

2018 Census External Data Quality Panel: Minutes of Meeting on 23 April 2019

Date and time	23 April 2019, 9am to 3:30pm
Location	AUT 46 Wakefield Street Auckland
Present - panel members	Alison Reid –chaired meeting Barry Milne Donna Cormack Ian Cope Len Cook Thomas Lumley (had to leave at 12:30)
Present – Stats NZ	Adele Quinn, Manager Census Analytics Carol Slappendel, Deputy Government Statistician Christine Bycroft, Principal Statistician Gareth Meech, Senior Manager Census Data Quality/ panel secretariat Kathy Connolly, General Manager Census Steph Prosser, Senior Analyst Census
Absent	Richard Bedford – co-chair Tahu Kukutai, panel member Vince Galvin, Chief Methodologist

Alison Reid chaired this meeting.

Review previous minutes and action points

While reviewing the 12 April minutes, it was noted the most recent version 0.3 was uploaded to the workspace the night before the meeting. The secretariat had uploaded v0.2 without taking the co-chairs final comments into account. The panel discussed the minor differences between versions and confirmed that the 12 April 2019 minutes contained a fair summary of discussions at the March meeting.

The action points were worked through with the action points closed but left (and greyed) in the minutes for completeness will be deleted. The remaining action points list have been rolled into this set of minutes.

Report back from 8 May in-committee meeting and additional feedback about 29 April announcement

At the 8 May meeting, the chair reported back that the panel discussed:

- 29 April announcement de-brief. They noted that the Minister was active that day. They had concerns about whether Stats was being too favourable with comparisons with 2013 Census data.
- Technical seminars. Some panel members had been to one of the public sessions while most had watched the webinar version of the same seminar. They were pleased to see the webinar is available on the website without barriers to access (ie. registering a name).
- Focus for the panel in the next period. The panel wants to understand Stats concepts of quality, they need as much as information as possible about the admin data sources used. The panel will likely want to get access to the data in some way.
- The panel is also likely (for the next couple of months) to prefer to meet in smaller groups or have phone calls, or have 'clinics' with selected Stats people to get more information about a specific topic.
- The panel is looking to do some planning for the report targeting a 23 Sept release date. They are considering whether they also get an external – to the panel – reviewer to peer review the report prior to publication. An independent review of the independent review would need to have an agreed process. The panel will be working hard on the report in July and August and will need to be

- in regular touch (but not necessarily all-day meetings) with Stats people.
- Kathy offered project management and/or editorial support for the report. The chair responded to say that they are building copy editor time in but may be interested in getting some help from Stats.
- Carol stated that Stats is looking to update customers in July and she is pushing for the organisation to be as transparent as possible about data quality.

Programme update

Kathy delivered a short session about the progress of the programme and covered the following key points:

- The background to the 29 April decision document
- Some panel members noted that the media release used the word 'comprehensive' when comparing to 2013 Census data quality. The members felt this was misleading the public and putting Stats 'out on a limb'. Future language needs to be much clearer that any improvement in data is around population structure and not characteristics.
- Kathy let people know that Stats briefed the Stats Minister and also two opposition MPs.
- When looking at the critical path, a question was raised about what the 'Data linkage survey starts' is. Christine said that 'survey' should say 'sample' and is part of Plan C workstream. It was also noted that the critical path should be updated with milestones involved with the publishing of the independent report.

AP8-1 Update critical path with independent panel report release, and milestones – after the 10 June in-committee meeting creates a schedule.

- The technical seminars and webinar were only briefly covered, but at least two panel members attended and made positive comments about the content and attendance of the presentations.

Admin data for non-response in prisons and defence establishments

Christine gave an overview of the updated methodology that was attached as a pre reading paper for this meeting - the 'Admin data for non-response in prisons and defence establishments' paper. Some questions were raised including:

- Why the institution ID variable was received in the dataset provided by Corrections but not in the admin data records? Was there a reason? Do we have an ethical basis on receiving it when it was not provided as part of the admin data going into the IDI and referenced in the PIA. Does that breach the original agreement?
- Stats responded to say we adhere to all the MOUs in place between data providers and that the MOUs are up to date. There could be several reasons to why the institution field was not included.

AP8-2 Find out and then update the panel about the reason the prisons data provided to the IDI does not have the institution ID variable.

