



2001 CENSUS OF POPULATION AND HOUSING

Census Paper

**2001 CENSUS: OCCUPATION
(Census Paper No. 03/06)**

2001 CENSUS: OCCUPATION
(Census Paper No. 03/06)

CHRIS KUNZ

Population Census Evaluation
October 2003

© Commonwealth of Australia 2003

This work is copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organisation. Apart from any use as permitted under the *Copyright Act 1968*, all other rights are reserved. Requests and inquiries concerning reproduction and rights in this publication should be addressed to:

The Manager
Intermediary Management
Australian Bureau of Statistics
Locked Bag 10
Belconnen ACT 2616

or telephone (02) 6252 6998 or fax (02) 6252 7102

or email <intermediary.management@abs.gov.au>.

In all cases, the ABS must be acknowledged as the source when reproducing or quoting any part of an ABS publication or other product.

For general inquiries about ABS products and services please call 1300 135 070. Overseas clients please call +61 2 9268 4909.

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

This 2001 Census Paper evaluated the quality of Occupation data. In general, Occupation data from the 2001 Census was of a higher standard than that of 1996:

- The non-response rate was lower (1.2% compared with 1.7% in 1996), making it once again the lowest rate for any released Census variable.
- 2001 data was more definitively coded, with 93.5% (as opposed to 90.4%) coded down to the lowest (6-digit) level of the ASCO classification.
- Discrepancy rates fell from 10.7% in 1996 to 5.4% in 2001, partly due to the advent of Automatic Coding, which coded 57% of all records.
- Automatic Coding discrepancy rates (at 4.6%) were lower than those of the human coder (at 6.2%), indicating that the system's introduction was generally successful.
- The specific form design change that gave examples of the type of farming, was a contributing factor in successfully increasing coding to the lowest level for farming type, by around a third.
- The 2001 Census and August 2001 Labour Force Survey results were similar and mirrored the relationship between their 1996 versions.

CONTENTS

LISTING OF FIGURES & TABLES

1. INTRODUCTION	1
1.1 About Census Papers	1
1.2 Background	2
1.3 Quality Issues Relating to Occupation Data	2
2. QUESTION DESIGN	3
2.1 Form Design	3
2.2 Differences between the 2001 and 1996 Forms	3
2.3 Full-time/Part-time Job - the Gateway Question	4
2.4 2001 Census Occupation-related Questions	5
2.5 1996 Census Occupation-related Questions	6
3. COLLECTION OF THE DATA	7
4. PROCESSING AT THE DATA PROCESSING CENTRE (DPC)	9
4.1 Data Capture (DC) and Intelligent Character Recognition (ICR)	9
4.2 Automatic Coding (AC)	9
4.3 Computer Assisted Coding (CAC)	10
4.4 The Raising of Queries	11
4.5 The Index and Classification	11
4.6 Edits Applied to the Data	12
4.7 Quality Management and Discrepancy Analysis	13
5. SAMPLE DATA	19
5.1 DQI Sample Analysis	19
6. FINAL DATA ANALYSIS	21
6.1 Non-response	21
6.2 Not Further Defined Coding	23
6.3 Case Study	26
7. RECONCILIATION OF 2001 CENSUS OCCUPATION DATA WITH LABOUR FORCE SURVEY DATA	29
7.1 Data Reconciliation Methodology	29
7.2 Deductions from Census Counts	30
7.3 Deductions from Labour Force Survey Counts	30
7.4 Results of Data Reconciliation	31

8. CONCLUSIONS	37
9. RECOMMENDATIONS	39
GLOSSARY	41
REFERENCES	43
APPENDIX 1 : Reconciliation between 2001 Census and August 2001 Labour Force Survey: More Detail	45
CENSUS PAPER LISTING	49

LISTING OF FIGURES & TABLES

Figure 1 : Full-time/Part-time Job (Gateway Question) 2001 Census	5
Figure 2 : Occupation, 2001 Census	5
Figure 3 : Tasks or Duties, 2001 Census	5
Figure 4 : Full-time/Part-time Job (Gateway Question) 1996 Census	6
Figure 5 : Occupation, 1996 Census	6
Figure 6 : Tasks or Duties, 1996 Census	6
Table 1 : Automatic Coding Rates for Occupation by Major Group, 2001 Census	10
Table 2 : Queries Raised for Occupation, by State and Australia, 2001 Census	11
Figure 7: Discrepancy Rates for Occupation, by Week Ending, 2001 Census	14
Table 3 : Discrepancy Rate and Distribution at ASCO Level, by Processing Type, 2001 Census	14
Table 4 : Discrepancies for Occupation by Group Level (1 to 4), 1996 and 2001 Censuses	15
Table 5 : Discrepancy Counts for Occupation by Major Group, 2001 Census	16
Table 6 : Top 10 Discrepancies by Occupation, by AC Percentage, 2001 Census	17
Table 7 : DQI Sample for Occupation, by Full-time/Part-time Job, 2001	19
Table 8 : Non-response Rate for Occupation, 1996 and 2001 Censuses	21
Table 9 : Non-response Rates for Occupation-related Census Questions, 1996 and 2001	21
Table 10: Occupation by Stated/Not Stated, by Sex, Income, Qualification & Birthplace, percent for applicable population	22
Table 11: Distribution of Not Further Defined Responses in the 1996 and 2001 Censuses	24
Table 12: Percentages for Farmers and Farm Managers by Group Level, 1996 and 2001 Censuses	26
Table 13: Percentages for Occupation by ASCO Level, 1996 and 2001 Censuses	27
Table 14: Adjustments made to August 2001 Labour Force Survey Benchmarks and 2001 Census to Derive a Common Population for Occupation Data	30
Table 15: Occupation Major Groups by Age, 2001 Census as a Percentage of August 2001 Labour Force Survey, for Persons, Australia	31
Table 16: Occupation Major Groups by Age, 1996 Census as a Percentage of August 1996 Labour Force Survey, for Persons, Australia	32
Table 17: Percentage Rates for Occupation Major Groups by Age, Persons, Australia, 2001 Census	33
Table 18: Percentage Rates for Occupation Major Groups by Age, Persons, Australia, August 2001 Labour Force Survey	33
Table 19: Percentage Rates for Occupation Major Groups by States and Territories, Persons, 2001 Census	34

Table 20: Percentage Rates for Occupation Major Groups by States and Territories, Persons, August 2001 Labour Force Survey	35
Table A1: Adjusted Counts for Occupation Major Groups by Age, 2001 Census	45
Table A2: Adjusted Counts for Occupation Major Groups by Age, August 2001 Labour Force Survey	46
Table A3: Adjusted Counts for Occupation Major Groups by States and Territories, 2001 Census	47
Table A4: Adjusted Counts for Occupation Major Groups by States and Territories, August 2001 Labour Force Survey	48

1. INTRODUCTION

1.1 *About Census Papers*

The ABS has a stated, corporate objective to provide the means for informed and increased use of statistics. This Paper is one 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 they should keep in mind, that have been identified as impacting on the quality of the census 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.

1.1.1 *This Paper*

This Paper's focus is Occupation, a question that has been asked in every Census since the first national census of 1911.

Data on occupation are used for analysing current and potential imbalances in the labour market. This information is then used to develop policies and programs in education, training, immigration, industry and industrial relations.

Occupation data are collected for employed persons of 15 years of age or older.

This paper contains information about question design for Occupation data in the 2001 and previous Censuses, and how the design and sequencing of questions can affect the quality of responses. Both 1996 and 2001 question content and format are shown in *2.4 and 2.5: Census Occupation-related Questions*.

A description of the Quality Management system, as applied to Occupation data, is provided. Further analyses examine the coding discrepancies for the different types of data processing, as recorded by the Data Processing Centre's (DPC) Management Information System (MIS).

There are frequent statistical references to 1996 data in this Paper. This has been done to provide a comparative measure between results gained from the Occupation questions in the two most recent censuses. The differences, while providing a guide to occupational change for Australia's population over the five year period, will be examined for an indication of the impact on data quality of any changes to question content, format or processing methods.

The Paper analyses Non-response rates for Occupation from the 1996 and 2001 Censuses. It describes procedures used to code data to the *Australian Standard Classification of Occupations (ASCO)*, Second Edition, and includes an analysis of the level of Not Further Defined coding allocated during the 2001 Census and a comparison with 1996.

Finally, the paper compares 2001 Census Occupation data to the August 2001 Labour Force Survey Occupation data.

1.2 Background

Prior to 1986, a single question was asked on title of Occupation. In 1986 a second question on the main tasks or duties that a person usually performed in his or her job was included to improve the quality of coding. The questions remained the same for subsequent censuses including the 2001 Census, but the examples and instructions were revised in attempts to improve reporting by respondents.

In 1986, for the first time, responses were coded using the *Australian Standard Classification of Occupations* (ASCO), and Computer Assisted Coding (CAC) was introduced for Occupation responses.

For the 1996 Census the coding system remained the same but Occupation data were coded using a revised (2nd Edition) version of the ASCO.

For the 2001 Census a new system was introduced to read and process information reported on Census Forms. Intelligent Character Recognition (ICR) software scanned the census forms, read the handwritten text, verified (and if deemed necessary, repaired) the text read from the form, and stored the form image and information for additional processing. Many Occupations were able to be automatically coded from the Occupation title response. Snippet images of responses unable to be automatically coded were sent to coding staff for resolution.

1.3 Quality Issues Relating to Occupation Data

The Census is ‘self-enumerated’ which means that the Census Form is completed by the respondent with minimal assistance from the Census Collector. Thus the way questions are presented in the Census Form, the sequencing, the instructions and the examples used to help respondents answer the questions contribute to a large extent to the response rate and to the ability to adequately code the responses.

Processing issues can also affect data quality. The main processing issues examined in this paper are:

- Changes to the method of data capture;
- The new automatic coding process; and
- Modifications to CAC.

2. QUESTION DESIGN

2.1 *Form Design*

Accurate and complete responses to census questions depend strongly on form design. The major aspects to consider when trying to improve form design are:

- clear sequencing of questions;
- clear and concise instructions;
- relevant examples in the questions;
- no leading or biased wording, and
- option and space for response.

