

2018 Census – Content determination process

Purpose

The purpose of this paper is to provide the methodology used for the 2018 census content determination process following the engagement and consultation period. It outlines how the process was run and who was involved. The main audience of the paper is for those working on any future censuses. However, other interested parties may include different teams within Statistics NZ and statistical agencies overseas (e.g. ABS, Pacific countries etc).

Background

Since the first official census in 1851, the content of Censuses has varied. Some topics have been included in every census, while others have been included or excluded to reflect changes in society over time. The last two censuses in 2011/13 and 2006 had minimal change.

Statistics NZ conducted a review of all current topics, which will enable the incorporation of appropriate new and revised topics in the 2018 Census, to meet the emerging needs of New Zealand's constantly changing and diversifying society. As a result of this review, new topics are expected to be included following their evaluation and successful testing.

The process began in November 2014 with the release of the 'Proposed View of Content'. This internal paper outlined potential content changes but was only for discussion within Statistics NZ. These discussions included face-to-face workshops and an online discussion forum. At the same time, Statistics NZ was undertaking face to face engagement with key customers to better understand their information needs.

Following feedback from those discussions, Statistics NZ released the Preliminary View of Content in April 2015. The report outlines current thinking on changes to the 2018 Census and was made publically available. Public feedback on the Preliminary view was then sought.

The engagement and consultation period for the 2018 Census began in April 2015, inviting online discussion and submissions from organisations and individuals.

Methods

The content determination framework

Once the official engagement and consultation period ended in June 2015, the results were analysed and summarised by topic. These summaries were the main source of information for all participants to use in the content determination process. They were heavily used when it came to scoring each topic using the 2018 **content determination framework**.

The censuses in 2006 and 2011/13 also used a content determination framework to help analyse the submissions and determine content for that year. Here is a link to the frameworks used in [2006](#) and [2011/13](#). The frameworks had been developed with the aim of being thorough, transparent, and objective.

The framework was reviewed in early 2015 by the census content teams and in conjunction with senior management. A review of international practice was also conducted ([link](#)). As a result, there were some small changes made to the framework for the 2018 Census. The framework is shown below.

Criteria to determine content for the 2018 Census

Does the proposed change add value to New Zealand's society and economy?

- Can the data be used by a wide range of decision makers?
- Will the data be used to inform decisions of national significance?
- Will the information support NZ's key uses of data?

Is the census the most appropriate information source?

- Is there wide geographical relevance across NZ?
- Is there wide relevance across the NZ population?
- Is small area or small population data needed?
- Does the Census provide information quickly and often enough?
- Does the Census provide the kind of information required?
- Is Census the best data source to meet customers needs?
- Will the Census provide data of sufficient accuracy?
- Will the Census provide data of enough depth?

Does the proposed change reflect an enduring information need?

- Does the proposed change align with the Statistics NZ's future thinking (e.g. the Census Transformation Strategy)?

Will the proposed change produce quality information?

- Are there minimal or no quality problems?
- Will the proposed change give better quality data?

Is there continuity with previous census data?

- Is it consistent and comparable with previous census data?

Is data consistent with other data collections?

- Is it consistent with other data collections (concepts, definitions, classifications)?

Is there general acceptance of the proposed change?

- Are there particular concerns for specific groups?
- Is it non-intrusive?
- Is it non-offensive?
- Are respondents willing and happy to answer?
- Do respondents feel the proposed change is of value?

Would the proposed questions be easy for respondents to complete?

- Are they easy to understand and interpret?
- Are they easy (simple) and quick to answer?
- Do people know the answer?