- The panel again reiterated that they are not entirely comfortable with the use of Corrections data for ethnicity, and have concerns about the quality of that data.
- When discussing the Results slide, the panel noted that this would indicate a drop in prisoner numbers and could cast doubt on the credibility of the census – especially in this current political environment. Stat responded to say that this data is not the official prisoner count, this is the census dataset – which is used by customers in specific ways.
- A panel member noted that the word 'anonymise' is used several times and is confusing. Christine agreed and will re-word.

AP8-3 Replace the use of the term 'anonymise' in the 'Admin data for non-response in prisons and defence establishments' paper for a more accurate and clearer term.

AP8-4 Update prisons and defence establishments methodology paper to better reflect quality of ethnicity and note how many people we keep the prisons recorded ethnicity for.

Data quality framework and customer use update

Gareth started the presentation with a reminder about the Quality Management Strategy that was introduced back in the first external data quality panel meeting in August. He then went on to describe the key customers that were drawn from the customer use index. The panel noted that uses such as mortality, crowding and other derivations were not present in the key use cases. They asked for a richer summary rather than the spreadsheet. Gareth referred the panel back to a previously presented document 'Fit for

purpose engagement summary as at 12 Feb 2019' (in shared folder Meetings\Panel meeting 5 – 14 Feb 2019) for a richer summary.

AP8-5 Panel to let Stats know if more information about census uses is required.

Data evaluations

Adele presented a session describing the census data evaluation process and examples. A question was asked about whether the data evaluated included admin sources or was response data only – Adele responded that the full dataset is used during evaluations. After Adele presented on the Warrant of Fitness tool, there were discussions about the comparability of 2018 and 2013 data. There was also a question about which cross-tabs are analysed given the variable based approach. Stats responded that for each variable age, sex and ethnicity cross tabs are run and analysed. In addition each variable checks which cross tabs were contained in 2013 NZ.Stats tables for other content requirements. For key demographic variables, additional analysis including migration and cohort analysis is also conducted. A request was made by the panel to update the shared space glossary with terms used in the session including 'macro-checks'.

The 2013 Information by variable metadata product was discussed and asked for it to be linked in the shared workspace.

AP8-6 Add a link to the 2013 Information by variable metadata product to the shared workspace.

AP8-7 Panel request to see proposed data products and metadata products in a future meeting.

AP8-8 Panel request to access the warrant of fitness reports.

Overall, the panel noted they were impressed with the evaluations approach described.

Quality rating scale

Adele presented on the 2013 and 2018 Quality Rating Scale (QRS) for variables. She noted that in 2013 the individual and overall ratings were not always strictly adhered to and sometimes a range of quality ratings was given for a variable (eg. Moderate-High). Stats noted that the 2018 approach is more mechanistic but is being reviewed due to several factors when arriving at the three specific and one overall rating. A question was raised whether we could show the QRS rating at sub-populations eg geography. Stats responded that we could be done but would take a lot of work.

During the discussion on data sources and coverage examples, questions were raised by the panel:

- When calculating the data source weighting, did Stats separate into ethnic groups? Stats responded no, that any more detail would have been too complex and one overall rating has been used. A panel member suggested Stats look at level 1 ethnicity separately.

AP8-9 Stats to look at how using level 1 ethnicities as well as subnational count could impact the ratings.

- Given the ratings, would it be best to use different hierarchy order of quality of data sources when filling missing data? Stats responded that a decision had already been made that the order of use of data sources would be 2018 Census data, 2013 Census data, admin enumeration and then imputation. The suggestion for the order to not be fixed but based on the quality of the source was a good suggestion, and something Stats NZ should consider for 2023.

When presenting the slides on individual form variables source breakdown, there was a question about whether we could produce the same slide with 2013 comparisons next to it and be able to break them down to regions and ethnicity. Stats responded to say that a pivot table is available that breaks this down to ethnicity and region. The 2013 comparison would take a little bit of work. It was also noted that the Post school qualification field of study bar on the source patterns graph is incorrect – this was noted and will be corrected.

AP8-10 Stats to load the pivot table version of the individual form source breakdown graphic that allows members to toggle level 1 ethnicity and region. Also provide national level with a 2013 comparison.