The current question structure was devised for the 1986 Census in conjunction with Computer Assisted Coding (CAC). Some changes were implemented for later censuses to increase the level of responses. During the 1996 Census processing there were concerns about the final form design because Question 32 about 'Occupation Title' and Question 33 about 'Tasks Performed' were on a different page to employer name and industry, requiring coders to flip between pages if coding data using both Occupation and Industry information. This might have led to a loss of information as coders had greater difficulty in identifying the correct data they should select. For the 2001 Census the use of imaging removed this problem: the coders could view snippets of all relevant images from the form on their computer screen.

Changes to the wording of the questions for the 2001 Census were minimal (see 2.2 *Differences Between the 2001 and 1996 Forms*). The questions about 'Job Title' and 'Main Tasks' contained additional examples: 'Sheep and Wheat Farmer' and 'running a sheep/wheat farm'. The aim was to reduce the number of respondents answering 'Farmer' or 'farming' which led to the allocation of the Not Further Defined ASCO code 131 'Farmers and Farm Managers' in the 1996 Census (see 6.3 *Case Study*).

The use of boxes for answers instead of dotted lines (refer to sections 2.4 and 2.5 respectively, for Occupation questions in the 2001 and 1996 Censuses) was necessary for the Intelligent Character Recognition (ICR) process introduced for 2001 [see 4.1 *Data Capture (DC) and Intelligent Character Recognition (ICR)*].

2.2 *Differences Between the 2001 and 1996 Forms*

The 2001 and 1996 Census Forms were close to identical in their Occupation-related content. In 1996, Occupation had one of the lowest non-response rates of all questions, so there was no trigger for major form design modifications.

The Occupation title question 'In the main job...' in 2001 differed from 1996 in only three minor respects:

- the example 'Pastrycook' became 'Pastry Cook'
- an additional example 'Sheep and Wheat Farmer' was provided
- boxes for writing individual letters in the response replaced lines.

Differences between the Tasks or Duties question ('What are the main tasks...') in 2001 and 1996 reflected the changes in the title question described above:

- the additional example of 'running a sheep/wheat farm'
- boxes for writing individual letters in the response, replaced lines.

2.3 Full-time/Part-time Job - the Gateway Question

The Full-time/Part-time Job (FPJP) question (No. 32 on the 2001 Census Household Form - see 2.4) was the 'gateway' through which those answering the Occupation questions (34 and 35) needed to pass.

Four groups of respondents were permitted through the 'gateway' to have their answer to the Occupation questions coded. These were those who answered to FPJP with:

- Yes, worked for payment or profit;
- Yes, but absent on holidays, on paid leave, on strike or temporarily stood down;
- Yes, unpaid work in a family business; or
- Those who did not respond to FPJP at all.

Occupation details supplied by those who did not answer the gateway question were also coded, to maximise the value of the data.

Those who marked the fourth or fifth options:

- Yes, other unpaid work; or
- No did not have a job

were sequenced to Question 42 (Actively Looking for Work - ATSP), and any responses made to the Occupation questions were not coded.

2.4 2001 Census Occupation-related Questions (Household Form)

Figure 1: Full-time/Part-time Job (Gateway Question), 2001 Census

<p>32 <i>Last week, did the person have a full-time or part-time job of any kind?</i></p> <ul style="list-style-type: none"> ▪ Mark one box only. ▪ A 'job' means any type of work including casual or temporary work or part-time work, if it was for one hour or more. ▪ See page 11 of the Census Guide for more information. 	<ul style="list-style-type: none"> <input type="radio"/> Yes, worked for payment or profit <input type="radio"/> Yes, but absent on holidays, on paid leave, on strike or temporarily stood down <input type="radio"/> Yes, unpaid work in a family business <input type="radio"/> Yes, other unpaid work ▶ Go to 42 <input type="radio"/> No, did not have a job ▶ Go to 42
---	---

Figure 2: Occupation, 2001 Census

<p>34 <i>In the main job held last week, what was the person's occupation?</i></p> <ul style="list-style-type: none"> ▪ Give full title. ▪ For example, Childcare Aide, Maths Teacher, Pastry Cook, Tanning Machine Operator, Apprentice Toolmaker, Sheep and Wheat Farmer. ▪ For public servants, state official designation and occupation. ▪ For armed services personnel, state rank and occupation. 	<p>Occupation</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> </table>																																																												

Figure 3: Tasks or Duties, 2001 Census

<p>35 <i>What are the main tasks that the person himself/herself usually performs in that occupation?</i></p> <ul style="list-style-type: none"> ▪ Give full details. ▪ For example, looking after children at a day care centre, teaching secondary school students, making cakes and pastries, operating leather tanning machine, learning to make and repair tools and dies, running a sheep/wheat farm. ▪ For managers, state main activities managed. 	<p>Tasks or duties</p> <table border="1" style="width: 100%; height: 100%; border-collapse: collapse;"> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> </table>																																																												

2.5 1996 Census Occupation-related Questions (Household Form)

Figure 4: Full-time/Part-time Job (Gateway Question), 1996 Census

30 Last week, did the person have a full-time or part-time job of any kind?

- Mark one box only.
- A job means any type of work including casual or temporary work or part-time work, if it was for one hour or more.

Yes, worked for payment or profit

Yes, but absent on holidays, on paid leave, on strike or temporarily stood down

Yes, unpaid work in a family business

Yes, other unpaid work

▶ Go to 39

No, did not have a job

▶ Go to 39

Figure 5: Occupation, 1996 Census

32 In the main job held last week, what was the person's occupation?

- Give full title.
- For example, Childcare Aide, Maths Teacher, Pastrycook, Tanning Machine Operator, Apprentice Toolmaker.
- For public servants, state official designation and occupation. For armed services personnel, state rank and occupation.

Occupation

Figure 6: Tasks or Duties, 1996 Census

33 What are the main tasks that the person himself/herself usually performs in that occupation?

- Give full details.
- For example, looking after children at day care centre, teaching secondary school students, making cakes and pastries, operating leather tanning machine, learning to make and repair tools and dies.
- For managers, state main activities managed.

Tasks or duties

3. COLLECTION OF THE DATA

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 created 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 was apparent for many 2001 Census variables (though not Occupation). Most of the change can be attributed to the increase in the proportion of System Created Records. A Fact Sheet has been produced 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. Discussion of the Non-response Rates for Occupation are in Section *6.1 Non-response*.

Australian censuses are self-enumerated which means that respondents fill in the forms themselves. Various reasons may prevent potential respondents from answering the questions relating to Occupation either completely or accurately. They may:

- provide insufficient or imprecise information;
- not answer because of their reluctance to disclose details of their Occupation;
- not answer because of the perceived difficulty of the questions;
- misinterpret sequencing of questions and therefore skip relevant ones;
- write multiple answers;
- mis-identify, or
- even seek to elevate the status of their Occupation or role.

Other factors may increase the level of non-response, such as random responding and the general tendency to omit write-in answers due to the effort required. These issues are reflected in the amount of non-response to the questions and in the number of 'Inadequately Described' and 'Not Further Defined' (NFD) codes (see *6.2 Not Further Defined Coding*) assigned by the process.

4. PROCESSING AT THE DATA PROCESSING CENTRE (DPC)

4.1 *Data Capture (DC) and Intelligent Character Recognition (ICR)*

Data Capture (DC) is the process of scanning Census Forms into the image and text files that are used for all subsequent processes.

At this stage, mark box responses are captured and coded. For the 2001 Census a new system was introduced as part of DC. The Intelligent Character Recognition (ICR) system read the hand-written text responses and translated them 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).

Records are automatically repaired where they are marked in such a way that they conform to initial tolerance guidelines for forming particular letters or numbers, as well as acceptable marking, if mark boxes.

Others that fail to meet such guidelines are sent to Manual Repair, where an operator studies the letter or number, initially in isolation, in an attempt to clarify the respondent's intention.

The record shifts through three further phases, if the Manual Repairer is unsure:

- Triplets, where the textual elements on either side are visible
- Fields, where the whole field for that question is visible
- Forms, where the whole form can be perused for similar letter/number formations.

Where the degree of surety was so low that neither Manual Repair nor Automatic Coding were possible, the field was sent to Computer Assisted Coding (CAC).

4.2 *Automatic Coding (AC)*

Automatic Coding (AC), introduced for the first time in the 2001 Census, was the next phase. The system sought to match a *basic* and a *qualifying* word from the response in the Census Form image to the Occupation Index. [This mirrored the CAC procedures.]

Basic words are single words that can stand alone as the title of the respondent's Occupation, e.g., Clerk. *Qualifying* words are those in a title that more specifically describe the basic word, providing a clearer idea of the type of Occupation e.g., Accounts, as in Accounts Clerk.

Responses not automatically matched due to their indecipherable nature, or the lack of entry in the Index, underwent Computer Assisted Coding (CAC) - a process similar to that used in the 1996 Census.

AC rates for each Occupation Major Group are shown in the following table:

Table 1: Automatic Coding Rates for Occupation by Major Group, 2001 Census

<i>Major Group</i>	<i>AC'd</i>		<i>Not AC'd</i>	
	<i>Number</i>	<i>%</i>	<i>Number</i>	<i>%</i>
Managers and Administrators	373,447	48.8	391,376	51.2
Professionals	887,930	58.6	626,166	41.4
Associate Professionals	533,413	54.7	442,240	45.3
Tradespersons and Related Workers	659,073	64.7	359,830	35.3
Advanced Clerical and Service Workers	200,222	64.6	109,746	35.4
Intermediate Clerical, Sales and Service Workers	745,148	54.5	621,553	45.5
Intermediate Production and Transport Workers	398,395	59.4	272,426	40.6
Elementary Clerical, Sales and Service Workers	503,105	63.5	289,273	36.5
Labourers and Related Workers	380,154	53.0	337,303	47.0
Inadequately Described	4,190	6.1	64,787	93.9
Total all Stated	4,685,077	57.1	3,514,700	42.9

An average 57% of records were coded by the AC system, taking a considerable workload from the manual coding process. For detail on the accuracy of AC and CAC, see 4.7 *Quality Management and Discrepancy Analysis*.

Tradespersons and Related Workers had the highest AC rate at 64.7%, while Managers and Administrators had the lowest at 48.8%.

As Automatic Coding was only introduced in 2001, it is not possible to provide comparative data for 1996.

4.3 *Computer Assisted Coding (CAC)*

CAC is the process of using procedures and rules to allow a coder to match the image of 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.

As with AC, the coder was required to identify *basic* and *qualifying* words from the response given on the Census Form.