A weighting and scoring system was applied to the content determination framework

Once the actual criteria were agreed on, a weighting and scoring system was applied to the criteria. The census team looked at what was done internationally, and came up with two different options. The first was a basic High/Medium/Low scoring system, where all criteria have the same weighting (as used by Australia). The second was by giving a specific weighting to the criteria and rating them from 0 to 5. The latter option was used in 2006 and 2011/13, has similarities to the system used by Canada, and was chosen to be used again this time around. Those with a higher value were

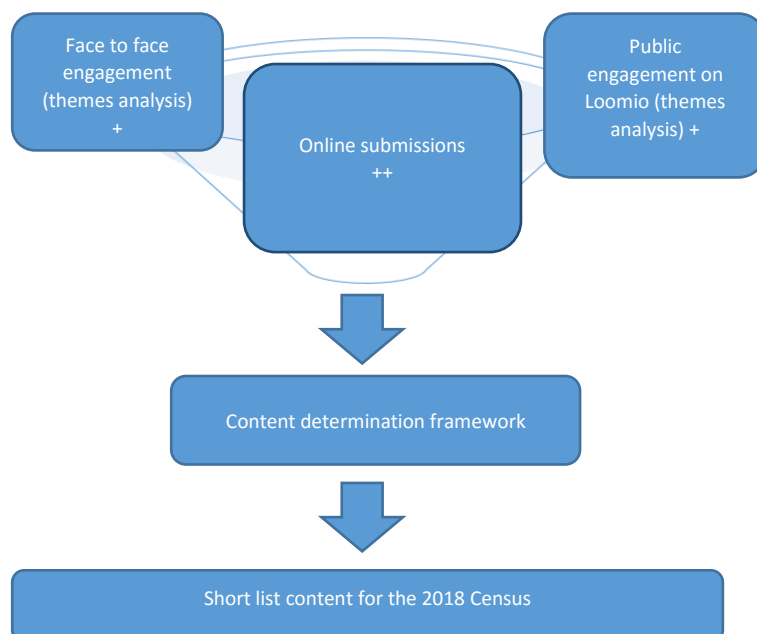
perceived to be of greater importance. The highest weightings were given to criteria around quality of information and whether the topic provided value.

Therefore, each criteria (upper level criteria only) was given a score from 0 to 5. The lower level criteria were used a guide on which the base the score on, but were not scored individually. That score was then multiplied by the weighting. Once all scores were completed, the final total (out of 90) was converted into a percentage to give a score out of 100. See appendix 1 for an example.

What went through the content determination framework

It should be noted that when it came to scoring, it is the topic that is being put through the framework, rather than individual submissions. Each individual submission was considered as a valuable source of information during the scoring of that topic.

In addition to formal submissions, information received from Loomio and face to face engagement were all used in the content determination process. However, formal submissions were given greater consideration when reviewing the feedback.



Participants

Three main groups were involved in the content determination process. The core group, made up of six members of the census content team, did the majority of the analysis. However, a wider group (made up of four other Statistics NZ staff outside of the census team) with specialist knowledge provided additional input, and a working group (some senior managers within Statistics NZ) also provided a quality check. See below for more information around the role of each group.

Core team: Their role was to:

- organise the assessment process
- summarise and organise submissions and engagement feedback
- assess submissions and engagement using the content determination framework
- attend all required meetings
- provide any necessary updates and reports.

Wider team: Their primary role was to:

- read the summary of submissions and engagement if required
- attend all required assessment meetings, where they will review and discuss the Core team's assessment of submissions and engagement. This included providing a check on the scoring given by the core group.
- while members of the wider group represented their other teams, they were expected to provide a neutral and objective view when assessing submissions and engagement and determining content.

Working group: Their role was to:

- review the work and methodology used to assess submissions and engagement
- provide a quality check on the findings of the core team and wider team.

A lot of consideration was given to ensure that the right people were included in each group. The individuals we thought suitable were contacted (either directly or through their manager), and were given a terms of reference ([link](#)) to show them what would be involved if they agreed to take part. This process gave us confidence we had worked across the organisation in a methodical manner.

Timing and meetings

The submissions period officially ended on the 30th of June 2015, although some late entries were accepted. The submissions (and Loomio and face to face engagement) were then analysed and summarised by the core team. In August 2015 a series of content discussion meetings began for the core team.

There were about 12 core group meetings. Most were two hours long, although there were some half day sessions. In total, about core group meetings totalled about 30 hours over a 4 week period. In these meetings, each topic was discussed and scored if necessary.

