

2018 Census data quality management strategy





Crown copyright ©

[See Copyright and terms of use](#) for our copyright, attribution, and liability statements.

Citation

Stats NZ (2017). *2018 Census data quality management strategy*. Retrieved from www.stats.govt.nz.

ISBN 978-1-98-852822-9 (online)

Published in July 2017 by
Stats NZ Tatauranga Aotearoa
Wellington, New Zealand

Contact

Stats NZ Information Centre: info@stats.govt.nz

Phone toll-free 0508 525 525

Phone international +64 4 931 4600

www.stats.govt.nz

Contents

Purpose and summary	4
Purpose	4
Summary	4
Introduction	6
Quality definitions	7
Six dimensions of quality	7
Four perspectives of quality	8
2018 Census quality goals	10
Strategies to achieve the 2018 Census quality goals	11
Appendix 1: Definitions of the six dimensions of quality	17
Appendix 2: Variables priority rating	18
Output data framework	18
Quality priority level variables/topics	19
Appendix 3: Data definitions	22

Purpose and summary

Purpose

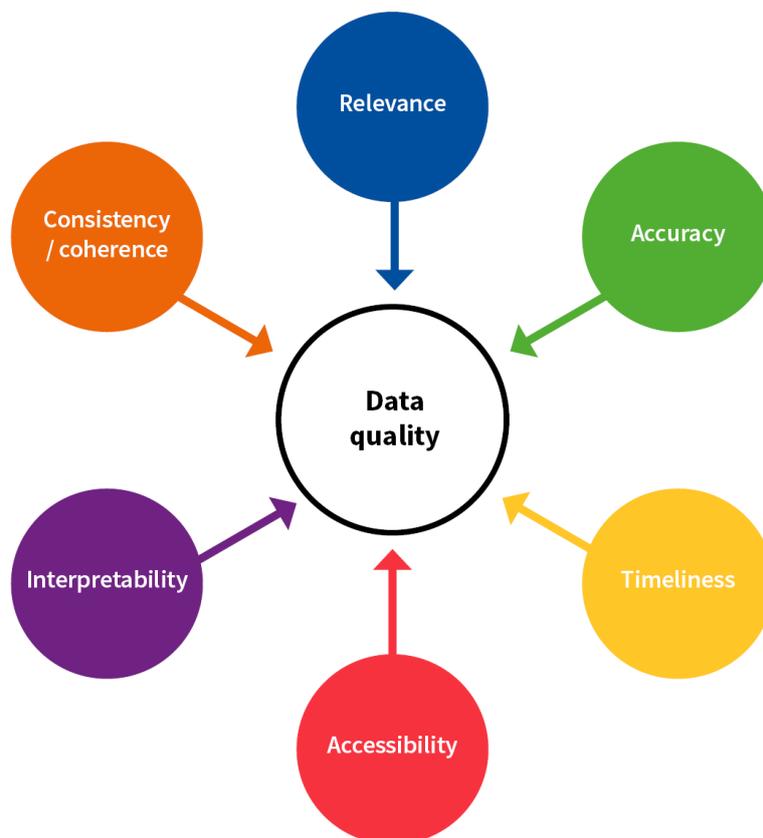
2018 Census data quality management strategy outlines how the 2018 Census programme will ensure key census data needs are met and outputs are ‘fit for purpose’.

Summary

This strategy aims to ensure quality is:

- clearly understood and communicated
- designed and built into every census process
- directed by the needs of internal and external customers.

Quality management best practice outlines six dimensions that contribute to the quality of data: relevance, accuracy, timeliness, accessibility, interpretability, and consistency/coherence.



For the 2018 Census, it is important that we place the right emphasis and priority on each of the **dimensions of quality with the ultimate goal of ensuring output data that is ‘fit for purpose’ and has value from our customers’ perspective.**

The 2018 Census has six quality goals and under these goals are 13 strategies with guidelines and actions for achieving the goals. The one-page summary below outlines these goals and strategies.

