IMPLEMENTING A QUALITY IMAGING AND RECOGNITION PROCESS FOR CENSUS 2001

PAPER PREPARED FOR
18TH POPULATION CENSUS CONFERENCE

By DIANNE MACASKILL
STATISTICS NEW ZEALAND

Implementing a Quality Imaging and Recognition Process for Census 2001

Introduction

At the 17th Population Census Conference I presented a paper on 'Imaging in the New Zealand Census'. The paper described the application of imaging and recognition technology to process the responses from the 1996 Census. The conference was held in October 1996 and at that time the processing was not yet completed so that the benefits of imaging had not been fully evaluated.

The use of imaging and recognition technology for the 1996 Census was the first such project to be undertaken in Statistics New Zealand. A number of significant benefits were achieved from the use of the technology and imaging and recognition technology will be used for processing Census 2001. It is recognised that the technology to be used for Census 2001 has progressed since the imaging and recognition systems were selected for Census 96. Many of the lessons learned on the impact of imaging and recognition on the statistical process can be applied to the systems being developed for Census 2001. However, it is also necessary to consider the impact of the updated systems and have processes in place to ensure that quality outputs are produced from Census 2001.

This paper covers the work that has been done within Statistics New Zealand until mid-July 1998. As planning for Census 2001 progresses it is likely that the processes described in this paper will develop and be adapted. However, the basic framework given for the management of quality will not change.

Overview of the 1996 Census and the use of imaging and recognition technology

The Census was held on 5 March 1996. Processing started 20 days later and the scanning phase to capture the data was completed on 21 June. The target date for completion of all unit record processing was 30 November 1996, with final data to be released by 28 February 1997.

Statistics New Zealand used imaging technology to process the 1996 Census of Population and Dwellings. This involved the scanning and recognition of tick boxes and numerics of an estimated 1.3 million dwelling and 3.6 individual double sided A3 forms, together with 5,500 fieldbooks (maximum of 165,000 double sided pages).

The data from both the forms and enumerator fieldbooks was captured. Intelligent character recognition was used to interpret the tick box and numeric responses. Only numeric and

tick box recognition was used as the recognition rates for alphabetic responses were too low at the time the testing took place in early 1994.

The images and interpreted data from the forms were then used for subsequent coding and editing. Images of the responses were displayed on a screen so that operators did not need to refer to responses from the paper forms. From this the operator repaired unrecognised numerics, coded the alphabetic responses and completed all editing.

The development of the image-based systems was contracted out to Hermes Precisa New Zealand which was subsequently purchased by Datamail another New Zealand company, during the processing phase. The contractor provided systems and staff to manage the scanning and recognition phases. In addition, the contractor provided the capital equipment that was needed for the short time and the specialist expertise in databases and network operation required for imaging technology.

The Capture and Code and Edit site was managed by Statistics New Zealand. This was to minimise risks of adverse public reaction to the perception of external agencies having access to confidential census information. Statistics New Zealand managed all operational issues apart from the system support and maintenance and the scanning operation. All processing activities were under-taken in a single centralised site. A two-shift, six-day-a-week operation using leased equipment. A single Sunday shift was worked when necessary.

To ensure data was of a high standard, acceptable levels of recognition were agreed on and written into the contract. A lot of time was put into developing the contract so that there were clear processes to follow if there were disputes over accountabilities or responsibilities. Penalty clauses were also written into the contract to cover the situation where contractual obligations were not met.

Imaging in the 1996 Census

There were four principal components to imaging in the 1996 Census. These were scanning, recognition, repair and the management of the images. Scanning for the purposes of the census can be described as taking an electronic image of a document and converting the dark and light areas into a map of 0's and 1's. The scanned document could then be viewed as an image although not used to manipulate data.

The recognition component took the defined areas of the map as described above and converted it into meaningful data which could be manipulated and analysed. The repair component involved the intervention of an operator who would key enter unrecognised responses and edit the data. The fourth component which was the Management of Images

involved the development of systems to retrieve images on demand and have them available for data capture, editing and coding.