The coder entered the first three letters of the *basic* and *qualifying* words. Matches from the words displayed on the computer screen were selected based on colour matching rules:

<i>Colour</i>	<i>Colour Matching Rule</i>
Red	Index entry can only be selected if all the words can be found in the Occupation Title
Blue	Coder can look for information in Occupation Task response or unused parts of Occupation Title
Black	Match from any of Occupation Title, Task or even Employer or Industry responses

Limitations imposed on differentiation by colour, were not applicable to AC directly, though its principles were reflected in its coding. AC could also access Industry and Employer information (equivalent to CAC’s Black colour match level) if stated.

4.4 *The Raising of Queries*

When the message 'Raise a query for this response' was displayed, it meant that a matching Index entry could not be found by the system for the Occupation title, and this was referred to an expert group of coders with access to a wide range of coding resources for resolution.

Table 2: Queries Raised for Occupation, by State and Australia, 2001 Census

	<i>NSW</i>	<i>VIC</i>	<i>QLD</i>	<i>SA</i>	<i>WA</i>	<i>TAS</i>	<i>NT</i>	<i>ACT</i>	<i>OT(a)</i>	<i>AUST</i>
<i>Queries</i>	407,214	271,740	187,334	94,823	128,731	28,045	16,189	32,399	220	1,166,695
<i>Records</i>	2,715,089	2,055,360	1,549,653	628,911	819,965	180,500	89,175	159,943	1,181	8,199,777
<i>Query %</i>	15.0	13.2	12.1	15.1	15.7	15.5	18.1	20.3	18.6	14.2

(a) Other Territories (OT), includes Christmas Island, Cocos (Keeling) Islands, and the Jervis Bay Territory.

The query percentage overall, equates to around one in seven.

ACT responses elicited the highest percentage of queries (20.3%) followed by OT (18.6%) and NT (18.2%), while QLD (12.1%) and VIC (13.2%) had the lowest.

It would be unwise to draw any conclusion relating to the types of Occupations predominating in particular states causing this variation in query. ACT, OT and NT were processed earlier in the processing cycle, when coders were less confident and more likely to raise a query. QLD and VIC were the last states processed, and, all other factors being equal, would have been expected to have the lowest query rate.

4.5 *The Index and Classification*

The processing system attempted to code all Occupation responses to entries in the *Australian Standard Classification of Occupations, Second Edition (cat. no. 1220.0)*.

To facilitate this process, an Occupation Index was created that could store a very broad range of responses, and direct each to specific entries in the Classification. For example, Occupation (Title) responses such as ‘well borer’ and ‘well sinker’ had entries in the Index, which directed any such responses to be coded to 4986-11 Driller.

While this Index was in existence from the previous Census, updates that incorporated new Occupations and variants used to describe Occupations were incorporated on an ongoing basis during 2001 processing.

During the Census, this process was triggered by coders completing Case Reporting Forms (CRFs) when they came across a response to Occupation that was not readily classifiable.

These forms (if the response was new) were then recorded with a new entry and highlight on the Index and reviewed by ABS Classifications Standards staff. The reviewer decided which Classification should apply to the response, and whether the coding process should be AC or CAC.

In preparation for the 2006 Census, the Index will be reviewed and the Classification itself is being revised with the issuing of a new classification, the Australian and New Zealand Standard Classification of Occupations.

4.6 Edits Applied to the Data

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 broad derivation used in the Labour Force Survey, to allow clients to more accurately compare data.

There are two key edits applied to Occupation data:

1. Only persons aged 15 years or over can have their Occupation details coded, and only if
2. they answer 'Yes' to one of the first three options in the 'gateway' question (No. 32: "Last week, did the person have a full-time or part-time job of any kind?"), or did not state an answer to this question.

Two further edits relate to Occupation response and derived Labour Force Status (LFSP):

1. where Occupation was stated as student, child, invalid pensioner, other pensioner, houseperson, retired, unemployed, honorary treasurer, drug dealer or worker's compensation, then set the response to all Labour Force and Occupation variables to Not Applicable and Labour Force Status to Not in the Labour Force (NILF);
2. where Occupation is worked for the dole, then set all Labour Force and Occupation variables to Not Applicable and Labour Force Status to Unemployed Looking for Full-time Work.

These edits are entirely logical and should be retained, as they comply with standard ABS definitions.

4.7 Quality Management and Discrepancy Analysis

4.7.1 The Quality Management Process

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.

During the processing of 2001 Census data, a sample of each coder's work was selected for reprocessing by another coder and any mismatches were looked at by an Adjudicator who would decide on the correct code. If the Adjudicator disagreed with the initial coder, a discrepancy would be recorded. There were 8,298,606 applicable Occupation responses from which 1,458,682 responses (17.6%) were recoded by QM coders. Altogether 78,459 Occupation discrepancies (5.4%) were recorded in the Management Information System (MIS) reports.

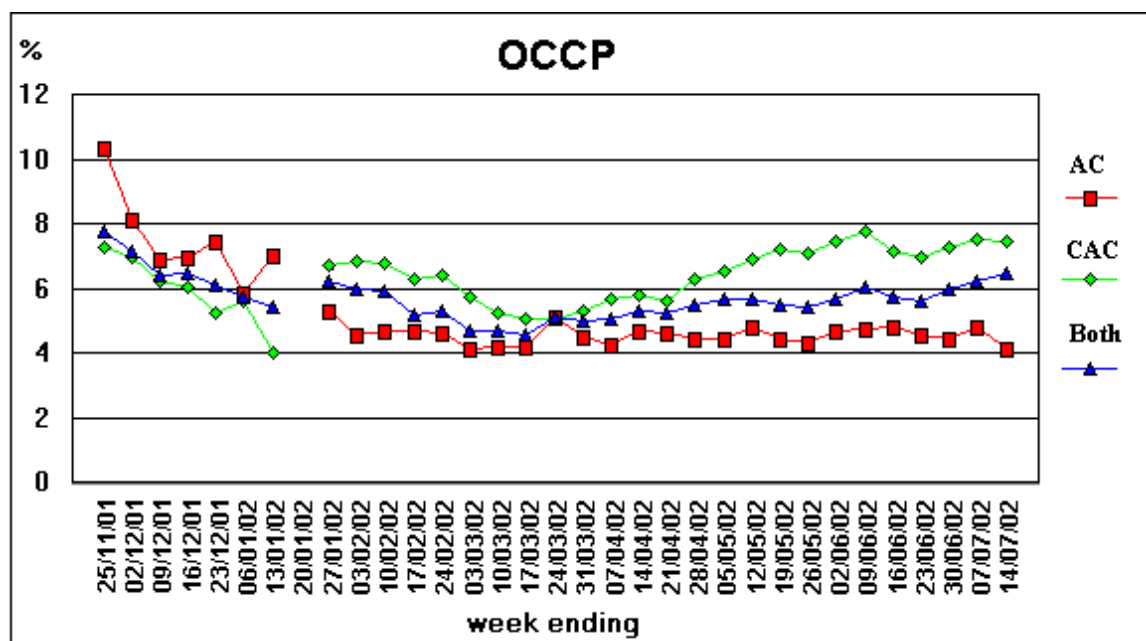
The Quality Management system in place during processing allowed the detection of discrepancies and the calculation of a crude discrepancy rate. This crude discrepancy rate differs from a true discrepancy rate for the following reasons:

- 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 discrepancy would not be detected;
- there is not always an absolutely correct code for every response; and
- discrepancies were also recorded for any QM coder discrepancy;
- Some discrepancies were far more serious than others. For example coding an electrical engineer (code 2125-11) to an electronics engineer (code 2125-13) was given the same weight as coding a tradesperson (Major Group 4) to a professional (Major Group 2).

4.7.2 Discrepancy Rates

Discrepancy rates for Occupation varied across the processing cycle, as shown in Figure 7, below.

Figure 7: Discrepancy Rates for Occupation, by Week Ending, 2001 Census



The initial weeks saw high rates, particularly for AC, as the system was ‘bedded down’ and systemic AC problems were resolved through either blocking of the AC option, or repair of particular letter combinations.

As some previously AC’d combinations were forced to CAC, the latter’s rate rose once more, only to be reduced with time and experience, until coders were encouraged to reduce their frequency of raising Queries and to try to code to the most detailed level possible.

The CAC average of 6% compared favourably with the 11% for 1996 (when all records were manually coded), indicating that improved training and documentation helped coding performance.

4.7.3 Discrepancy Rates by Processing Type

Each different type of processing has a different Discrepancy Rate. In 2001, the distribution across the processes by classification level looked like this:

Table 3: Discrepancy Rate and Distribution at ASCO Level, by Processing Type, 2001 Census

Processing Type	Rate (%)	Discrepancy Distribution (%) by Digit Level				
		1	2	3	4	6
AC	4.6	46.8	9.3	10.5	10.7	22.7
CAC	6.2	44.3	8.9	10.6	8.7	27.4
AC & CAC	5.4	45.4	9.1	10.6	9.6	25.4
QM	5.4	43.1	8.6	10.3	9.8	28.1
QM & CAC	5.7	43.5	8.7	10.4	9.4	27.9

Discrepancy distribution by digit level shows that nearly half of all discrepancies are made at the initial 1-digit level, placing the response into the wrong Occupation Major Group category.

As would be expected, codes at the 6-digit level are more prone to error than 2, 3 or 4, given the finer degree of difference at the Classification's lowest level. It was probably the CAC coder's determination to try to code to the 6-digit level that led to their proportionately higher Discrepancy Rate at the lowest level.

Nevertheless, the distribution of discrepancies was almost identical for AC and CAC, with the overall Discrepancy Rate for AC (at 4.6%) - a slight improvement - being the greatest distinguishing factor.

The similarity in result of the two systems is reasonable as AC can only use the words of a response, such as 'Sales' and 'Manager', the same words that might result in a coder placing an Occupation in an inappropriate Major Group (see Top 10 table below).

AC's advantage is in its ability to act consistently. Given good Repair work (where letter formation for ICR is clarified) and logical programming, AC stands an excellent chance of at least matching the quality of work of the human coder. This, it achieved.

4.7.4 Discrepancy by Group

Comparing 2001 and 1996 distributions is not so easy. 1996 analysis often included discrepancies by QM coders themselves - normally excluded from 2001 analysis. To facilitate a comparison, QM discrepancies for 2001 have been included in the table below.