2018 Census data quality management strategy – one-page summary

Goals		Strategies
1.	Quality is a design focus of the end-to-end census model	<ul style="list-style-type: none"> 1. Design and build quality and quality measures into processes in the end-to-end model 2. Collect and use operational data to measure quality and influence the management of processes
2.	Ensure accurate population and dwelling counts to the same or better quality than the 2013 Census	<ul style="list-style-type: none"> 3. Ensure an accurate census dwelling frame that is fit for purpose and contributes to the continuous improvement of the property data frame 4. Achieve consistent and acceptable response rates across all geographic areas and target response groups 5. Set operational thresholds to ensure Māori coverage and response is well managed 6. Manage the use of unit imputation
3.	Collect data on population and dwellings to meet customer needs	<ul style="list-style-type: none"> 7. Consult customer groups to determine relevant content 8. Variables and topics will be prioritised to determine where most effort should be applied (see Appendix 2 for more detailed information) 9. Design to maximise the completeness and accuracy of response
4.	Data is available to contribute to continuous, adaptive, and timely improvement	<ul style="list-style-type: none"> 10. Processing and evaluation processes are designed to provide insight into the quality and coverage of the data in a timeframe that enables action 11. Use administrative data to improve quality of data and efficiency of processes
5.	Deliver relevant data to customers in ways that are accessible	<ul style="list-style-type: none"> 12. Ensure accessibility of data by providing a range of products and services in forms data users need
6.	External and internal data users understand the data and the quality of data	<ul style="list-style-type: none"> 13. Develop tools and services to assist users to interpret census data

Introduction

The UN Economic and Social Council states that ‘a quality management programme is an essential element in the overall census programme’. Quality management is driven by customer needs.

The census aims to produce the best possible count of the population and dwellings and is uniquely placed to produce a rich, useful, and accessible set of related attribute data at various geographic levels.

We are redesigning and re-engineering our processes for the 2018 Census to significantly improve the efficiency and the effectiveness of the census. These changes, particularly to the collection model and the increased focus on using corporate tools and systems in the census, will bring new quality risks, challenges, and opportunities.

[The 2018 Census strategy](#) includes a goal to ‘improve data quality while modernising’. Our aim is to ensure quality is understood, designed, and built into every census process, so that each 2018 Census project’s contribution towards the programme goal of ‘delivering the same or better data quality as the 2013 Census’ is known, integrated, and managed.

Quality definitions

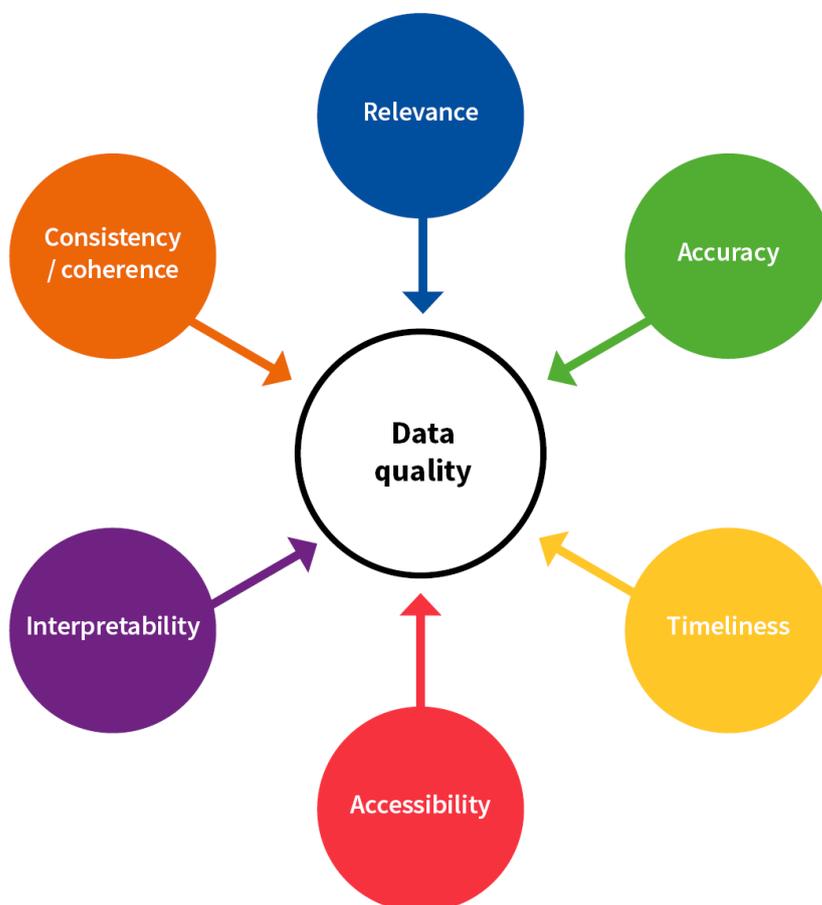
Six dimensions of quality

Quality management best practice¹ outlines the six dimensions that contribute to the quality of data: relevance, accuracy, timeliness, accessibility, interpretability, and consistency/coherence.