Benefits Achieved from Imaging the 1996 Census

New Zealand was one of the few countries to use imaging for the 1995/96 round of Censuses. At that time some development work was being done by countries to investigate and research imaging and recognition for the 2000 round of Censuses.

When the decision was made to use imaging a large test was undertaken using forms from 1991 to determine the benefits and the costs of using imaging. At the time that this work was being done Statistics New Zealand was not aware of any other countries planning to use imaging and recognition for the 1996 Census. Testing showed that the expected benefits from using imaging included the earlier availability of data for publication, cost savings in data capture, systematic control of quality, reduction in staff numbers and hence staff management costs, reduction in paper handling and storage, easier access to responses on questionnaires and easy access to data for the Post Enumeration Survey.

There were significant benefits gained from using imaging although not as many as anticipated. These included the achievement of the scanning timetable and the implementation of a paperless office. When the images had been scanned they were copied to compact discs and these discs were used by the operators during the code and edit process. No paper questionnaires were used during this phase of the process and data was available earlier than in 1991. Processing was completed in late December 1996 and some key census counts released by the end of February 1997. It was, however, early May before more detailed analyses were available which was over two months later than the target date. The complete dataset was not available until four and half months after the February target date. The main reason for the delay was the large amount of work needed to clear problems that had not been identified during unit record processing.

It was difficult to estimate the savings that resulted from using imaging and recognition. The use of imaging had an impact on all parts of the process and the process and systems development required that the impact of using imaging was considered at all phases. This meant that there were a number of changes to work flows and work practices in 1996 which make it difficult to do direct comparisons. In addition, the form used in 1996 was also longer than that used in 1991 and there were 8% more households and 7.2% more dwelling forms than in 1991. In 1993 when imaging was being considered a test was done using 1991 forms to compare the cost of key entry against the cost of scanning. At this stage it was estimated that the use of imaging for the capture of responses would save $470,000 over the traditional key entry approach. At the time that Census budgets were being

prepared and the project was being funded $1.2 million was deducted by government to compensate for the efficiencies gained from the use of the technology.

There was a reduction in the number of staff. In addition, it was important to make the best use of the capital investment and additional shifts were run rather than employing more staff on a single shift. Approximately seventy staff per shift were employed. This compares to 425 staff being employed for processing the 1976 Census and 175 staff for the 1991 Census.

The use of images rather than paper forms was successful. Code and edit was carried out by operators using the images. While operators did not need paper forms there were examples where forms needed to be re-scanned. This meant that there needed to be an efficient system for retrieving the paper forms for the times when they were required. Forms were efficiently retrieved for use in the Post Enumeration Survey comparisons.

Recognition Rates Achieved

The recognition rates that were achieved for some key variables are provided in Appendix 1.

Problems associated with Imaging

The recognition and error rates as specified in the contract were met. However, the impact of recognition problems was not random, which resulted in biases in some of the final data which were not detected until near the end of processing. For example in the fertility question 8,200 records were incorrectly recognised as 9's. Many 3's were recognised as 9's but not the reverse. As not many women have nine children in New Zealand this resulted in an apparent threefold increase in the number of women who had given birth to nine children.

Outputs from the Census are used at very low levels of aggregation. When outliers were examined at these low levels the quality was unacceptable. This resulted in the retrieval of the image or paper form for each suspect field which was time consuming and was a major contributor to the delay in publishing the detailed results from Census 96.

The optimisation of the setting of accuracy parameters for recognition, error rates and edits is critical to obtaining quality census outputs which are accurate, timely and within budget. The recognition parameters have two parts, the overall level of accuracy and the confidence level associated with recognition. Setting the confidence level too high can result in expensive unnecessary operator intervention while having too many unit record edits to identify incorrect recognition can slow down the rate of processing resulting in increased costs and a loss of timeliness. It is important to find an optimal balance between the rate of

recognition to be achieved, the confidence with which any variable can be said to have been recognised correctly and the number of edits.

The optimisation of the setting of confidence intervals for recognition, error rates and edits is critical to obtaining quality census outputs which are accurate, timely and within budget. Setting confidence intervals too high can result in expensive unnecessary operator intervention while having too many unit record edits can slow down the rate of processing resulting in increased costs and a loss of timeliness.