Table 4: Discrepancies for Occupation by Group Level (1 to 4), 1996 and 2001 Censuses

<i>Group Type</i>	<i>1996</i>		<i>2001 (a)</i>	
	<i>Number</i>	<i>% of All Discrepancies</i>	<i>Number</i>	<i>% of All Discrepancies</i>
1-Digit: Major Group	70,091	68.9	69,645	60.4
2-Digit: Sub-major Group	11,533	11.3	13,911	12.1
3-Digit: Minor Group	10,013	9.8	16,472	14.3
4-Digit: Unit Group	10,046	9.9	15,217	13.2

(a) Includes QM discrepancies.

The above table gives the impression (with gross discrepancy numbers roughly equivalent) that discrepancy rates were similar. This is not the case. The 1996 discrepancy sample was 6.8% of applicable records, while that of 2001 was 17.6%. Excluding QM discrepancies, the overall Discrepancy Rate improved significantly, almost halving, from 10.7% in 1996 to 5.4% in 2001.

Sub-major, Minor and Unit Group Levels featured proportionately more discrepancies in 2001.

If figures for the 6-digit level were to be included (and the breakdown at this level was not available in 1996), the Major Group proportion of discrepancies in 2001 would have been reduced to 44.3% and those at the additional 6-digit level, 26.8%.

Given that discrepancies at the Major Group level are the more serious and that discrepancy rates, overall, nearly halved, 2001 coding can be said to have been significantly better than its 1996 counterpart - even though there was little change in the discrepancy balance across the Groups.

Table 5: Discrepancy Counts for Occupation by Major Group, 2001 Census

	2001	
	<i>Number</i>	<i>% of All Discrepancies</i>
<i>Occupation Major Group</i>		
Managers and Administrators	8,235	12.4
Professionals	7,322	11.1
Associate Professionals	11,654	17.6
Tradespersons and Related Workers	4,212	6.4
Advanced Clerical and Service Workers	2,965	4.5
Intermediate Clerical, Sales and Service Workers	13,492	20.4
Intermediate Production and Transport Workers	4,553	6.9
Elementary Clerical, Sales and Service Workers	7,315	11.1
Labourers and Related Workers	6,371	9.6
Sub-total	66,119	
Inadequately Described	2,845	
Not Stated	681	
Total All	69,645	

The highest proportion of discrepancies occurred in the Intermediate Clerical, Sales and Service Worker (20.4%) and the Associate Professionals (17.6%) Major Groups.

Further analysis (through Major Group to the second digit) shows these were predominantly coded in error to Intermediate Clerical Workers (within the former Major Group) and Business and Administration Associate Professionals, as well as Managing Supervisors (Sales and Service), within the latter.

4.7.5 Top Ten Discrepancies

Table 6: Top 10 Discrepancies by Occupation, by AC percentage, 2001 Census

<i>6-Digit Occupation</i>	<i>Incorrectly Coded</i>	<i>AC%</i>	<i>Not Coded, in Error</i>	<i>AC%</i>	<i>Net Discrepancies as shown by QM Sample</i>
8211-79 - Sales Assistant nec	2,708	23.8	1,657	39.3	1,051
6211-79 - Sales Representative nec	2,249	48.7	1,167	37.5	1,082
8211-11 - Sales Assistant (food and drink products)	1,885	63.1	1,911	47.8	-26
6111-11 - General Clerk	1,656	5.3	2,725	47.0	-1,156
1231-11 - Sales and Marketing Manager	1,540	57.9	1,577	61.6	-928
8211-15 - Sales Assistants (Other Personal and Household Goods)	1,526	45.1	2,345	41.8	819
9111-79 - Cleaners nec	1,496	0.1	815	20.1	681
9111-11 - Commercial Cleaner	1,323	47.7	1,313	23.4	10
8211-00 - Sales Assistants n.f.d.	1,191	0.1	1,243	33.3	-52
3311-11 - Shop Manager	1,104	32.1	1,208	32.5	-104

The table above, apart from indicating the Occupations subject to the highest discrepancies, quantifies the direction of that discrepancy for the QM Sample.

Positive net discrepancies indicate that more records were coded to a particular classification in error than coded erroneously to another classification. For example, as shown in the table above, 1,051 records were coded to 8211-79 Sales Assistants nec in this manner rather than to other classifications.

The predominance of Sales Assistants in the list indicates the difficulty of determining exactly the variant involved. The relatively homogenous nature of the Top 10 listing for AC and CAC errors also indicates that discrepancies were not uniformly spread or random.

Just within these ten Classifications, there are three Major Groups incorporating the term 'Sales' (1, 6 and 8) and two with 'Manager' (1 and 3).

Training for 2006 could incorporate a focus on differentiating between appropriate Major Groups and lower level groups for variants of 'Sales', 'Managers' and 'Cleaners', potentially reducing discrepancy rates at a number of levels.

While overall discrepancy rates and level distributions for AC and CAC were similar, there were differences in discrepancy rate for individual classifications. The table above shows AC discrepancies were all but non-existent for 8211-00 - Sales Assistants n.f.d and 9111-79 - Cleaners nec, whereas for 8211-11 - Sales Assistant (food and drink products) they were 63.1%.

4.7.6 *Query-related Discrepancies*

A total of 4,326 records (5.5% of discrepancies) should have been sent to query, but were coded to the Classification, deemed Inadequately Described or coded as Not Stated. Of these 1,239 (28.6%) were processed by Automatic Coding.

These figures are far lower than in 1996, when between 23 and 34% of discrepancies were for not raising queries at one of the Classification Levels.

In 1996, raising a query was an involved process that required transcription of details of the record. For 2001, this approach was changed to a press-button option. Raising a query was not perceived as a sign of coder incompetence, however, excessive query-raising was discouraged through feedback when a level of experience had been reached that should have lead to sound decision making. For many responses, the data supplied did not allow the Query Resolution coder to obtain a more detailed outcome than the CAC coder could.

The impact of these varying methods resulted in the lower number and percentage of query-related discrepancies in 2001.

5. SAMPLE DATA

5.1 DQI Sample Analysis

A 2% statistically derived sample of CDs (totalling 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 Investigations (DQIs) were carried out at the DPC, which 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 network.

5.1.1 The Sample

The Occupation DQI focused on non-standard, or unusual, responses to the Occupation questions.

A total of 366,667 records were in the DQI Sample. Of these 53.4% were not applicable for the Occupation question. Of the balance, 166,112 (97.1%) made a response that was standard and beyond the focus of the DQI.

It is the analysis of the remaining 4,889 records (2.9%) that are featured in the table below:

Table 7: DQI Sample for Occupation, by Full-time/Part-time Job, 2001

<i>Occupation</i>	<i>Response Option to Full-time/Part-time Job (a)</i>						<i>Total Number</i>	<i>%</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>NS</i>		
Total in DQI Sample							366,667	
Responded in Sample (and applicable)	155,678	5,353	3,309	681	3,976	2,004	171,001	100.0
Standard responses	154,529	5,295	2,985	495	1,164	1,644	166,112	97.1
Non-standard responses:	1,149	58	324	186	2,812	360	4,889	2.9
Student	134	10	23	8	558	100	833	0.5
Retired	66	4	22	12	907	95	1,106	0.6
Casual	48	5	3	0	4	1	61	0.0
Volunteer/Charity	2	3	0	34	34	6	79	0.0
Carer	439	8	8	16	50	10	531	0.3
Home Duties/Mother/Father/Housewife	134	14	190	94	1,141	118	1,691	1.0
Other Unpaid Work	4	1	29	6	32	3	75	0.0
Other Non-standard	322	13	49	16	86	27	513	0.3

(a) See *Figure 1* for detail on options 1 to 5. *NS* is Not Stated.

The single largest non-standard response for Occupation/Tasks and Duties related to Home Duties/Mother/Father/Housewife. They constituted 1% of all responses to Occupation in the DQI Sample, though further investigation has revealed that 67% of the 1,691 in fact selected 'No, did not have a job' - the fifth option from the Full-time/Part-time Job, gateway question.

This meant that they would have been sequenced to Qn 42 (Actively Looking for Work). Similarly, a further 5.6% of the 1,691 responded to FPJP with the fourth option 'Yes, other unpaid work', and were also sequenced to Qn 42.

Only 7.9% of those stating Home Duties etc., answered FPJP with the option 1: 'Yes, worked for payment or profit'.

Still, 26.5% of those who were sequenced out of answering Occupation by their response to FPJP, but still answered Occupation, wrote Home Duties etc., while 19.7% wrote that they were 'Retired'.

It should be noted that 82.7% of the Carers in the DQI stated that they worked for payment or profit. Some of these may be professional Carers, not just those paid an allowance by government to care for a friend or relative. There is not enough information from the DQI to clarify their situation.

The fact that 1% of all responses to Occupation included Home Duties etc., 0.7% Retired and 0.3% Carers, indicates either confusion or a mild degree of protest or frustration from groups who wish to see their preoccupation reflected in the Census format.

Specific instructions in the Full-time/Part-time Job question, directed at those who are exclusively involved in Home Duties/Parenting, Caring for a friend or relative or retirees, may help reduce the level of unnecessary reporting of Occupation by these groups.

Further attention to the sequencing (arrow) and instruction, may be required to minimise respondent error.

A separate question on 'unpaid work' is being tested for inclusion in the 2006 Census.

Possibly, this could be added after FPJP, for those who answer with the fourth or fifth option, (which would allow for multiple marking as the three categories are not mutually exclusive) featuring *Home Duties*, *Carer* and *Retired* options, as well as *None of the above*.

6. FINAL DATA ANALYSIS

6.1 Non-response

The questions about Occupation data were only applicable to the 8,298,606 persons (excluding Overseas Visitors) who were fifteen years or over, and were employed (answered one of the first three ‘Yes’ options to Qn 32) in 2001. If this was the case and if Occupation (Qn 34) and Task and Duties (Qn 35) were left unanswered, a code for ‘Not Stated’ was assigned.

Note that non-response to the Occupation question alone, was not enough to be classified as ‘Not Stated’ for Occupation as a topic.

Table 8: Non-response Rate for Occupation, 1996 and 2001 Censuses

	1996		2001	
	Number	%	Number	%
Not Stated for Occupation	128,595	1.7	98,829	1.2

The relative performance of Occupation Non-response is shown in Table 9.