The dimensions of quality are interrelated and need to be balanced depending on the data items and an understanding of the needs of census data users. The tensions that arise between different dimensions of quality need to be managed. Quality, in turn, often needs to be balanced against other considerations, such as cost and respondent burden.

[See Appendix 1](#) for the definition of each dimension.

Figure 1: Six dimensions of quality



¹ The United Nations Economic Commission for Europe (UNECE) describes the six dimensions of data quality in a census context and categorises three types of data errors. These concepts are internationally recognised as a best practice model and have been developed in the draft overarching documents *Statistics New Zealand quality management policy and guidelines* and the *Statistical data quality model*.

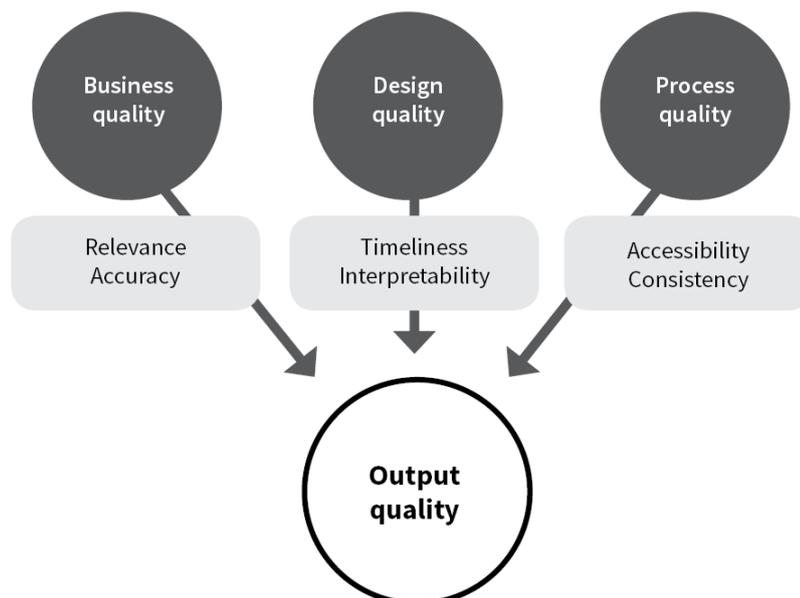
As with previous censuses the definition of ‘quality’ for census data is defined as ‘fit for purpose’ for customers. A ‘quality’ deliverable or process within the census programme, therefore, will be one that sets and places the right emphasis and priority on the dimensions of quality with the ultimate goal of ensuring that output data is fit for use and has value from customer perspectives. This will contribute to our strategic refresh aim of doubling the value of data we provide to New Zealand by 2018.

Four perspectives of quality

To make the dimensions of quality easier to apply in practice, we distinguish between the following four perspectives:

- Business quality – some aspects of quality are determined by the way we function.
- Design quality – the design phase in all parts of the census has a big influence on the quality of data and is critical to achieving our goal of delivering the same or better quality data than the 2013 Census.
- Process quality – ensuring that all processes operate according to their design. If this is achieved the final product should be fit for use.
- Output quality – the quality of the final data and data products is the ultimate goal of quality management and the contribution of each part of the census to output quality should be clear.

Figure 2: Four perspectives of quality and six dimensions of quality



These six quality dimensions and four perspectives of quality are interrelated and have the following implications for the census:

- Relevance is mostly determined during the design phase, by ensuring that questions included in the census questionnaires will produce the data that users need. Once the data are received, there is less opportunity to affect relevance.

- Accuracy is strongly influenced by both design and process quality. All processes must be well designed and implemented effectively to achieve the required accuracy for outputs.
- Timeliness is part of the design decisions that are made in planning stages. Our processes have to be constructed to meet timeliness constraints and deadlines that have been set. The tension between timeliness and accuracy is a key one for the census.
- Accessibility is a business-, design-, and process-quality issue and is important for adding value for customers. External access will depend on how effective our dissemination systems are (business quality). Internally, design that makes data accessible will assist the programme with internal handovers (eg project to project; and to other Stats NZ areas) and for use in future census development.
- Consistency issues are part of business, design, and process quality. Consistent use of our standards and classifications throughout census will ensure alignment with other surveys. Consistency is determined during the design phase, is applied in processes, and needs to be monitored and evaluated throughout the census cycle. For the census, the tension between relevance and consistency over time needs to be managed.
- Interpretability means the customer should have good information about data and data quality (ie what does this data mean and how good is it, how can I use it?). Business quality means following our quality reporting standards and policies. It is important to design processes across the census that provide information about data and data quality.