There were also some problems associated with ensuring that all forms had been scanned. This was a particular problem for large non private households. This highlighted the need for ensuring that there were systems in place for tracking each form and for ensuring that each form went through all the required phases of processing.

Census 2001

The above gives a brief description of the outcome of the imaging and recognition process for the 1996 Census and provides the starting point for developing improved systems and quality evaluation for Census 2001.

Process for 2001

Statistics New Zealand will be using imaging and recognition for Census 2001 and images will be used as the source of information for the code and edit phase.

All phases of Census 1996 were evaluated. The following is the list of high level issues that were identified as needing attention in order to improve the quality of the imaging and recognition process in 2001. The list has also been developed to take account of the fact that the technology will be more advanced than that selected in 1994 and any quality management system needs to allow for the identification of new, unexpected issues or problems.

- The quality requirements for each variable based on key uses need to be clearly specified. Testing is to be undertaken to ensure that the quality requirements are met and that all phases of the census process are designed to meet these requirements. The quality requirements for each variable will also drive the level of unit record editing that is applied to each variable.

- Quality management systems should recognise that all phases of the census process contribute to the quality of the imaging and recognition process. This includes instructions given to enumerators on the management of the forms, the clarity with which forms are completed, the design of the forms and the design of the questions.

- Documents need to be tracked through all stages of the Census process, from distribution to collection, processing and destruction. The expanded use of an address register is being investigated to use for the management of the forms and the data through all stages of processing.

- Documents need to be stored in such a way that individual documents can be readily retrieved if any problem occurs during processing.

- Recognition of alphabetic responses will be investigated. The coding process used in 1996 worked well. The use of alpha recognition for fields such as occupation and industry will depend on whether the investment that needs to be made to implement alpha recognition is smaller than the benefits from changing a process that is working effectively.

- More use needs to be made of experts in imaging and recognition technology so that the business requirements of Statistics New Zealand for high quality outputs are more readily met.

Managing the quality of the imaging and recognition process for Census 2001

Managing quality in Census 2001 will involve optimising the balance between cost, timeliness and accuracy. The level of funding available for Census 2001 is fixed and all Census processes must contribute to ensuring that the outputs from the Census will be provided within budget and meet the standards required.

Quality is achieved when outputs are produced within budget and meet the accuracy and timeliness requirements of key users. Quality involves setting standards for the quality of the output and being able to provide users with an assessment of the impact of relevant design and operating decisions on the resulting data accuracy and timeliness. Decisions need to be made on where the greatest impacts are likely to be so that work can focus on managing these impacts. For example, in 1996, the exact specification of form colour was seen as being an important contributor to achieving high levels of recognition. However, this needs to be balanced with obtaining colours that encourage response. In comparison, we may not want to spend long evaluating the effect of different wording in Help Notes on the quality of recognition given that it is known that very few people read the notes.

Statistics New Zealand is using a framework as an overall high level model for managing and improving the quality of Census outputs. The model has an application across all Census processes although this paper deals with its application to the imaging and recognition process. The model has four main components and is based on the premise that quality needs to be specified at the start of the project. Processes used are then evaluated for their impact on the quality standards at each stage of the Census cycle. In particular it requires that the points where errors can occur are clearly identified and processes are put in place to ensure that the errors are managed in a way that is consistent with meeting the required quality of the output.

Quality is managed by both a top down and a bottom up approach. The top down approach requires that systems and processes are in place to ensure that the impact on quality is considered at each stage of the process. This involves the use of a standard project management system which requires a quality management plan for each stage of the Census.

The bottom up approach for managing quality requires the identification of possible sources of error and the formulation of plans to manage them which are consistent with balancing quality, cost and timeliness. As described below this part has two components. The first part is the identification of those parts of the census process which can impact on the quality of imaging and recognition, as for example the design of the questionnaire, and managing their development in a way which when integrated with other processes optimises the efficiency with which the quality standards are achieved. The second part involves testing and evaluating the quality throughout the development and operational phases.

