e.c.	202	0	202	0.0	162 \ 164
Moldovan	140	0	140	0.0	169 \ 171
Mongolian	415	300	715	41.9	144 \ 137
Montenegrin	771	0	771	0.0	127 \ 134
Moroccan	1,161	350	1,511	23.1	117 \ 116
Mozambican	112	0	112	0.0	174 \ 176
Namibian	41	0	41	0.0	183 \ 184
Native North American Indian	1,856	1,847	3,703	49.9	104 \ 92
Nauruan	280	0	280	0.0	157 \ 161
Nepalese	2,946	0	2,946	0.0	94 \ 99
New Caledonian	173	200	373	53.6	166 \ 150
New Zealander	123,314	38,047	161,361	23.6	16 \ 13
Ni-Vanuatu	311	350	661	52.9	153 \ 138
Nicaraguan	456	150	606	24.7	140 \ 140
Nigerian	1,160	150	1,310	11.4	118 \ 119
Niuean	1,301	549	1,850	29.7	113 \ 112
North American, n.e.c.	261	50	311	16.0	159 \ 157
Northern European, nec	957	50	1,007	5.0	122 \ 124
Norwegian	17,293	12,283	29,576	41.5	54 \ 43
Oromo	398	0	398	0.0	146 \ 149
Other North-East Asian, n.e.c.	5	0	5	0.0	189 \ 189
Other North-African and Middle Eastern, n.e.c.	289	0	289	0.0	155 \ 159
Pakistani	12,618	300	12,918	2.3	61 \ 65
Palestinian	7,001	300	7,301	4.1	78 \ 79
Papua New Guinean	9,441	1,847	11,288	16.3	72 \ 66
Pathan	175	0	175	0.0	165 \ 167
Peruvian	4,772	449	5,221	8.6	84 \ 84
Polish	150,900	32,155	183,055	17.5	13 \ 10
Polynesian, nec	2,101	1,698	3,799	44.7	102 \ 90
Portuguese	35,687	7,839	43,526	18.0	34 \ 35
Punjabi	2,263	0	2,263	0.0	100 \ 104
Romanian	16,121	2,646	18,767	14.1	57 \ 56
Roma/Gypsy	603	300	903	33.2	135 \ 127
Russian	60,213	15,978	76,191	21.0	26 \ 25
Salvadoran	6,617	100	6,717	1.5	79 \ 80
Samoan	28,091	2,247	30,338	7.4	38 \ 40
Saudi Arabia	181	0	181	0.0	163 \ 165
Scottish	540,046	333,383	873,429	38.1	7 \ 5
Serbian	97,315	8,288	105,603	7.8	18 \ 21
Seychellois	2,104	799	2,903	27.5	101 \ 100
Sikh	1,097	0	1,097	0.0	120 \ 121
Sinhalese	58,602	3,395	61,997	5.5	27 \ 29

Ancestry (cont.)	*2001 Census Ancestry Count*	*Lost (a) Ancestry Estimate (Freq. Aust.)*	*Revised Ancestry Estimate*	*% Lost*	*Frequency Ranking: Count \ Estimate*
Slovak	7,054	799	7,853	10.2	77 \ 76
Slovene	14,189	1,648	15,837	10.4	60 \ 59
Soloman Islander	769	0	769	0.0	128 \ 135
Somali	5,007	150	5,157	2.9	83 \ 85
South African	52,119	9,836	61,955	15.9	29 \30
South American, n.e.c.	765	250	1,015	24.6	130 \ 123
South Eastern European, n.e.c.	115	0	115	0.0	173 \ 175
Southern and East African, n.e.c.	2,410	549	2,959	18.5	98 \ 98
Southern Asian, n.e.c.	915	250	1,165	21.4	124 \ 120
Southern European, n.e.c.	218	100	318	31.4	161 \ 156
Spanish	75,237	33,902	109,139	31.0	23 \ 20
Sudanese	3,788	0	3,788	0.0	88 \ 91
Sundanese	84	0	84	0.0	181 \ 182
Swedish	24,424	21,420	45,844	46.7	40 \ 33
Swiss	22,151	9,936	32,087	30.9	42 \ 38
Syrian	10,213	549	10,762	5.1	68 \ 69
Taiwanese	4,416	0	4,416	0.0	86 \ 88
Tamil	7,706	0	7,706	0.0	75 \ 78
Tanzanian	269	0	269	0.0	158 \ 162
Thai	20,606	1,248	21,854	5.7	45 \ 48
Tibetan	289	0	289	0.0	156 \ 160
Timorese	5,491	300	5,791	5.2	81 \ 82
Tongan	14,889	1,098	15,987	6.9	58 \ 58
Torres Strait Islander	9,791	399	10,190	3.9	70 \ 70
Trinidian (Tobagonian)	356	0	356	0.0	149 \ 152
Tunisian	258	50	308	16.2	160 \ 158
Turkish	54,596	1,248	55,844	2.2	28 \ 31
Ugandan	337	0	337	0.0	150 \ 153
Ukrainian	33,960	5,243	39,203	13.4	35 \ 36
Uruguayan	5,196	599	5,795	10.3	82 \ 81
Uzbek	160	0	160	0.0	167 \ 168
Venezuelan	448	0	448	0.0	142 \ 146
Vietnamese	156,581	1,198	157,779	0.8	12 \ 14
Walloon	14	50	64	78.1	187 \ 183
Welsh	84,246	65,358	149,604	43.7	20 \ 17
Western European, n.e.c.	7,917	1,897	9,814	19.3	74 \ 72
Yoruba	103	0	103	0.0	178 \ 180
Zambian	328	0	328	0.0	152 \ 155
Zimbabwean	2,896	449	3,345	13.4	95 \ 96