2018 Census quality goals

To deliver quality throughout the census cycle, we've developed six 2018 Census quality goals that reflect the generic statistical business process model of 'Need, design, build, collect, process, analyse, disseminate' and the six dimensions of quality.

The 2018 Census quality goals are:

1. Quality is a design focus of the end-to-end census model.
2. Ensure accurate population and dwelling counts to the same or better quality than the 2013 Census.
3. Collect data on population and dwellings to meet customer needs.
4. Data is available to contribute to continuous, adaptive, and timely improvement.
5. Deliver relevant data to customers in ways that are accessible.
6. External and internal data users understand the data and the quality of data.

Strategies to achieve the 2018 Census quality goals

Goal 1: Quality is a design focus of the end-to-end census model

All quality dimensions need to be factored into the design of each phase of the census.

Strategy 1: Design and build quality and quality measures into processes in the end-to-end model

Quality should be embedded into our thinking and be an activity that we build in at the start of the process and during development rather than as merely an outcome measure. We need to make robust and evidence-based design decisions, and test and validate them to ensure quality outcomes.

To ensure quality we will:

- design and build quality throughout the census
- look for opportunities to build in quality at the earliest possible stage
- ensure that all areas of the census programme understand their contribution to the dimensions of quality and the impacts on data
- identify who our customers are and be responsive to them by making decisions based on their needs
- collaborate when making design decisions to identify the flow-on effects on downstream quality, particularly handovers to the next process in the chain
- use quality measures to manage data and data quality across all phases of the census
- look for innovative solutions to manage tensions between quality dimensions
- use test-driven development and testing to provide confidence and assurance that the right technology and processes are selected and that they operate as intended
- be flexible and adaptable within strategies throughout the census, rather than adopting a ‘one size fits all’ approach where standard approaches are not fit for purpose
- make sure that trade-offs between time, cost, and quality are made with the whole design in mind and are well documented.

Strategy 2: Collect and use operational data to measure quality and influence the management of processes

Large amounts of digitised data (including raw response, processed data, operational data, metadata, and paradata – [see Appendix 3](#) for definitions of different data types) will be captured or documented in the 2018 Census.

We need to design and implement systems and processes to enable us to collect and analyse operational data in a consistent way, particularly where it has an impact on the quality of output data. Doing this will ensure that quality can be measured and managed across the end-to-end model, and used to influence and improve processes. We need to agree and document expectations so that operational reporting can be effective in influencing the operation to achieve the desired level of quality.

Operational data may have more than one purpose. Key purposes, including customer usefulness, should be identified and considered to optimise quality. Operational data that has a direct effect on the quality of outputs needs to have clear standards, classification, and quality specifications to ensure it is coherent with output dataset uses and operational needs.

Operational data from testing and operation phases needs to be accessible for design decisions, and made accessible and compatible with output datasets for use in the future development of the census and wider Stats NZ initiatives.²

Goal 2: Ensure accurate population and dwelling counts to the same or better quality than the 2013 Census

The accuracy of the counts is of paramount importance for achieving high quality outputs.

Strategy 3: Ensure an accurate census dwelling frame that is fit for purpose and contributes to the continuous improvement of the property data frame

A key focus for the 2018 Census is achieving high coverage of dwellings. Different teams will collect address and dwelling data at different points throughout the end-to-end process. To enable us to manage the address frame, we need to develop robust methodologies and design of all interactions across the census programme to maximise accuracy and minimise the introduction of errors.

We need to ensure consistency and accuracy of dwelling occupancy determination and classifications. Where we cannot determine occupancy in the field, leveraging off our methodological expertise will assist the maintenance of quality.

We will use modern technology and existing systems to find solutions for identifying (and collecting census information about) the population who do not live in dwellings and are not covered by the dwelling frame.

Strategy 4: Achieve consistent and acceptable response rates across all geographic areas and target response groups

Maintaining acceptable response rates across all areas, including small geographies, ensures no area is under-represented. Collection methodology needs to be innovative and designed to maximise responses by reducing barriers to participation.