The following are the main components of the quality framework for the Census and for the imaging and recognition process.

1. **Quality Specification**

   The quality required for each output will be specified at the time that the content is specified. At present the end of August 1998 is the target date for completing this phase. The quality requirements will be based on the known requirements of the users of Census data and set so that there is some optimal balance obtained across the possible conflicting requirements of accuracy, cost and timeliness. The processes that impact on the quality of imaging and recognition will be designed and evaluated to meet the output quality specifications. Output quality specifications will not be the same for each variable.

2. Project Management Systems

Statistics New Zealand uses a project management methodology called Statistics Project Management for all major survey developments. This methodology is based on one called LBMS which was successfully used for the implementation of a significant IT upgrade within Statistics New Zealand. The project management methodology uses templates to ensure that the scope for the project or sub project is agreed, that roles and responsibilities of each person working on the project are defined, that project plans and timetables are prepared and that quality and risk management plans are an integral part of overall planning. The use of this project
management methodology therefore means that there is a formal process for ensuring that quality is considered at all parts of the development.

The sub-projects that make up the total Census development and operation have been identified and a separate set of project management documents are being developed for each of these sub projects.

3. Impact on Imaging and Recognition

All phases of the Census that can impact on the quality of imaging and recognition need to be designed so that when bought together the final outputs meet the stated quality standards. It is recognised that this may involve trade-offs and that these trade-offs need to be understood and known. For example, at the time of selecting a company to provide the imaging and recognition services the quality standards need to be made clear so that the optimal interaction of manual edits and recognition criteria is developed. Specifying confidence levels for recognition that are too widecould result in a large number of edits being needed to identify problems with the data. Alternatively the costs of putting too high a confidence requirement on the recognition for any individual field could result in too much planned operator intervention.

4. Monitoring and Evaluation

The problems identified during development, testing and production will be logged and measured and solutions identified and developed where appropriate. This will involve a comprehensive testing and evaluation programme to ensure that where possible problems are identified prior to the start of processing. The monitoring and evaluation programme will also take account of the fact that some problems will only be identified in the operational phase. Evaluation systems need to be developed so that these problems are identified early in processing in time for early corrective action to be taken.

Monitoring and evaluation also requires that the impact of any changes which could impact on the continuity of outputs should be measured and understood. The processes used to edit data after imaging and recognition will need to be evaluated to ensure that any changes to those which were used in 1996 will not have unexpected impacts on the outputs and their comparability between the two Census years.

Quality Specification

During May, June and July 1998 there was extensive consultation with users. A discussion document on the content of Census 2001 was written and circulated and used as the basis for discussion at a number of consultation meetings. The document listed
the criteria for including a topic in the Census. Some of the criteria relevant to quality were that the Census should be a suitable vehicle for collection and that the topic could be included within the overall budget.

The population census is a key part of the wider integrated social and population statistical system that covers the Census, household surveys, demographic estimates and statistics derived from administrative records. Analysis of administrative records and household surveys has shown that there are four variables which are more important than others because they are the key variables used to integrate with administrative datasets and with other social surveys. These four variables which are common to almost all datasets are age, location, ethnicity and sex. These are the variables which allow statistics from different data sources to be effectively integrated and which need to be measured in a standard way and as accurately as possible. These variables need to have the highest quality standards.

Much government social policy centres on support for the family. The Census is the key source of information on the family. The family code is derived from a series of questions in both the individual and dwelling form on relationships within the household. A review of the uses of the family classification is currently being undertaken in recognition of the fact that the complexity of the family is changing and in 1996 the cost of coding the more complex family types was high. The review will identify the level to
which family will be coded and then every effort will be made to ensure that high quality outputs can be obtained for the variables that are used to classify family type. The quality standard of variables which are used to code family along with the four variables described above will be specified and then processes designed to measure the variables as closely as possible to the specification.

There is another group of variables which are used directly for policy and electoral purposes. In New Zealand the electoral system is based on a general roll and a Maori roll. The number of electorates is based on the electoral populations for both the Maori and the non Maori

populations. It is critical that these populations are as accurate as possible to ensure equity of representation in the electoral process.