Table 9: Non-response Rates for Occupation-related Census Questions, 1996 and 2001

<i>Question (with 2001 Qn Number)</i>	1996		2001	
	Number	%	Number	%
Job Last Week (Q33)	168,246	2.2	111,870	1.4
Occupation (Qns 34 and 35)	128,595	1.7	98,829	1.2
Industry Sector (Qn 36)	183,064	2.4	202,177	2.4
Industry of Employment (Qns 38 and 39)	151,739	2.0	144,613	1.7
Hours Worked (Qn 40)	169,430	2.2	248,204	3.0
Method of Travel to Work (Qn 41)	138,171	1.8	152,129	1.8

As can be seen from the list above, Occupation has the lowest Non-response rate of all the employment related variables.

It cannot be compared with variables that should be completed by the whole population, like Birthplace of Individual (5.5%), nor the Occupation ‘gateway’ question Full-time/Part-time Job (2.4%), which need only be completed by those 15 years of age or older.

Even amongst variables with the same population (all employed persons), as in Table 9, Occupation has the highest response rate. It does have the advantage of requiring a non-response to both questions 34 and 35 to be coded to Not Stated - though this is equally true for Industry of Employment, with a rate of 1.7%.

6.1.1 Characteristics of Non-respondents

Table 10: Occupation by Stated/Not Stated, by Sex, Income, Qualification & Birthplace, percent for applicable population

Variable	OCCP (%)			
	Stated		Not Stated	
	No.	%	No.	%
Sex:				
Male	4,494,864	98.9	51,919	1.1
Female	3,704,913	98.7	46,910	1.3
Age:				
15 to 19	537,843	97.7	12,383	2.3
20 to 29	1,778,632	99.0	17,802	1.0
30 to 39	2,011,951	99.1	18,636	0.9
40 to 49	2,054,130	99.0	18,337	0.9
50 to 59	1,405,706	97.6	14,675	1.0
60 to 69	347,665	90.3	8,637	2.4
70 to 79	52,673	79.2	5,681	9.7
80 to 89	8,776	79.2	2,306	20.8
90 to 99	2,145	87.2	316	12.8
100 and over	256	82.1	56	17.9
Income:				
Negative	27,659	96.1	1,129	3.9
Nil	37,358	89.7	4,289	10.3
\$1-399	2,104,830	98.2	38,590	1.8
\$400-999	4,351,262	99.4	28,283	0.6
\$1,000 or more	1,503,866	99.6	5,685	0.4
Not Stated	174,802	89.3	20,853	10.7
Qualification:				
Deg or Higher	1,547,556	99.7	5,319	0.3
Adv Dip/Dip	639,573	99.5	3,148	0.5
Certificate	1,672,972	99.3	11,743	0.7
Inad Desc	11,847	99.2	994	0.8
Level NS	347,127	94.8	19,066	5.2
No Qual	3,874,102	98.5	58,559	1.5
Birthplace:				
Australia	6,091,234	98.9	67,167	1.1
Overseas	1,985,645	98.7	26,561	1.3
Not Stated	122,898	96.0	5,101	4.0

Persons who did not state their Occupation were more likely to be over 60 years of age, have Nil or Not Stated Income and be Not Stated to education level. In terms of Sex and Birthplace, there was very little difference between respondents and non-respondents.

6.2 Not Further Defined Coding

6.2.1 Description

The principles of coding to ASCO required responses given on Census Forms to be coded to the most detailed level of the classification possible. If the response was not detailed enough to allow coding to the 6-digit level, a Not Further Defined (NFD) code was allocated. The coding was structured as follows:

- the Occupation level (for example 3491-11) called the 6-digit level; or
- the NFD category of the unit group to which it belonged (3491-00) called the 4-digit level; or
- the NFD category of the minor group to which it belongs (3490-00) called the 3-digit level; or
- the NFD category of the sub-major group to which it belongs (3400-00) called the 2-digit level; or
- the NFD' category of the major group to which it belongs (ie 3000-00) called the 1-digit level; or
- the Inadequately Described category.

When a code other than the Occupation level (6-digit) is allocated, this is referred to as NFD coding. Major reasons why NFD coding occurs are:

- the level of information provided by respondents is not detailed enough;
- the response is made in a 'colloquial' form familiar to the respondent, but not present in the Index or the formally structured Classification;
- poor written language skills enable only the broadest interpretation of the response;
- multiple responses in the forms cause the system to code to a higher code so that fine level information is lost. For example a manager describing his or her tasks as managing both building construction (ASCO code 1191-11) and engineering (ASCO code 1221-11) would be allocated the NFD code for the major group 'Managers and Administrators'; and
- coders may not follow correct procedures to classify the response given, or may not use all the information in the forms.

6.2.2 Comparison between 1996 and 2001 NFD data

The table below compares 1996 and 2001 NFD counts from Levels 1 to 4, as well as the balance remaining at the lowest, 6-digit level of coding. The difference is expressed in percentage points:

Table 11: Distribution of Not Further Defined Responses in the 1996 and 2001 Censuses

<i>ASCO, Second Edition major groups</i>		<i>% of Responses Coded To NFD by Level:</i>					<i>Occupation Group (6-digit)</i>	<i>Total persons</i>
		<i>Major Group (1-digit)</i>	<i>Sub-major Group (2-digit)</i>	<i>Minor Group (3-digit)</i>	<i>Unit Group (4-digit)</i>			
Managers and Administrators	2001	5.5	0.8	1.3	2.5	89.9	764,823	
	1996	10.9	1.2	6.1	4.6	77.3	709,925	
	Diff (a)	-5.4	-0.4	-4.8	-2.1	12.6	54,898	
Professionals	2001	1.3	0.4	3.1	2.6	92.4	1,514,096	
	1996	1.5	0.3	4.0	1.6	92.6	1,309,468	
	Diff	-0.2	0.1	-0.9	1.0	-0.2	204,628	
Associate Professionals	2001	0.3	0.8	1.9	0.4	96.6	975,653	
	1996	0.7	1.1	2.0	0.4	95.9	861,169	
	Diff	-0.4	-0.3	-0.1	0	0.7	114,484	
Tradespersons and Related Workers	2001	1.3	0.3	0.9	2.2	95.3	1,018,903	
	1996	1.5	0.1	0.5	2.6	95.3	997,010	
	Diff	-0.2	0.2	0.4	-0.4	0	21,893	
Advanced Clerical and Service Workers	2001	0.3	0.1	0.1	0.2	99.4	309,968	
	1996	0.1	0.0	0.0	0.1	99.9	329,844	
	Diff	0.2	0.1	0.1	0.1	-0.5	-19,876	
Intermediate Clerical, Sales and Service Workers	2001	0.3	0.4	0.9	2.3	96.1	1,366,701	
	1996	0.4	0.3	0.5	3.5	95.3	1,222,735	
	Diff	-0.1	0.1	0.4	-1.2	0.8	143,966	
Intermediate Production and Transport Workers	2001	1.5	5.0	2.7	2.0	88.8	670,821	
	1996	2.3	10.9	3.4	1.8	81.6	661,425	
	Diff	-0.8	-5.9	-0.7	0.2	7.2	9,396	
Elementary Clerical, Sales and Service Workers	2001	0.3	0.2	0.3	4.0	95.2	792,378	
	1996	0.4	0.5	0.4	12.3	86.5	677,395	
	Diff	-0.1	-0.3	-0.1	-8.3	8.7	114,983	
Labourers and Related Workers	2001	3.7	2.0	0.8	6.2	87.3	717,457	
	1996	7.2	0.3	1.6	5.8	85.2	667,250	
	Diff	-3.5	1.7	-0.8	0.4	2.1	50,207	
Not Stated	2001						98,829	
	1996						128,595	
	Diff						-29,766	
Inadequately Described	2001						68,977	
	1996						71,503	
	Diff						-2,526	
Total	2001						8,298,606	
	1996						7,636,319	
	Diff						662,287	

(a) Difference in percentage points.

In 2001, 99.4% of Advanced Clerical and Service Workers were coded to the most detailed level (with the balance to broader, NFD levels). While this is a high number, the coding in 1996 was even more detailed. Only this Major Group and Professionals were greater in total in NFDs in 2001.

Labourers and Related Workers (12.7%), Intermediate Production and Transport Workers (11.3%) and Managers and Administrators (10.1%), had the highest 2001 levels, in total, of NFDs.

Despite a reduction of nearly 50% in the number of Managers and Administrators left at the 1-digit level, 5.5% of this group remained at that broadest level. This made them by far the largest of all Major Groups not more definitively coded.

In 2001, there were 662,287 (8.7%) more employed persons, but marginally less in the Inadequately Described category (68,977 in 2001 compared to 71,503 in 1996). The rise in employed was spread across all Major Groups except one - Advanced Clerical and Service Workers, which actually fell by 19,876.

Table 11 also displays percentage point change from 1996 and this best shows progress towards definitive coding:

- Overall, for the Managers and Administrators major group, there was a 12.6 percentage point increase in more detailed coding, beyond NFDs.
- The fact that two other groups (Elementary Clerical, Sales and Service Workers and Intermediate Production and Transport Workers) also increased significantly (by 8.7 and 7.2 percentage points respectively) in more specific coding, indicates that form design changes alone were not fully responsible. Other factors such as better coding instruction and query support are likely to have assisted coders to code more Occupation responses to the Occupation level.
- The significant reduction in 1 and 3-digit NFD group percentages for Managers and Administrators would to a large degree reflect the form change that included a farming example encouraging differentiation between the type of farming undertaken (see 2.2 *Differences Between the 2001 and 1996 Forms*).
- While the number of employees coded to a Farmer classification decreased in 2001 by 2.1% (to 194,883), those coded to the most detailed, 6-digit Occupation level, increased by 30.4% (to 175,250).
- Managers other than farmers increased in number by 11.5% from 1996, though their more detailed coding only rose by 23.7% - clearly showing that the extra wording for farmers on the form had its desired effect (see 6.3 *Case Study*).

6.3 Case Study

The main change in form design (see 2.2 *Differences between the 2001 and 1996 Forms*), was the addition of ‘Sheep and Wheat Farmer’ as an example of Occupation.

Under Tasks or Duties, the additional wording added was ‘running a sheep/wheat farm’.

The Occupation question might be interpreted as implying mixed farming (doing both), while the tasks wording indicates one or the other.