The panel noted that these two graphs are powerful and show the difference between iwi and other variables.

Challenges with the application of the 2018 quality rating scale were presented by Stats. The panel suggested that there is value in showing the overall and individual ratings – audience segmentation will determine who needs to see detail. Most users will only need a statement, so keep the scales simple. The panel also suggested that it is better keep the rating system comparable to 2013 – better to stick with something previously done to increase trust in user community.

A request was made by the panel for Stats to publish our detailed quality evaluation process in full so the panel can refer to it in their report.

AP8-11 Stats to add a document describing data evaluation process to the list of papers to be published so that the independent report can refer to it in its findings.

The graph showing the interim quality ratings was presented. Several panel members said they liked this graph and it was helpful to tell the story of the data. Stats noted that the 2018 data in the graph contained Absentees which is not an output variable so will be removed from future versions.

When discussing the new variables around damp and mould, it was noted that the benchmarks being used by the evaluations team were using the 2017 test data. A panel member suggested that Stats could re-test the reliability of the 2017 data and should also ask experts in their field. Phillipa Howden-Chapman was suggested an expert that could help.

AP8-12 Stats to consider whether re-testing whether 2017 test data could be used as a benchmark for damp and mould.

Data quality decisions making and output options

Gareth presented on the guidelines and considerations being used to determine what data is publishable and how it may be supported. During the discussion about guidelines, the panel noted that there is still value in lower quality data, but need controls to ensure correct usage, or at least minimise misuse. Metrics should include an equity measure – equity in trying to correct the inequity of data collection in the 2018. This should also be considered in the risk and impact assessment process. In the example of iwi data, discussion is needed to work through what use and benefits can be realised.

Principle should be to release as much as possible with suitable caveats and metadata – transparency is very important for this census. Variables may need to be segmented into population structure and characteristic groups and different actions needed for both.

When discussing the risk and impact assessment process, the panel made the point that it is extremely important to understand customer uses well. When discussing whether to withhold data, the panel suggested that as much data as possible goes into the IDI and that withheld could be subject to Official Information Act requests.

When discussing whether it would be ok to still work on some data after first release – for example if we could improve once in the datalab with more recent data sources or new methods. The panel noted that there is a difference between early data that we think is correct but needs to be confirmed compared to producing experimental data. Whatever type of data is produced justification will be needed.

Output rules could be applied in a datalab environment that gives customers access to the full range of data but that rules like collapsing categories could be developed (on top of confidentiality rules). The panel suggested we need to create some worked examples for each quality rating.

Customer education will be important to help them use the data and understand limitations. Stats

mentioned that the programme is working on an Open API (Application Programming Interface) that would allow external customers to directly query the census unit record dataset that will produce automatically confidentialised tables. There was some interest in that possibility among the panel, and there could be a lot of interest from technical customers. Kathy noted that Stats hadn't talked about it much as there are still some development challenges. This may impact whether customers need to put in a customised request, so we need to give them notice but be sure of the publication date.

The panel noted again they would like to see the approach to products and services and metadata products (previously noted as AP8-6).

Specific data sources

Stats was only able to upload an updated spreadsheet that detailed the specific IDI variables used for admin enumeration the night before the meeting. Adele took questions from the panel:

- The panel asked whether the ethnicity data from MOE data came from secondary or tertiary data, as the different sources have different perceived data quality. To be actioned and updated in the spreadsheet.
- Suggested adding sorting aids to the spreadsheet.

AP8-13 Stats to update spreadsheet as per above notes.

Panel plans for independent report

The meeting chair discussed the next steps for the independent report (described earlier in the minutes). Stats asked what areas the panel requires to be able to write their report. The chair noted that the project planning to be done in the next two weeks will help determine what is needed and when. The panel asked when the international peer review of the methods was going to happen outside of the external reviewers report.

There was a brief discussion about 2023 Census business case and planning and whether some panel members could talk to Stats people who have been planning the Māori and iwi engagement.