Many other variables in the census are used to describe and differentiate population sub groups. While the quality in terms of coverage and accuracy needs to be high many of the variables do not need the very high quality of those described above in order to be fit for purpose.

The specification of quality impacts on the imaging and recognition processes in that it will be used to determine the optimum balance between the confidence levels set for recognition and the amount of unit record and macro editing that will be undertaken after recognition.

Project Management Systems

Effective project management systems are necessary to ensure that the overall approach to the management of a project is comprehensive and that information is available to ensure that the project is on target to deliver what is expected within budget. Statistics New Zealand has developed its own project management methodology called Statistics Project Management. Project management will not ensure that the required quality is achieved but it does provide a framework for ensuring that there are systems in place to manage and consider quality during all phases of a development.

Statistics Project Management addresses six primary aspects of project management. These are the:

- organisation of the project which includes roles, responsibilities, authorities, etc.
- planning which includes scoping, project plans and identification of milestones
- progress control which includes status identification and reporting on progress
- change control
- risk management
- quality management

Templates have been developed for all these aspects and managers need to complete them and then follow a formal process to have them reviewed and approved.

Statistics Project Management is used for all survey developments and is being applied to the Census for the first time. It has, however, been successfully used for a number of large statistical developments within Statistics New Zealand. A large project such as the Census is broken down into a number of sub-projects. The documentation required for the Statistics Project Management system requires that the success criteria for any sub-project are identified and that a quality plan is developed for each sub-project which details the way in

which quality will be managed so as to produce outputs which meet the criteria and are fit for the purpose for which they were produced.

The quality plan has a number of specified components. These are the clear specification of an agreed end product, the description of the end product, the documentation of the quality standards to be meet, the review mechanisms for determining whether the planned process will produce the required quality of output. The quality plan also includes review dates and the dates by which milestones will be completed. There is also a requirement for the quality plan to be signed off by those with accountability for producing or using the specified output.

A quality plan for all Census sub projects is currently being developed. This will guide the individual Census managers who will refer to it and build on it when incorporating quality plans into sub project plans during development.

The implementation of Statistics Project Management methodology for Census 2001 means that there are processes to be followed for each part of Census which need to be documented and are subject to review. This does not in itself guarantee quality but does mean that every manager has a template to follow which leads them to consider the key quality and risk issues associated with their phase of the project. For the imaging and recognition sub-projects, managers will need to consider all the Census processes and ensure that the quality plan for their sub-project is consistent with the quality plans for the sub-projects that impact on the quality of imaging and recognition.

Impact on Imaging and Recognition

The overall quality plan required by the project management system is an important tool for ensuring that quality is considered throughout the development. However, ensuring that the quality plans are completed to an adequate standard and are comprehensive is more complex.

The interactions between the Census sub projects also require consideration and the quality of imaging and recognition cannot be considered only as part of the processing phase. The form design, printing methods, preparation for scanning methods, enumeration procedures, document management systems, code and edit systems and validation and evaluation methods all have an impact on the quality of imaging and recognition.

Appendix 2 lists the summary components of the census development and the identified interactions of each of these components with the imaging and recognition process. Work has just started on this aspect of Census development and the list will continue to be

developed. The following are some examples of where a range of components of the Census impact on the quality of imaging and recognition.

- The form is critical to the quality of the imaging process. It must be in a 'drop out' colour to ensure the background text is not noticed by the imaging technology. At the same time it needs to be consistent with the requirements of good survey design and not have a negative impact on the respondent's ability or willingness to complete the form. While the 'drop out' of the background colours performed well in 1996 this interaction with imaging will need to be reviewed again for 2001.

- In Census 1996 it was found that better scanning results were achieved if the paper weight was 90-100gm which is slightly heavier than normal. In addition it was found that for pre-printed fields there were some print fonts that gave better results.

- All forms that are to be scanned need some form of registration mark to set a reference point for scanning and recognition. As with the drop-out colour these marks need to be consistent with good form design so as not to impact on the quality of the answer.