(a) The 2% DQI Sample may have led to some - particularly smaller - Ancestry groups being significantly under or over-represented in the extrapolated Lost Ancestry count.

APPENDIX C - Multiple Response Rate (a) to Ancestry, by Birthplace, 1986 & 2001 Censuses

Birthplace	*1986 Rate % (a)*	*2001 Rate % (a)*	*Birthplace*	*1986 Rate % (a)*	*2001 Rate % (a)*
Australia	14.2	25.9	Bosnia & Herzegovina (2001)		3.5
Norfolk Island		36.7	Bulgaria	5.1	5.8
New Zealand	17.2	28.5	Croatia (2001)	N/A	3.9
Papua New Guinea	24.4	35.0	Cyprus	14.3	6.9
Fiji	12.7	11.0	Greece	1.9	1.8
Other Oceania & Antarctica	11.7	13.2	FYROMacedonia (2001)	N/A	1.7
Total Oceania & Antarctica		*25.9*	Romania	5.7	7.3
			Yugoslavia (1986)	8.1	N/A
England	5.6	10.8	Yugoslavia (2001)	N/A	5.5
Northern Ireland	6.7	9.3	Other South Eastern Europe		5.6
Scotland	5.5	11.2	*Total South Eastern Europe*		*3.5*
Wales	10.4	18.6			
Other United Kingdom		23.0	Belarus (2001)		9.8
Total United Kingdom		*11.0*	Czechoslovakia (1986)	5.2	N/A
Ireland	4.3	5.2	Czech Republic (2001)	N/A	6.5
Total UK and Ireland (1986)	5.7		Estonia (2001)	N/A	6.0
			Hungary	3.6	4.8
Austria	7.9	11.1	Latvia (2001)	N/A	6.0
Belgium	11.3	18.1	Lithuania (2001)	N/A	5.0
France	5.1	18.3	Poland	2.5	3.8
Germany	6.7	8.7	Russian Federation (2001)	N/A	8.0
Netherlands	3.6	5.0	Slovakia (2001)	N/A	6.5
Switzerland	11.3	18.3	Ukraine (2001)	N/A	8.0
Other Western Europe		21.9	*Total Eastern Europe*		*5.5*
Total Western Europe		*8.9*			
			Egypt	8.5	11.6
Denmark	5.3	8.9	Other North Africa		13.3
Finland	2.3	4.4	*Total North Africa*		*12.0*
Norway	8.9	10.3			
Sweden	9.3	15.8	Iran	3.7	3.4
Other Northern Europe		11.4	Iraq	4.0	5.5
Total Northern Europe		*9.5*	Israel	9.9	20.1
			Lebanon	3.1	2.6
Italy	1.2	1.7	Syria	5.4	5.7
Malta	3.2	4.7	Turkey	2.4	3.0
Portugal	1.8	2.3	Other Middle East		10.6
Spain	2.7	4.3	*Total Middle East*		*4.6*
Other Southern Europe		25.9			
Total Southern Europe		2.4	(*Continued ....*)		

(See footnotes at end of table)