² When we share individual or household information either within Stats NZ or with other government agencies for statistical and public interest research, we remove all names, addresses, and other identifiable information. We never share identifiable information for operational purposes with other agencies (eg Inland Revenue, New Zealand Police) without your consent. For more information visit [Privacy, security, and confidentiality of information supplied to Statistics NZ](#).

We will implement strategies for particular population groups who may be hard to reach, including for individuals who are not covered by the dwelling frame. As a result, participation in the census will be maximised, by ensuring that there are appropriate avenues for response.

We will use learnings from the 2013 Census post-enumeration survey research on target response groups, and processing of timely data during the operation phase to identify the geographic areas and population groups where we can implement tailored response strategies. Different respondent needs and behaviours will require us to be responsive and adapt our approach.

Designing thresholds for target response groups will help us manage coverage and response rates of population groups to achieve acceptable quality levels. These thresholds will also indicate where and when operations can consider moving field resources to other priority areas.

The 2018 Census communications and marketing strategy of tailored, responsive communications will be a key driver of response quality. It will assist the public and targeted response groups to understand and engage with the new collection model.

Strategy 5: Set operational thresholds **to ensure Māori coverage and response is well managed**

This strategy is aligned with strategy 4. This and other goals specific to Māori data quality are contained in **Te Rautaki Māori, the 2018 Census Māori strategy**³. The primary goal of this strategy is to increase the value and use of data for and about Māori through increased Māori participation in the census.

Strategy 6: Manage the use of unit imputation

The level of response achieved influences the amount of unit imputation required. To achieve the programme goal of delivering the same or better data quality as the 2013 Census, response rates need to be maximised and unit imputation (previously known as substitution) should remain the last option for ensuring quality of population and dwelling counts.

Imputation rates are key quality indicators for the census as a whole, and for data users. Unit imputation must follow a statistically robust, transparent process, and be well documented including having an audit trail ([see Appendix 3](#) for a definition of unit and item imputation).

³ We have developed a number of strategies internally, including Te Rautaki Māori, to guide the 2018 Census. For more information about the tikanga, aims, and outcomes within Te Rautaki Māori, please contact the Stats NZ information centre.

Goal 3: Collect data on population and dwellings to meet customer needs

Census variables are fit for purpose across the six quality dimensions with a focus on relevance, accuracy, and consistency.

Strategy 7: Consult customer groups to determine relevant content

Consultation with census data users ensures that data is relevant for changing customer needs. This consultation may result in proposals for new content or changes to existing content. These proposals need to be assessed against the content determination framework to balance the need against respondent burden and the consistency of census data over time.

Final decisions regarding 2018 Census content have now been made.

[See the 2018 Census report on final content.](#)

Strategy 8: Variables and topics will be prioritised to determine where most effort should be applied

To manage the balance between the six quality dimensions and ensuring data that is fit for use, **we'll assign a quality priority (priority one, two, or three**, where one has the highest priority) to output variables or topics. These quality priorities are driven by customer needs and the key reasons for a census, and will be described within a variable specification. Priority levels indicate the relative importance of the variable and determine the relative amount of effort that will be spent on each variable.

We need to apply the prioritisation of variables throughout all parts of the census, from the design and build stages through to collection, processing, evaluation, and output.

All output data for customers, no matter which priority level, must be designed to reach minimum quality standards of being fit for purpose and meeting customer needs.

[See Appendix 2](#) for more on these priority levels.

Strategy 9: Design to maximise the completeness and accuracy of response

To manage the quality of responses the 2018 Census programme will:

- measure, baseline, and report on the quality of responses on a simple scale
- identify the places throughout the end-to-end process where we can improve the quality of responses and use design to minimise item non-response
- use imputation to improve output quality and to remove bias
- investigate increasing the range of variables that are imputed⁴
- minimise changes to responses received except where clear evidence suggests change and, where relevant, ensure consistency by using standard tools and automation

⁴ In the 2013 Census item imputation was used for age, sex, usual residence, and work and labour force status.

- design and test to minimise introduced errors
- identify and collect key information to get the best possible count in the absence of any questionnaire responses in the field. This may be a balance of what information can be directly collected from the respondent and how well missing information can be imputed through various methods.
- consider reduction of respondent burden when designing to increase the completeness of response
- use key performance indicators to guide acceptable coverage levels and response rates.