Following the completion of the core group meetings, the wider group meetings then commenced. These four meetings were held in early September 2015 over a two week period.

Once the core and wider group had come to an agreed position on census content, these results were then shown to working group in a single meeting in mid-September.

Topic scoring

What topics were scored

As mentioned before, it was topics rather than submissions that were scored using the content determination framework. Although not all topics were scored. The core team decided if there was value in putting the topic through the framework.

To simplify things, the preliminary view was used as the baseline for deciding what would be scored using the content determination framework. If it was Statistics NZ's position that the topic would remain unchanged, then that topic would not be scored, unless customer feedback suggested there should be a change. In that case, the change raised by the customer(s) would be scored. If Statistics NZ's position was a topic should change/be excluded/new topic included, then these were all scored using the framework.

This meant that topics that were almost guaranteed for inclusion (such as age, name, etc) were not required to be put through. Also, some ideas did not go through if prior knowledge (e.g. the topic has come up and been rejected in many previous censuses) meant that it was unlikely to be successful.

How the topics were scored

Once it was determined which topics required scoring, topics were divided up and individual team members were responsible for scoring the topic as a first attempt. The individual responsible would then take the core group through their scores and the reasoning behind it. Some adjustments would typically be made as a result of these discussions. Once the core group had agreed to the score, they

moved onto the next topic. While most topics would take 15-30 minutes, more complex ones could take over 2 hours to come to an agreement.

Once all topics were scored, they were grouped into the following categories:

Less than 60 percent	Excluded
60 to 69 percent	More work required to recommend inclusion/exclusion
70 to 74 percent	Recommended for inclusion but some issues may need resolving
75+ percent	Recommended for inclusion

The cut off points for each category were agreed on following the completion of all topic scoring. The groups above are the same ones that were used in both 2006 and 2011/13.

Topics in the 60 to 69 percent categories were then included in an 'Initiatives plan'. The [plan](#) outlined the further work required before recommending inclusion or exclusion.

Assumptions made when using the content determination framework

It should be noted that some flexibility was required when using the content determination framework. Some practice scoring of topics was required before a workable and consistent approach to scoring was found. The core team also needed to adjust to aspects that were not working well or efficiently throughout the process. For instance, the scoring of entire topics (new content) was considered separately to scoring small changes to existing questions (existing content). This was because it became difficult to compare the results from each. See appendix 2 for more details regarding the assumptions.

Discussion

It is our view that overall the content determination process was successful. It allowed for a consistent and methodical approach in evaluating content for the 2018 Census. While it did take some time early on to apply that consistency (and work out all the assumptions used), once these were resolved the process was efficient. However, should a similar system be used in any future censuses, some questions to ask are:

- 1) Should the content determination process have been done earlier? One issue that arose was when scoring suggested that a topic should be excluded, but the preliminary view recommended inclusion (and vice-versa). However, at times this may have been due to good quality information being received in the submissions and engagement process.
- 2) On some occasions it was noted that while topics would be a valuable inclusion, there was not enough information to know if a question would work well enough in practice. Testing or further research was required. Would it have been good to try and flag these topics prior to the content determination process, in order to allow for better information when it came to scoring? This may have resulted in less topics being in the "more work required for inclusion" category.

Results

This paper does not intend to go into an in depth discussion around the results of the content determination process. However, the table below outlines which band each topic fitted into following the conclusion of the content determination process. It should be noted that these are the

final results – following meeting with both the wider and working groups. For information on exact scoring and how results changed as a result of discussions with the wider and working groups, see [this link](#).