APPENDIX C *(continued)*

Birthplace	*1986 Rate % (a)*	*2001 Rate % (a)*	*Birthplace*	*1986 Rate % (a)*	*2001 Rate % (a)*
Burma	24.8	17.8	Afghanistan		2.2
Cambodia	5.3	7.3	Armenia		7.4
Laos	2.7	5.6	Other Central Asia		11.8
Thailand	8.0	10.3	*Total Central Asia*		*3.7*
Viet Nam	2.5	2.9			
Total Mainland S-E Asia		*5.0*	Canada	24.2	37.5
			United States	31.6	40.3
Brunei Darussalam		12.5	Other North America		26.1
Indonesia	10.2	11.3	*Total North America*	27.8	*39.3*
Malaysia	8.3	10.1			
Philippines	9.2	10.4	Argentina	21.3	29.0
Singapore	12.5	13.6	Brazil	18.3	24.8
East Timor	N/A	13.9	Chile	8.2	12.2
Timor (1986)	11.3	N/A	Colombia		9.2
Total Maritime S-E Asia		*11.0*	Ecuador		9.4
			Paraguay		19.9
China	3.3	1.5	Peru		16.8
Hong Kong	4.3	3.9	Uruguay	13.7	22.1
Taiwan		2.3	Venezuela		23.1
Other Chinese Asia		5.0	Other South America		18.3
Total Chinese Asia		*2.3*	*Total South America*	13.1	*18.1*
Japan	5.4	6.5	El Salvador		6.6
Korea (1986)	1.5	N/A	Guatemala		18.4
Korea DPR (North)		0.0	Mexico		27.9
Korea, Rep (South)		1.1	Other Central America		14.3
Total Japan & the Koreas		*3.2*	*Total Central America*		*9.7*
Bangladesh	11.5	3.5	*Total Caribbean*		*28.4*
India	21.0	9.8			
Pakistan	12.8	8.2	Kenya		16.8
Sri Lanka	13.6	10.5	Mauritius		13.5
Other Southern Asia		3.3	South Africa	16.8	23.5
Total Southern Asia		*9.5*	Zimbabwe		27.4
			Other Southern & East Africa		13.9
Nigeria		13.4	*Total Southern & East Africa*		*20.7*
Other Central & West Africa		10.0			
Total Central & West Africa		*11.2*	**Total all Birthplaces**	**12.6**	**22.3**

(a) Persons in each Birthplace category responding with more than one Ancestry as a percentage of all persons in that Birthplace category**.**

(b) Only 1986 birthplace data that are directly comparable with 2001 boundaries have been included. Most 1986 regional totals have been omitted due to the variance in composition.

Census Papers

2001 Census Papers:

03/04 *2001 Census: Income*
03/03 *2001 Census: Computer and Internet Use*
03/02 *2001 Census: Housing*
03/01b *2001 Census: Ancestry - Detailed Paper*
03/01a *2001 Census: Ancestry - First and Second Generation Australians*
02/03 *2001 Census: Form Design Testing*
02/02 *Report on Testing of Disability Questions for Inclusion in the 2001 Census*
02/01 *2001 Census: Digital Geography Technical Information Paper*

1996 Census Working Papers:

00/4 *1996 Census Data Quality: Income*
00/3 *1996 Census Data Quality: Industry*
00/2 *1996 Census Data Quality: Qualification Level and Field of Study*
00/1 *1996 Census Data Quality: Journey to Work*
99/6 *1996 Census Data Quality: Occupation*
99/4 *1996 Census: Review of Enumeration of Indigenous Peoples in the 1996 Census*
99/3 *1996 Census Data Quality: Housing*
99/2 *1996 Census: Labour Force Status*
99/1 *1996 Census: Industry Data Comparison*
97/1 *1996 Census: Homeless Enumeration Strategy*
96/3 *1996 Census of Population and Housing: Digital Geography Technical Information Paper*
96/2 *1996 Census Form Design Testing Program*

1991 Census Working Papers:

96/1 *Income*
95/1 *Housing*
94/4 *Ancestry*
94/3 *Disability*
94/2 *Education*
94/1 *Labour Force Status*
93/6 *Aboriginal/Torres Strait Islander Counts*
93/5 *Public Communications*
93/4 *Comparison of Census and PES Responses*
93/3 *Posted-in Forms*
93/2 *Self Coding*
93/1 *Sequencing Instructions*

These papers are available on the ABS web site at <http://www.abs.gov.au>. From the ABS home page, select **Census -> (Census Information) Fact Sheets and Census Papers -> (Other Publications) Census Papers.**

If you have further data quality queries, please contact the Assistant Director, Census Evaluation by telephone: (02) 6252 5611 or email: <joanne.healey@abs.gov.au>.