Goal 4: Data is available to contribute to continuous, adaptive, and timely improvement

Timeliness of response and operational data will allow the census to be responsive and adaptive during operations to make continuous, adaptive improvement.

Strategy 10: Processing and evaluation processes are designed to provide insight into the quality and coverage of the data in a timeframe that enables action

With a target of 70 percent online responses, large amounts of respondent data will be received directly in a digitised form. This will have a positive effect on data quality and processing.

Automation is a key design principle of the high-level processing strategy to achieve timeliness and accuracy goals. However, some manual intervention may be required. Ensuring the right balance between automation and manual intervention to maximise the quality of processed data will be crucial. This balance can be achieved by using effective statistical methods and tools, conducting robust testing, and by setting thresholds and priorities for manual coding and editing.

Strategy 11: Use administrative data to improve quality of data and efficiency of processes

Administrative data must be fit for purpose and should be used where it adds sufficient value for the investment of effort. Investigation and use of administration data will be guided by: how it improves quality or particular dimensions of quality, whether it can be used as an alternative to imputation, if it clearly reduces respondent burden, and efficiency gained.

Goal 5: Deliver relevant data to customers in ways that are accessible

Data, and information about data, needs to be presented in ways that are clear, easy to understand, and find.

Strategy 12: Ensure accessibility of data by providing a range of products and services in forms data users need

The census has a unique role as an important source of demographic, social, and economic data for small geographic areas and population groups.

Consulting with customer groups in the first instance should be a key driver in developing content for data products that make best use of census data (and information about data).

Providing a range of products for different types of users is crucially important to the goal of maximising the value of census data. We need to ensure that the availability of census data and products is widely communicated and they are easy to obtain and find on our website.

Census output data will need to be consistent with survey data within Stats NZ and other datasets, by using Stats NZ standards and classifications.

Goal 6: External and internal data users understand the data and the quality of data

Information about data should be transparent and easy to use. Ensuring this throughout the end-to-end process will help users to interpret data quality.

Strategy 13: Develop tools and services to assist users to interpret census data

Tools and services that help interpret the data and the quality of the data should be easy to understand, use, and find. They need to be developed for internal and external customers and should assist users to interpret data and provide a guide to appropriate use.

Underlying concepts, definitions, standards and classifications used, methodologies of data collection and processing, and indications of the accuracy of the information need to be easily available to census data users.

Appendix 1: Definitions of the six dimensions of quality

Timeliness: Data are released within a time period that permits the information to be of value to users.

Accessibility: Statistics and information about data (metadata) are presented in a clear and understandable way, they are provided in suitable mediums for access, users are aware of their availability, and they are easily obtained and widely disseminated.

Consistency/coherency: The census data and information is consistent and coherent within census datasets and metadata, it can be successfully brought together with other statistical data and information within a broad analytical framework and in time series. The use of standard concepts, definitions, and classifications promotes consistency.

Accuracy: Source data and statistical techniques are sound and statistical outputs sufficiently portray the reality they are designed to represent.

Relevance: The relevance of statistical information reflects the degree to which it meets the needs of the users. The challenge for a census programme is to balance conflicting user requirements so as to go as far as possible in satisfying the important needs within resource constraints. This dimension of quality is particularly important in census content development and in dissemination.

Interpretability: The interpretability of statistical information reflects the availability of supplementary information and metadata necessary to interpret and use it. This information usually covers the underlying concepts, definitions, variables, and classifications used, the methodology of data collection and processing, and indications of the accuracy of the information.

Appendix 2: Variables priority rating

Output data framework

The majority of the variables will retain the same priorities that were assigned in the 2013 Census, with a small number of variables proposed for an increase in priority level due to their increased importance in the new census model or increased relevance due to changes in public interest and policy.

In previous censuses, most variables that were used in output products were sourced from response data. For the 2018 Census, there is an opportunity to source some data from administrative sources as well. In addition, the quality of the output data can be influenced by the quality of operational data collected.

Quality priority levels will determine the relative importance of the variables and the focus of effort required across the census programme to ensure output data is fit for purpose.

Priority one

Priority one variables/topics are given the highest priority in terms of quality (accuracy, relevance, timeliness, consistency, interpretability, and accessibility), time, and resources across all phases of the census.

Priority one includes:

- core census output variables – these are the key reasons for a census.
- population and dwelling counts
- data for electoral needs
- demographic and location data that are essential for census counts and electoral needs.