It is interesting that the word ‘wheat’ does not appear anywhere in ASCO - neither as an Occupation title for a farmer, nor in any detailed explanation of possible category contents. ‘Wheat Farmer’ does feature in the Index, where it is automatically coded to 1313-11 - Grain, Oilseed and Pasture Grower.

To examine whether the changed wording may have had any positive impact on Occupation coding, the breakdown of counts at the 3-digit, 4-digit and the 6-digit (most detailed) level, need to be examined and compared with those for the same groupings in 1996.

Table 12: Percentages for Farmers and Farm Managers by Group Level, 1996 and 2001 Censuses

<i>Group Type</i>	<i>1996</i>		<i>2001</i>	
	<i>Number</i>	<i>%</i>	<i>Number</i>	<i>%</i>
To 3-Digit only:				
1310-00 Minor Group: Farmers and Farm Managers nfd	41,409	20.8	7,658	3.9
To 4-Digit only:				
1312-00 Livestock Farmers nfd	18,392		8,152	
1313-00 Crop Farmers nfd	4,795		3,823	
Total	23,187	11.7	11,975	6.1
To 6-Digit only:				
1311-11 Mixed Crop and Livestock Farmer	34,956		44,459	
1312-11 to 1312-79 (farming various animals)	80,075		88,267	
1313-11 to 1314-11 (farming various crops)	42,553		54,499	
Total	134,397	67.5	175,250	89.9
Total All Coded to Minor Group 131	198,993		194,883	

Despite a marginal decline in farm manager numbers overall (down 4,110), the percentage coded to the 6-digit Occupation level rose by over 22 percentage points and by 40,853 persons. Mixed Crop and Livestock Farmer (1311-11), the classification most likely to benefit from the ‘Sheep and Wheat Farmer’ example, rose 27.2%, from 34,956 in 1996 to 44,459. On the surface, this appears to have overwhelmingly justified the wording changes, aimed at more definitive coding.

Given that AC rates for Managers and Administrators were lower than for any other Major Group (see 4.2 *Automatic Coding*), AC's part in this positive change is less than might have been presumed.

Table 13: Percentages for Occupation by ASCO Level, 1996 and 2001 Censuses

<i>ASCO Type</i>	<i>1996</i>		<i>2001</i>	
	<i>Number</i>	<i>%</i>	<i>Number</i>	<i>%</i>
1-Digit: Major Group	188,999	2.5	121,340	1.5
2-Digit: Sub-major Group	104,132	1.4	78,502	1.0
3-Digit: Minor Group	118,120	1.6	123,145	1.5
4-Digit: Unit Group	300,753	4.0	205,084	2.5
6-Digit: Occupation	6,724,224	90.4	7,602,731	93.5
Total Coded to ASCO (a)	7,436,228		8,130,802	

(a) Excludes the 71,503 and 68,977 coded to Inadequately Described in the 1996 and 2001 Census, respectively.

The table above shows that Occupation was more definitively coded in 2001 than in 1996, with 93.5% coded to the most detailed level. At each of the four broadest level groups, 1996 percentages (if not populations) were greater, while at the Occupation level, 2001 had greater than 3 percentage points more.

While this seems to be a further positive change, a comparison of the Discrepancy Rate figures for both Censuses will provide confirmation (see 4.7 *Quality Management and Discrepancy Analysis*).

For a breakdown by Major Group by the various Digit levels, see 6.2 *Not Further Defined Coding*.

7. RECONCILIATION OF 2001 CENSUS OCCUPATION DATA WITH AUGUST 2001 LABOUR FORCE SURVEY DATA

7.1 *Data Reconciliation Methodology*

The purpose of this section is to explain the differences in the collection of Occupation data between the August 2001 Labour Force Survey and the 2001 Census, to outline the steps taken to reconcile these two data collections and to present the findings from this reconciliation. The methodology used to reconcile Census and Labour Force Survey data is based on an internal paper called *Comparing Labour Force Survey and Population Census Data*, prepared by the ABS' Labour Force Section and that of Census Development and Field Organisation, in January 1998.

Although the Census and Labour Force Survey both collect data on Occupation, they are not strictly comparable due to differences in the scope, coverage, timing, measurement of underlying labour force concepts and collection methodology. Factors contributing to differences in estimates include under-enumeration in the Census for which Census Occupation data have not been adjusted, the use in the Labour Force Survey of population benchmarks derived from incomplete information about population change, differing methods of adjustment for non-response to the Survey or Census, the personal interview approach adopted in the Survey as opposed to self-enumeration in the Census, and sampling variability. State comparisons are affected by the unit of output: State of Enumeration for Census and State of Usual Residence for the Labour Force Survey.

Differences in the underlying definition of 'employed' between the two collections should also be borne in mind when comparing figures. Census questions are not as detailed, nor as comprehensive as the Labour Force Survey questions. This is largely due to space limitations on the Census Form, as well as constraints imposed by self-enumeration. The differences in definition of 'employed' between the two collections relate specifically to absences from work. To determine the labour force status of persons absent from work without pay, the Survey applies a test of duration of absence from work. Therefore, a respondent who had been away from work for four weeks or more without pay is regarded as not employed. By contrast, the Census does not apply tests of duration for absence from work, and as a result, all persons away from work are most likely to be classified as employed. This of course depends on how the respondent has completed the Census Form. As a consequence, a proportion of Census respondents would be regarded as employed by the Census whereas these same respondents would be regarded as unemployed or not in the labour force by the Labour Force Survey. As there is no clear way of identifying the Occupation of persons classified as employed by the Census but unemployed or not in the labour force by the Survey, it is not possible to remove this population from Census data.

[For further background information on the Census and the Labour Force Survey, see *Labour Statistics: Concepts, Sources and Methods, 2001* (cat no. 6102.0).]

To enable reconciliation, the scope of both the 2001 Census and the August 2001 Labour Force Survey were first reduced to a 'common', broad population. *Table 14* (below), shows the adjustments made to August 2001 Labour Force Survey benchmarks and to the 2001 Census, for Occupation data comparison of those 15 years or older and 'employed'.

Table 14: Adjustments made to August 2001 Labour Force Survey Benchmarks and 2001 Census to derive a Common Population for Occupation Data

<i>Population group</i>	<i>Deducted from Labour Force</i>	<i>Deducted from Census Counts</i>
Jervis Bay Territory and external territories		1,145
Defence Force Personnel		61,139
Not enumerated in Census (a)	289,777	
Residents temporarily overseas (a)	302,323	
Inadequately Described (a)	68,941	
Not stated for Occupation (a)	98,808	

(a) Excludes Other Territories, to balance with Labour Force Survey

7.2 Deductions from Census Counts

As the Labour Force Survey excluded Other Territories and Defence Force Personnel, these groups had to be identified and tables created by State and also by Age Group, to remove them from the Census Occupation Major Group counts.

The Other Territories component (Jervis Bay and external territories) of the Census count (1,145) needed to be removed.

To uncover the Defence Force component, Census Occupation counts were cross-classified with Industry and those specifically in 'Defence' (61,139), were excluded.

Each of these actions were taken for the respective populations by ASCO Major Group and by State and Age Range.

The resulting totals were deducted from Census counts to leave the figures in Tables A1 and A3 in *Appendix 1*.

7.3 Deductions from Labour Force Survey Counts

As Labour Force Survey Occupation counts were based on full 'State' estimates, those in Australia but not enumerated in the 2001 Census (known as the Undercount) and Residents Temporarily Overseas had to be excluded from the Labour Force population. ABS Demography provided a breakdown of these numbers, by State (excluding Overseas Territories) and by Age Group.

Two groups who had stated in the Census that they were employed were also excluded from Labour force counts as their data could not be matched to the ASCO Major Group Classification. Those groups were: Not Stated for Occupation, who indicated they were employed, but did not answer either of the two Occupation questions (see *6.1 Non-response*); and the Inadequately Described who responded, but not clearly enough to be coded to a Major Group.

The revised August 2001 Labour Force figures, excluding those four groups mentioned above, are shown in the Tables A2 and A4 in *Appendix 1*.

7.4 Results of Data Reconciliation

It should be noted that any attempt to find a common population from the two sources is unlikely to arrive at one figure, due to differences such as those described previously (in 7.1 *Data Reconciliation Methodology*).

The Labour Force August 2001 estimate of the civilian population aged 15 and over was 15,442,000. This was 2.7% higher than the raw, unadjusted Census count of 15,038,339. If adjustments such as those identified in *Table 14* were to be made, the Census count would in fact be higher. After deducting those elements outlined in 7.2 and 7.3, the total population for Occupation according to Labour Force is 3.7% larger than the Census count of 8,068,516. In 1996, this gap was 3.0%.

As Labour Force Occupation figures are larger than Census ones, this indicates that there is a greater tendency to employment in Labour Force figures generally. This should be kept in mind when reviewing the comparison, which is best viewed at the broader Major Group percentage level.

7.4.1 Data Comparison of Occupation Major Groups by Age, 2001 and 1996

The following analyses are based on comparisons of tables *A1*, *A2*, *A3* and *A4* in *Appendix 1*:

Table 15: Occupation Major Groups by Age, 2001 Census as a Percentage of August 2001 Labour Force Survey for Persons, Australia

Occupation major group	Age Group						Total
	15-19	20-24	25-34	35-44	45-54	55 and over	
	%						
Managers and Administrators	213.9	133.0	123.4	115.0	112.9	124.4	118.4
Professionals	106.1	92.5	95.7	93.5	96.4	105.8	96.1
Associate Professionals	121.4	104.2	97.9	98.9	95.9	103.9	99.2
Tradespersons and Related Workers	95.5	85.8	88.9	95.1	93.3	117.8	93.6
Advanced Clerical and Service Workers	99.7	82.1	81.7	82.0	79.3	97.4	83.2
Intermediate Clerical, Sales and Service Workers	83.6	87.1	93.8	95.9	96.3	116.6	94.7
Intermediate Production and Transport Workers	81.2	91.7	87.3	84.6	89.6	127.1	90.6
Elementary Clerical, Sales and Service Workers	86.9	83.7	100.9	99.5	106.4	136.1	96.4
Labourers and Related Workers	77.8	88.8	88.8	99.3	103.0	111.0	94.7
Australia	87.2	89.6	94.7	96.1	97.3	114.7	96.5

Census counts for Managers and Administrators generally, were 118.4% of the Labour Force version. This, along with the age group 55 and Over's average of 114.7% and the higher counts for Professional and Associate Professional categories amongst 15-19 year olds were the main areas where Census exceeded Labour Force. This may have had something to do with Census responses being largely the respondent's interpretation - something that the interview situation of the Labour Force Survey could question and clarify.