Currently recommending exclusion:
Veteran population
Generational attachment
Citizenship
Mortgage payment amount
Birthplace addition
Language proficiency
Well-being indicator
Paternity
Ownership of other dwellings
Industry training
Private dwellings in retirement villages
More content development and evidence required to recommend inclusion:
Stepfamilies
Types of unoccupied dwellings
Second residence
Sexual orientation/Gender identity
Housing quality
Cigarette smoking behaviour
Access to telecommunications – collect information on internet quality
Post school qualification (collect overseas qualifications separately)
Currently recommended for inclusion, but some minor issues to resolve
Unpaid activities (volunteering + hours spent)
Legally registered relationship status
Main means of travel to education + Educational institution
Fuel types used to heat dwellings (appliances)
Currently recommended for inclusion
Usual residence one year ago
Years at usual residence
Number of children born alive

Appendix one

Score	Second residence	Weighting	Total
3.5	<p>Does the proposed change add value to New Zealand's society and economy?</p> <ul style="list-style-type: none"> Helps with civil emergency management Greater understanding of children in shared custody arrangements Greater understanding of regional population fluctuations Housing stock information 	4	14/20
4.5	<p>Is the census the most appropriate information source?</p> <ul style="list-style-type: none"> Difficult to get from other sources Needed to low levels of data for emergency planning Maybe not wide relevance across the NZ population 	3	13.5/15
4	<p>Does the proposed change reflect an enduring information need?</p> <ul style="list-style-type: none"> Potentially in terms of changing family dynamics and home ownership 	2	8/10
2.5	<p>Will the proposed change produce quality information?</p> <ul style="list-style-type: none"> Could be a number of issues as seen by testing in Australia, such as invalid responses (wrong address, using usual address, time spent at address not computing). Was address fatigue which saw issues with response rates Will children's second address be filled out correctly by one parent? 	4	10/20
3*	<p>Is there continuity with previous census data?</p> <ul style="list-style-type: none"> Not in the previous census 	1	3/5
3	<p>Is data consistent with other data collections?</p> <ul style="list-style-type: none"> No other data collected about this, but is just as address, so should be consistent with the other address categories 	1	3/5
3.5	<p>Is there general acceptance of the proposed change?</p> <ul style="list-style-type: none"> Not offensive, but some people may find slightly intrusive with another address question. Most happy to attempt to answer (as in Australia) although unsure how much value they will see from it 	2	7/10
2	<p>Would the proposed questions be easy for respondents to complete?</p> <ul style="list-style-type: none"> Depends on the exact questions asked, but a number of issues were had in Australia's testing when answering this question 	1	2/5
	Total		60.5/90

Final score (%)

67%

As mentioned in the main part of this document, results were split and compared depending on whether we were scoring the inclusion/exclusion of an entire topic, or if it was a change to existing content. For the latter, it was the change in relation to the original content that is being scored, not the entire topic. The table below outlines the assumptions for each criteria in the content determination framework. Note that all are ranked from 0 (low) to 5 (high).

Criteria	Inclusion/exclusion	Changes to existing content
Does the proposed change add value to New Zealand's society and economy?	0 was given if it provided no value, 5 if high value (e.g. name and sex variables).	A 2.5 was given if there was no extra value. A 2 would mean a slight decrease in value, a 3 a slight increase. This was done as so many changes were minor, most would get a low value (e.g. 1 or 2) if scored the other way.
Is the census the most appropriate information source?	0 if not appropriate, 5 if only place it can and should be collected.	
Does the proposed change reflect an enduring information need?	Some difficulties arose when it came to determining how to take the census transformation strategy into account. This criteria became solely about if the information would still be relevant in the not to distant future (e.g. smoking relevant now, but will it be in 10 years time?).	
Will the proposed change produce quality information?	Same logic as the value question.	As with the value question, 2.5 was used as a starting point (if the change would produce information of the same quality).
Is there continuity with previous census data?	For proposed new topics, this question was not relevant. A 3 was given in these instances.	A 5 was given if the change would not affect continuity, and 0 if all previous data was not comparable.
Is data consistent with other data collections?	This was given a 3 if there were no inconsistencies with other data (ie the data is not collected elsewhere). The logic behind this is that it would not reduce the chance of the topic at least getting into the "needs further work band" if the question was not relevant.	
Is there general acceptance of the proposed change?	0 is a low acceptance, 5 high	A 2.5 midpoint is where acceptance would remain the same. 3 is more acceptable, 2 less.
Would the proposed questions be easy for respondents to complete?	0 is difficult to complete, 5 easy	Again a 2.5 midpoint is where the question would be as easy/difficult to complete as the original question(s).