Priority two

Priority two variables/topics are given second priority in terms of quality (accuracy, relevance, timeliness, consistency, interpretability, and accessibility), time, and resources across all phases of the census.

Priority two includes:

- definitions of key subject and population groups of high public interest
- data that is closely linked to the main purpose of the census
- data that is important for policy development, evaluation, and monitoring
- data that is used frequently in cross tabulations with priority one variables.

Priority three

Priority three variables/topics are given third priority in terms of quality (accuracy, relevance, timeliness, consistency, interpretability, and accessibility), time, and resources across all phases of the census.

Priority three includes:

- data that we would not run a census solely for, but we couldn't get this information for population groups without it being in a census
- data that is important to certain groups
- data that can be used to create frames for higher-quality sample surveys.

Minimum quality standards (eg within the variable specifications) have to be met to ensure the output data in all three priority levels is fit for purpose.

Quality priority level variables/topics

Priority one

Count of the population (final)
Count of dwellings (final)
Meshblock location of each dwelling in New Zealand
Age of all respondents in New Zealand on census night
Sex of all respondents in New Zealand on census night
Location of all respondents in New Zealand on census night to meshblock level
Usual residence to meshblock level of all usually resident in New Zealand
Ethnicity of all respondents in New Zealand on census night
Number of occupants on census night
Māori descent
Absentees
Unoccupied dwellings

Priority two

Families and households	Family type (and associated variables/derivations)
	Child dependency status (and associated variables/derivations)
	Household composition (and associated variables/derivations)
	Extended family type (and associated variables/derivations)
Usual residence one year ago	
Occupied dwelling type	
Tenure of household	Ownership of dwelling (dwelling form question, variable used in the derivation of tenure of household)
	Mortgage payments (variable used in the derivation of tenure of household)

	Weekly rent paid by household (variable is included in output and is used in the derivation of tenure of household)
Iwi affiliation	
Work and labour force status (wklfcs)	Job indicator (key variable for the derivation of work and labour force status)
	Hours worked in employment per week (variable used in the derivation of wklfcs)
	Job search methods (variable used in the derivation of wklfcs)
	Available for work (variable used in the derivation of wklfcs)
	Seeking work (variable used in the derivation of wklfcs)
Birthplace	
Status in employment	
Legally registered relationship status	
Partnership status in current relationship	
Total personal income	
Sources of personal income	
Sector of landlord	
Highest qualification	Highest secondary school qualification (variable is included in output and used in the derivation of highest qualification)
	Level of post-school qualification (variable is included in output and used in the derivation of highest qualification)
Field of study	
Overseas qualification indicator	
Study participation	
Main means of travel to work	
Main means of travel to education	
Educational institution address	
Workplace address	

Priority three

Occupation
Industry
Sector of ownership
Languages spoken
Number of rooms/bedrooms
Number of children born
Years since arrival in New Zealand
Years at usual residence
Main types of heating
Dwelling mould indicator
Dwelling dampness indicator
Access to basic amenities
Access to telecommunication systems
Number of motor vehicles
Unpaid activities
Individual home ownership
Religious affiliation
Cigarette smoking behaviour
Disability/activity limitations

Appendix 3: Data definitions

The term ‘data’ within this document covers all purposes (for use internally and externally to Stats NZ) and types of information that the census collects and uses (from respondents, information about the data, outputs). Definitions of these types and purposes include:

Administrative data: All data that is collected by government agencies or private organisations when conducting their business or services. It is data that is not collected primarily for statistical purposes (eg tax returns, births/deaths/marriages, building consents).

Imputation: A type of error treatment that involves determining and assigning replacement values for some or all variables for a record. Imputation may occur when we have no response from the unit (dwelling or person) – unit imputation; or when a respondent has submitted an incomplete form with one or more questions unanswered – item imputation.

Metadata: Information that describes data or methodology (eg definition of a variable).

Operational data: All data (information about data – metadata and paradata) that is not used directly for output purposes. Used for end-to-end census operations and census transformation.

Output data: Finalised output variables that have been collected, processed, and derived in the various phases of the census and are fit for purpose. They are **available in ‘clean’** databases/datasets for use by internal users and extracted to create data products for internal and external customers. Response data, administrative data, and operational data all contribute to output data.

Paradata: Data that is collected about a survey (eg how long it took to collect something).

Response data: Respondent answers to questions on the census dwelling and individual forms.