The relatively higher Census counts for those aged 55 and over probably reflects the more relaxed criteria for employment from a Census perspective and a greater willingness on the part of respondents to identify casual or part-time work as employed - whereas Labour Force interprets an absence of four weeks or more without pay as being not employed. This is likely to have limited the count of those in that age grouping in the LFS and therefore their numbers at each Major Group level.

If distributions across the table were to be proportionate, all Census percentages should reflect the national average of Census being 96.5% of Labour Force. In total, Elementary Clerical, Sales and Service is virtually that, but when viewed by age, there are significant variations. Advanced Clerical and Service Workers recorded the lowest relative percentage. This can in part be explained by the relative excess of Managers and Administrators, indicating there was a degree of bias that did not occur in the same direction in the Labour Force Survey.

The relative positions of the Census and the Labour Force Survey are not a 2001 phenomenon. 1996 comparison figures reveal a fairly similar story:

Table 16: Occupation Major Groups by Age, 1996 Census as a Percentage of August 1996 Labour Force Survey for Persons, Australia

<i>Occupation major group</i>	<i>Age Group</i>						<i>Total</i>
	<i>15-19</i>	<i>20-24</i>	<i>25-34</i>	<i>35-44</i>	<i>45-54</i>	<i>55 and over</i>	
	%						
Managers and Administrators	323.9	209.4	138.6	119.3	118.1	105.6	121.7
Professionals	125.6	102.3	100.9	99.2	104.0	102.6	101.4
Associate Professionals	117.6	109.9	109.8	106.4	100.7	105.3	106.2
Tradespersons and Related Workers	104.9	93.8	91.8	90.5	93.8	98.1	93.5
Advanced Clerical and Service Workers	95.9	84.2	94.1	86.7	83.8	87.1	87.9
Intermediate Clerical, Sales and Service Workers	91.4	93.9	92.9	97.7	93.2	91.1	94.1
Intermediate Production and Transport Workers	87.1	80.5	93.3	91.3	93.1	88.0	90.4
Elementary Clerical, Sales and Service Workers	81.5	81.9	87.7	81.9	89.7	90.6	84.6
Labourers and Related Workers	65.4	92.4	85.8	89.7	97.6	86.6	86.4
Australia	85.3	94.5	98.3	97.9	99.6	98.0	97.1

The similarities with 2001 analysis broadly validates the 2001 Census Occupation data. A key difference was the 55 and Over age group, which in 1996 was lower than its Labour Force counterpart.

7.4.2 *Comparison of Occupation Major Groups by Age, 2001 Census and August 2001 Labour Force*

Perhaps the greatest area of concordance between the two measures of Occupation can be seen in Major Group percentage at the national level. Comparison of the Total columns from the following two tables, reveals only very marginal differences - further indication of the acceptability of the data.

Table 17: Percentage Rates for Occupation Major Groups by Age, Persons, Australia, 2001 Census

<i>Occupation major group</i>	<i>Age Group</i>						<i>Total</i>
	<i>15-19</i>	<i>20-24</i>	<i>25-34</i>	<i>35-44</i>	<i>45-54</i>	<i>55 and over</i>	
	%						
Managers and Administrators	0.1	0.3	1.8	2.7	2.7	1.9	9.4
Professionals	0.1	1.4	5.1	5.2	4.7	2.1	18.6
Associate Professionals	0.3	1.0	3.1	3.3	3.0	1.4	12.0
Tradespersons and Related Workers	0.9	1.6	3.2	3.1	2.4	1.2	12.4
Advanced Clerical and Service Workers	0.1	0.3	1.0	1.0	1.0	0.5	3.8
Intermediate Clerical, Sales and Service Workers	1.1	2.4	4.1	4.2	3.6	1.5	16.9
Intermediate Production and Transport Workers	0.5	0.7	1.9	2.3	2.0	1.1	8.3
Elementary Clerical, Sales and Service Workers	2.6	1.6	1.7	1.7	1.5	0.9	9.8
Labourers and Related Workers	1.1	1.0	1.8	2.1	1.9	1.1	8.9
Total	6.6	10.1	23.6	25.5	22.7	11.6	100.0

Table 18: Percentage Rates for Occupation Major Groups by Age, Persons, Australia, August 2001 Labour Force Survey

<i>Occupation major group</i>	<i>Age Group</i>						<i>Total</i>
	<i>15-19</i>	<i>20-24</i>	<i>25-34</i>	<i>35-44</i>	<i>45-54</i>	<i>55 and over</i>	
	%						
Managers and Administrators	0.0	0.2	1.4	2.3	2.3	1.5	7.6
Professionals	0.1	1.5	5.2	5.3	4.7	1.9	18.7
Associate Professionals	0.2	0.9	3.0	3.2	3.0	1.3	11.6
Tradespersons and Related Workers	0.9	1.8	3.5	3.1	2.5	1.0	12.8
Advanced Clerical and Service Workers	0.1	0.4	1.1	1.2	1.2	0.5	4.5
Intermediate Clerical, Sales and Service Workers	1.3	2.6	4.2	4.2	3.6	1.3	17.1
Intermediate Production and Transport Workers	0.5	0.7	2.1	2.6	2.1	0.8	8.8
Elementary Clerical, Sales and Service Workers	2.9	1.8	1.6	1.6	1.4	0.6	9.8
Labourers and Related Workers	1.3	1.1	2.0	2.0	1.8	0.9	9.1
Total	7.3	10.9	24.0	25.6	22.5	9.7	100.0

As in 1996, the Major Group ‘Managers and Administrators’ recorded the largest percentage rate difference, with 9.4 per cent for the Census and 7.6 per cent for the LFS. Census rates were higher across every age range in this Group..

Main age group differences were in the ranges 15-19 years (Census 6.6% and LFS 7.3%) and 55 and Over (Census 11.6% and LFS 9.7%).

Differences between figures in the collections were statistically minor, with the percentage rates comparison showing an overall similarity in the distribution of data.

7.4.3 Comparison of Occupation Major Groups by States and Territories, 2001 Census and August 2001 Labour Force

Tables A3 and A4 in *Appendix 1* provide adjusted figures by States and Territories for both collections. The percentage rates in *Tables 19* and *20* have been calculated as proportions of the total number of persons employed in the labour force in each State and Territory.

Table 19 : Percentage Rates for Occupation Major Groups by States and Territories, Persons, 2001 Census

Occupation major group	States and Territories (%)							
	NSW	Vic.	Qld	SA	WA	Tas.	NT	ACT
Managers and Administrators	9.6	9.7	8.7	9.6	8.8	8.8	8.5	10.5
Professionals	19.6	19.6	16.4	17.1	17.4	17.2	18.3	26.7
Associate Professionals	11.8	11.6	12.2	11.7	12.5	11.8	13.9	14.4
Tradespersons and Related Workers	12.0	12.4	12.8	12.4	13.5	12.7	12.1	8.3
Advanced Clerical and Service Workers	4.3	3.7	3.6	3.4	3.9	2.9	3.4	3.5
Intermediate Clerical, Sales and Service Workers	16.9	16.5	17.3	16.9	16.4	17.1	16.7	18.6
Intermediate Production and Transport Workers	8.1	8.3	8.7	8.5	8.7	9.2	7.5	3.8
Elementary Clerical, Sales and Service Workers	9.6	9.8	10.4	9.5	9.7	10.2	8.8	10.0
Labourers and Related Workers	8.2	8.5	10.0	10.9	9.0	10.0	11.0	4.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 20: Percentage Rates for Occupation Major Groups by States and Territories, Persons, August 2001 Labour Force Survey

<i>Occupation major group</i>	<i>States and Territories (%)</i>							
	<i>NSW</i>	<i>Vic.</i>	<i>Qld</i>	<i>SA</i>	<i>WA</i>	<i>Tas.</i>	<i>NT</i>	<i>ACT</i>
Managers and Administrators	7.6	8.5	6.7	8.5	6.5	8.00	5.5	8.5
Professionals	19.6	20.1	15.8	17.1	17.5	16.3	19.6	27.9
Associate Professionals	11.0	11.5	12.1	11.9	12.4	11.9	14.7	13.2
Tradespersons and Related Workers	12.2	12.9	13.0	12.5	14.7	13.7	10.8	8.8
Advanced Clerical and Service Workers	4.9	4.1	4.7	3.9	4.4	2.8	2.5	4.0
Intermediate Clerical, Sales and Service Workers	18.3	16.3	17.2	16.0	16.1	17.9	17.0	18.1
Intermediate Production and Transport Workers	8.7	8.9	9.3	9.2	8.5	9.2	8.4	5.0
Elementary Clerical, Sales and Service Workers	9.2	9.9	10.6	9.8	10.2	9.0	9.7	10.3
Labourers and Related Workers	8.6	7.8	10.6	11.1	9.7	11.2	11.9	4.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Only in three cases was there a percentage point difference of more than 2. Each of these was in the Manager and Administrator Major Group where NT (8.5% compared to 5.5%), WA (8.8% to 6.5%) and ACT (10.5% to 8.5%) all displayed the extra Census bias referred to earlier.

It could well be that sampling variability for the LFS exacerbated the differences which were only marginal elsewhere - though the greater likelihood of an employee claiming to be a manager in the Census and being unable to support such a claim at an LFS interview, still seems the most likely cause.

8. CONCLUSIONS

This paper has examined the quality of occupation data from the 2001 Census. The conclusions are outlined below:

The decision to modify form design to include ‘Sheep and Wheat Farmer’ as an example, had a dramatic and positive impact on moving classification of responses by farmers to a lower level, with 6-digit farmer classification rising by 30%, though farmer numbers overall declined by 4,110.

Occupation Non-response, at 1.2% was lower than the 1996 figure of 1.7%. As in 1996, Occupation had the lowest Non-response of all Census variables.

Occupation was more definitively coded in 2001 than in 1996. In the 2001 Census, percentages at the 1, 2, 3, and 4-digit level were lower, while at the most detailed Occupation level (6-digit), the 1996 figure of 90.4% was exceeded by 2001’s 93.5%.

Managers and Administrators reduced their 1-digit coding by nearly 50%, but still had the highest percentage (5.5%) not more definitively coded, though 6-digit coding increased by 12.6 percentage points.