- There are a number of quality issues associated with enumeration which need to be understood and managed. For example, in Census 1996 some enumerators added sticky yellow labels as reminders to office staff to take some specified action. The unexpected outcome of this was that some glue remained on the form after the label was removed and the form could not be readily scanned. Educating enumerators in the requirements of imaging is important, as for example torn or stained forms can impact on scanning rates as can staples or pins which damage the forms. Most forms were returned by households in good condition. However, further research needs to be done on the best ways of ensuring that householders return the forms in good condition.

- Good form design needs to minimise the incidence of changed responses which may result in an incorrect response being recorded or require operator intervention during processing.

The above are just some examples of issues that will need to be considered to ensure a quality imaging and recognition process.

The high volume of processing associated with the Census means that systems need to be developed to ensure that processes continue to flow and that one delay does not result in delays later in the process. There are a number of ways of managing this including ensuring that systems are designed so that wherever possible they are not dependent on an earlier part of the system. The imaging and recognition systems will need to be designed so that those forms which are not processed can be put aside and dealt with quickly but in such a way as to not impact on the overall flow of work Another example of where this principle might

impact on scanning and recognition is that processing should be able to start as soon as possible after scanning for those households for which all forms have been received and not be dependent on having all forms for all households.

Lessons from Census 1996 also show that it is not efficient to have imputation processes which require all the data to be processed and which run the risk of having problems identified when it is too late or too expensive to correct them. For example, imputation methods will be used which ensure that imputation for non-response can be undertaken at the time a record is processed and does not need to wait until the end of processing.

An important aspect of any quality plan is peer review by experts. Statistics New Zealand built up a lot of expertise with the 1996 Census. However, it is recognised that the technology has changed and improved since decisions were made for that Census. Consultants will be used to help with both the tender process for selecting the company to undertake the imaging and recognition and to assist with identifying the best technical solutions for Statistics New Zealand. Technical experts will be retained throughout the development to provide advice on the solution paths being taken. In particular, advice will be sought on the statistical impact of any decisions that are made. Close contact will also be maintained with other statistics agencies who will be using imaging and recognition for the 2000 round of Censuses. An expert in management engineering (operations research) will also be used to advise on optimising work flows.

Monitoring and Evaluation

Monitoring of error and evaluation of outputs is a key component of the management of quality.

A testing programme is currently being developed for Census 2001. The testing programme will focus on identifying all the areas that impact on imaging and ensuring that processes are in place for their management. There will be a number of pilot tests undertaken over the next two years and these will be specifically designed to test certain processes such as help notes and instructions to enumerators to determine whether the proposed solutions are effective in improving the quality of imaging and recognition. There are a very large number of points where testing could be undertaken and information already available needs to be used to determine whether investment in a particular test would result in improved quality.

A major part of the testing programme will be the Dress Rehearsal which will be a test of about 5,000 households one year before the census. All processes will need to be in place by this time so that they can be fully evaluated in a real 'Census-like' situation. There are a number of problems which don't become apparent until tested in a volume situation or tested

as part of a whole process. Replicating the Census process as closely as possible in the Dress Rehearsal will be one of the important tests undertaken in preparation for Census 2001.

A key component of evaluating quality will be to work through the best mix of setting the levels for the recognition of fields, the accepted error rate and the editing strategy. A number of individual tests will be done before determining a final solution. The final decisions will be made in partnership with the company undertaking the imaging and recognition. To date initial tests show that the more advanced equipment and technology available is giving superior results to those achieved for Census 1996.

Although it is intended to have a rigorous testing plan it needs to be recognised that unidentified problems may occur in the census proper when large volumes of forms flow into the census processing centre. An evaluation plan will be developed aimed at allowing us to identify the problems early and take remedial action early in processing. Part of the problems that occurred in 1996 were caused by problems not being identified soon enough with the result that there was considerable reprocessing required.

The evaluation plan will include an evaluation at the micro level. Individual records will be checked after edit to ensure that problems which could impact on the quality specification have been identified and corrected by the processes in place. Macro checks will also be developed so that outputs can be compared with expectations based on the last census and other relevant data.