The next meetings dates are:

- in-committee 10 June in Auckland
- ON HOLD: full meeting 24 June in Wellington, likely to be removed once panel report planning started
- later meetings to be confirmed

Action log

Ref	Date raised	Description	Owner	Date required	Progress	Status
Meeting 4 actions						
AP 4-9	7/12/18	Update panel once more detail has been completed on the quality framework including an ordering of decision making criteria.	Gareth	By 14 Feb meeting	14/2 Planned item on 6 March agenda. 6/3 Variable quality rating scale paper to be sent to panel before 12 April. 11/4 Paper loaded, set time in May agenda. Leave open. 23/5 Quality rating, data evaluations discussed, chair to decide whether to close – Alison?	Open
Meeting 6 actions						

AP6-1	6/3/19	Provide the panel information on smoking question data as an example of a variable without other admin enumeration sources.	Adele	By 12 April	11/4 Unable to provide counts of lower priority variables for April meeting. Suggest smoking data and associated quality rating topic for May meeting. 23/5 Smoking Warrant of fitness had not been completed in time for meeting, age WOF used in presentation instead. Panel has requested access to all WOFs, suggest close here as new action AP8-7 will cover smoking.	Close
AP6-2	6/3/19	Present list of papers and approx. completion (or draft) timing so panel can base own reports off Stats documents.	Gareth	By 12 April	11/4 Updated list loaded to shared space however panel requires more specific dates quickly to know when to book upcoming meetings. Leave open. 23/5 No more information available about method paper programme or dates. To be forwarded to panel by 10 June. Panel noted they would like to see drafts earlier rather than wait for final versions.	Open
Meeting 7 actions						
AP7-1	12/4/19	Produce document about what the IDI spine is and is not. Include date ranges for each spine data source used in the 2018 Census.	Christine	By 22 May	23/5 Document not produced in time for meeting. To be completed in June, exact date to provided to panel by 10 June. Christine noted that a paper covering the timeliness of data sources used for admin enumeration is being drafted.	Open
AP7-2	12/4/19	Analysis of ethnic distribution for NZ European and Asian ethnic groups and the	Adele	By 22 May	23/5 Not yet completed. Target date now 14 June. Adele to check	Open

		percentage of each source that contributes to Māori descent.				
AP7-4	12/4/19	Update method paper timeline spreadsheet.	Gareth	By 29 April	23/5 Duplicate of AP6-2, close.	Close
Meeting 8 actions						
AP8-1	23/5/19	Update critical path with panel independent report milestones	Kathy /Alison	By 14 June		Open
AP8-2	23/5/19	Update the panel about the reason the prisons data provided to the IDI does not have the institution ID variable	Adele	By 14 June		Open
AP8-3	23/5/19	Replace the use of the term 'anonymise' in the 'Admin data for non-response in prisons and defence establishments' paper for a more accurate and clearer term	Christine	By 14 June		Open
AP8-4	23/5/19	Update prisons and defence establishments methodology paper to better reflect quality around ethnicity and note how many people we keep the prisons recorded ethnicity for	Christine	By 14 June		Open
AP8-5	23/5/19	Panel to let Stats know if more information about census uses is required	Alison	By 14 June		Open
AP8-6	23/5/19	Add a link to the 2013 Information by variable metadata product to the shared workspace	Gareth	By 14 June		Open
AP8-7	23/5/19	Panel request to see proposed data products and metadata products in a future meeting	Gareth	By 30 June	23/5 Gareth to first confirm date a plan can be given to the panel.	Open
AP8-8	23/5/19	Give panel access the warrant of fitness reports	Gareth	By 14 June		Open
AP8-9	23/5/19	Determine how using level 1 ethnicities as well as subnational	Adele	By 30 June		Open

		count could impact the ratings				
AP8-10	23/5/19	Load the pivot table version of the individual form source breakdown graphic that allows members to toggle level 1 ethnicity and region. Also provide national level with a 2013 comparison	Gareth /Adele	By 30 June (after dataset is finalised in mid June)		Open
AP8-11	23/5/19	Stats to add a document describing data evaluation process to the list of papers to be published so that the independent report can refer to it in its findings	Gareth	By 30 June		Open
AP8-12	23/5/19	Report back to whether re-testing whether 2017 test data could be used as a benchmark for damp and mould	Steph	By 30 June		Open
AP8-13	23/5/19	Update data source spreadsheet to answer whether the ethnicity data from MOE data came from secondary or tertiary data and add sorting aids to the spreadsheet.	Adele	By 14 June		Open

Released under the Official Information Act