Elementary Clerical, Sales and Service Workers had 4-digit coding reduced by 8.3 percentage points, with 6-digit coding rising by 8.7 percentage points.

The use of Automatic Coding for the first time, coding 57.1% of all Stated records, proved marginally more successful than Computer Assisted Coding. Discrepancy rates for AC were 4.6%, while for CAC they were 6.2%. The overall 2001 Discrepancy rate of 5.4% was a significant improvement over the 10.7% of 1996.

Classifications relating to ‘Sales’, ‘Managers’ and ‘Cleaners’ dominated those featuring the highest discrepancy rates.

While overall Discrepancy Rates nearly halved in 2001, there were similar proportional breakdowns at each level of the Classification.

Comparison of the 2001 Census Occupation counts and those from the August 2001 Labour Force Survey revealed a similar relationship to that in 1996. Overall, results generally validated both approaches, though with Census counts for Managers and Administrators and those employed and 55 years of age or over, exceeding that of the Survey. These differences are mostly related to the differing methods of enumeration and definitions of employed.

9. RECOMMENDATIONS

- Given the success of the former changes for 2001, it is suggested that those responsible for form design seriously consider the possibility of including the following examples in the two other Occupation-related questions, to assist in Sales detail and differentiation:
 - * Sales Assistant (Occupation), and
 - * sell food and drink products (Tasks)
- Coder training should emphasise a range of specific examples that differentiate between 'Sales', 'Manager' and 'Cleaner' titles at varying levels of the Classification.
- The DPC's Management Information System (MIS) needs to have greater flexibility to allow 'drilling down' for each variable. Reports should be able to be easily produced that give Discrepancy Rate counts by Level by Processing Type by Occupation Major Group. If possible any Classification should be able to be substituted for Major Group.

Quick access to management information such as this will provide extra knowledge and assistance to those monitoring the Census processing operation - and to those who evaluate the accuracy of the data.

- Investigate methods of recognising 'work' that is not necessarily paid employment.

GLOSSARY

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

ASCO - Australian Standard Classification of Occupations. The Second Edition (released in July 1996), was used to code the 1996 and 2001 Census and August 1996 and August 2001 Labour Force Survey responses.

CAC - Computer Assisted Coding. The process of using procedures and rules to allow a human (manual) coder to match the image of text responses to entries on an index for that topic.

CD - Collection District. The smallest unit for collection, processing and the output of data.

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

DC - Data Capture. The process of scanning Census Forms into the image and text files that are used for all subsequent processes.

Discrepancy Rate - Generally, the rate at which Quality Management and subsequent Adjudication coding differed from initial coding. Expressed as a percentage, it is regarded as the discrepancy rate within final data.

DPC - Data Processing Centre. A centralised facility that was located in Ultimo, Sydney to process 2001 Census forms.

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

Index - the list of responses to Occupation - found on the IUU Db. Individual Index entries are matched to a Classification code.

IUU Db - the Index Utility Update Database. An ABS internal database that contains an alphabetic listing of all different Occupation responses and the code to which each has been assigned.

LFS - Labour Force Survey. Conducted quarterly by the ABS Labour Force Section.

MIS - Management Information System. A DPC-based system that accumulated and output statistics on the progress and quality of the processing operation.

Non-response - failure to answer (in the case of Occupation), Occupation, as well as Tasks and Duties (Questions 34 and 35 on the Household Census Form).

NFD - Not Further Defined. For Occupation, a classification existing at each of the 1 to 4 levels (Major Group, Sub-major Group, Minor Group and Unit Group), containing records that could not be coded to a lower level.

Occupation - employment (full-time or part-time), held by an individual 15 years of age or older.

QM - Quality Management. The process of regular review of a percentage of coding work, though also a term for broader DPC-wide ongoing reviews.

QR - Query Resolution. A specialist group with access to extra resource material, who were used to resolve difficult coding issues.

Repair - where changes are made after initial scanning.

REFERENCES

Australian Bureau of Statistics (1997) *Australian Standard Classification of Occupations (ASCO) -- Statistical Classification, Second Edition* (cat. no. 1220.0)

Census Working Paper 99/6 *1996 Census Data Quality: Occupation*

Labour Force, Australia, August 2001 (cat no. 6203.0)

APPENDIX 1: Reconciliation between 2001 Census and August 2001 Labour Force Survey: More Detail

Table A1 : Adjusted Counts for Occupation Major Groups by Age, 2001 Census

<i>Occupation Major Group</i>	<i>Age group</i>						<i>Total</i>
	<i>15-19</i>	<i>20-24</i>	<i>25-34</i>	<i>35-44</i>	<i>45-54</i>	<i>55 and over</i>	
Managers and Administrators	4,114	21,055	141,415	218,269	218,875	152,579	756,307
Professionals	11,878	114,040	413,086	418,782	377,651	167,172	1,502,609
Associate Professionals	20,646	80,290	248,090	265,108	238,410	112,376	964,920
Tradespersons and Related Workers	73,364	126,912	259,052	249,882	191,383	99,093	999,686
Advanced Clerical and Service Workers	6,478	26,112	77,015	82,968	76,841	40,044	309,458
Intermediate Clerical, Sales and Service Workers	87,700	190,730	330,328	335,222	291,637	123,696	1,359,313
Intermediate Production and Transport Workers	36,334	55,288	150,276	183,515	157,304	84,918	667,635
Elementary Clerical, Sales and Service Workers	206,818	124,791	135,353	133,327	123,025	68,315	791,629
Labourers and Related Workers	84,780	77,878	145,563	169,301	153,460	85,977	716,959
Total	532,112	817,096	1,900,178	2,056,374	1,828,586	934,170	8,068,516

**Table A2 : Adjusted Counts for Occupation Major Groups by Age, August 2001
Labour Force Survey**

<i>Occupation Major Group</i>	<i>Age group</i>						<i>Total</i>
	<i>15-19</i>	<i>20-24</i>	<i>25-34</i>	<i>35-44</i>	<i>45-54</i>	<i>55 and over</i>	
Managers and Administrators	1,923	15,826	114,561	189,814	193,922	122,657	638,703
Professionals	11,192	123,252	431,563	447,800	391,732	157,949	1,563,488
Associate Professionals	17,010	77,073	253,300	267,983	248,724	108,144	972,233
Tradespersons and Related Workers	76,806	147,921	291,351	262,881	205,215	84,127	1,068,302
Advanced Clerical and Service Workers	6,496	31,822	94,301	101,223	96,858	41,103	371,804
Intermediate Clerical, Sales and Service Workers	104,876	219,069	352,269	349,676	302,845	106,068	1,434,802
Intermediate Production and Transport Workers	44,766	60,320	172,139	217,021	175,544	66,806	736,597
Elementary Clerical, Sales and Service Workers	237,972	149,030	134,134	133,982	115,632	50,190	820,940
Labourers and Related Workers	108,975	87,731	163,942	170,428	148,953	77,452	757,481
Total	610,016	912,044	2,007,559	2,140,809	1,879,425	814,498	8,364,350

Table A3 : Adjusted Counts for Occupation Major Groups by States and Territories, 2001 Census

<i>Occupation Major Group</i>	<i>States and Territories</i>							
	<i>NSW</i>	<i>Vic.</i>	<i>Qld</i>	<i>SA</i>	<i>WA</i>	<i>Tas.</i>	<i>NT</i>	<i>ACT</i>
Managers and Administrators	257,595	197,024	132,071	59,640	71,465	15,777	6,931	15,804
Professionals	523,046	397,234	249,510	105,925	141,192	30,725	15,004	39,973
Associate Professionals	316,540	234,575	186,168	72,404	101,208	21,115	11,357	21,553
Tradespersons and Related Workers	321,906	251,579	194,986	76,939	109,291	22,612	9,901	12,472
Advanced Clerical and Service Workers	113,983	75,655	54,618	20,871	31,134	5,218	2,777	5,202
Intermediate Clerical, Sales and Service Workers	451,694	333,865	264,528	104,438	132,797	30,436	13,708	27,847
Intermediate Production and Transport Workers	215,276	167,882	132,631	52,840	70,771	16,410	6,116	5,709
Elementary Clerical, Sales and Service Workers	256,158	199,747	158,110	58,563	78,627	18,275	7,210	14,939
Labourers and Related Workers	218,655	171,691	152,699	67,418	73,233	17,879	8,980	6,404
Total	2,674,853	2,029,252	1,525,321	619,038	809,718	178,447	81,984	149,903

Table A4 : Adjusted Counts for Occupation Major Groups by States and Territories, August 2001 Labour Force Survey

<i>Occupation Major Group</i>	<i>States and Territories</i>							
	<i>NSW</i>	<i>Vic.</i>	<i>Qld</i>	<i>SA</i>	<i>WA</i>	<i>Tas.</i>	<i>NT</i>	<i>ACT</i>
Managers and Administrators	211,387	181,065	105,502	53,072	55,787	14,293	4,774	13,097
Professionals	541,960	426,986	248,710	106,616	149,736	29,124	17,139	42,895
Associate Professionals	303,434	244,637	189,905	74,052	105,797	21,292	12,782	20,365
Tradespersons and Related Workers	337,438	274,262	204,980	78,047	126,108	24,608	9,407	13,485
Advanced Clerical and Service Workers	135,142	87,549	73,470	24,247	37,761	5,060	2,151	6,192
Intermediate Clerical, Sales and Service Workers	506,530	345,181	269,701	99,936	138,199	32,112	14,807	27,904
Intermediate Production and Transport Workers	239,956	188,783	146,649	57,415	72,563	16,436	7,352	7,732
Elementary Clerical, Sales and Service Workers	255,846	209,571	166,899	61,217	87,230	16,152	8,413	15,860
Labourers and Related Workers	238,103	164,813	165,994	68,970	82,922	20,114	10,381	6,294
Total	2,769,796	2,122,846	1,571,810	623,572	856,103	179,191	87,207	153,825

Census Papers

2001 Census Papers:

- 03/09 *2001 Census: Level, Main Field and Year of Completion of Highest Non-School Qualification*
- 03/06 *2001 Census: Occupation*
- 03/05 *2001 Census: Labour Force Status*
- 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*

A range of 1991 Census Working Papers, from 93/1 to 96/1 are also available.

These Papers can be accessed 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 -> (Fact Sheets and Information Papers) 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>.