One of the key quality issues that need to be considered is the issue of continuity so that information from one Census is comparable to information from another Census. There is always a tension between continuity and making efficiencies. Throughout the process we will need to be aware of the decisions that are being made that could impact on time series data. In New Zealand Census data is released at low geographic areas sometimes containing fewer than 100 people. While confidentiality is preserved in any release the low levels at which data is provided also means that any anomalies in the data become apparent and require explanation.

While this paper has focused on Census 2001 the planning will also involve collecting information which will be valuable input into Census 2006. For example, recording why operator intervention is needed to repair some fields will give useful information for 2006 on the cause of the intervention and will help direct improvements for Census 2006.

Summary

The quality management strategy for the Census in New Zealand is aimed at providing data required by key users to a minimum standard to suit use, using best practice processes to achieve efficiency as well as effectiveness. To be assured that this objective is being met, key outputs from the Census will be defined along with standards for accuracy and timeliness and then an assessment made of the impact of relevant design and operation decisions on the resulting data accuracy and timeliness against those standards.

Appendix 1

Results of tick box responses

	Requirements	**Achieved**
Recognition Rate	100%	100%
Error Rate	0.1%	0.2%

Results of numeric responses

	Requirements	**Achieved**
Recognition Rate	70%	85%
Error Rate (excluding blank fields)	Maximum 3.0%	Average 2.6%

The rates in the tables above are averages. The following table shows recognition and error rates per digit for 11 numeric fields.

Field	**Recognition rate**	**Error rate per digit**
Fertility (I29)	93%	2.2%-2.9%
Qualyrs (I32)	85%	2.9%-3.6%
Weekly Rent (D8)	76%	3.7%-5.0%
Rooms (D11)	86%	3.1%-3.8%
Bedrooms (D12)	91%	2.5%-3.2%
Usual Hours 1,2 (I48)	83%	2.5%-3.0%
Date of birth (I7)	81%	2.2%-2.3%
Dwelling form age (D3)	82%	2.6%-2.8%
Years at address (I3)	84%	3.5%-4.0%
Years in NZ (I9)	81%	2.3%-2.8%
No. of absentees (D18)	82%	2.3%-3.3%

Appendix 2

The list below specifies some high level Census components and the factors of each component that impact on the quality of imaging and recognition

A decision has not yet been made on whether alpha recognition will be used and the interactions with alpha recognition have not been identified in the following lists

Questionnaire (Design and Printing)

Scanning Colour drop-out
Registration marks
Length and width of forms
Paper thickness
Bar codes for receipt

Numeric Response zones clear of text
Constrained borders
Instructions on forms and in help notes

Tick Box Response zones clear of text
Constrained borders
Instructions on forms and in help notes
Clear unambiguous response options

Capture

Numeric Setting of rejection rates for every variable
Setting of error rates for every variable
Macro and micro editing strategy

Tick Box Setting of mark size
Decision rules for multiple marks

Enumeration

Scanning Decision on address register will impact on the management of the forms
Preparation of the fieldbooks prior to scanning
Managing state of the forms so that they are in good condition

Numeric Publicity and promotion of the Census
Training of enumerators

Tick Box	Publicity and promotion of the Census Training of enumerators

Printing

Scanning	Clear fonts Minimum background noise Paper thickness Print type

Editing

Numeric	Dependent on quality specification Dependent on optimising recognition rates, error rates and editing Balance between micro and macro editing Specification of misrecognition bias
Tick Box	Dependent on quality specification Dependent on optimising recognition rates, error rates and editing Balance between micro and macro editing Strategy for multiple marks

Appendix 3

Background Census 2001

Date	To be decided but probably March 2001
Population	Estimated to be 3.9 million
Dwellings	Estimated to be 1.5 million
Number of enumerators	Approximately 6,000
Budget	Approximately $30 million over 5 years Cost of enumeration approximately $15 million
Language	Two languages will be used English and Maori
Statutory topics	Address on Census Night. Address of Dwelling Age Ethnicity Maori descent Name Number of Occupants on Census Night Number of Rooms Sex Tenure of Dwellings Usual Residence