Stats NZ
Tatauranga Aotearoa

2023 Post-enumeration Survey: High-level design

Citation
Stats NZ (2023). *2023 Post-enumeration Survey: High-level design*. Retrieved from www.stats.govt.nz.

ISBN 978-1-99-104930-8

Published in March 2023 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

List of tables and figures

List of tables

List of figures

Purpose and introduction

2023 Post-enumeration Survey: High-level design outlines the key methodological elements of the survey and how they will be implemented.

It details the high-level design for the 2023 Post-enumeration Survey (PES), administered by the Census Coverage project. It includes an overview of the key methodological elements that will enable the delivery of the PES strategy, vision, goals, and investment objectives; how they will be implemented; and how progress will be measured.

Introduction

This document provides an overarching perspective on the Census Coverage project.
By nature, a large multi-year project evolves through its lifecycle as strategies get refined, environmental changes require adaptation, and engagement with stakeholders leads to shifting priorities. This document therefore reflects a snapshot of the project during continuous delivery, operation, and assessment.

There may need to be a change from the high-level design as specified here, to mitigate risk, resolve issues, adapt to new knowledge, adjust to funding changes or respond to changes in the macro-environment. This then becomes a managed change process, where all the implications of the change are spelt out and agreed upon through the census coverage governance framework, as outlined in the next section. For example, there will be impacts on the PES design arising from Cyclone Gabrielle and the subsequent extension and amendments to census operations in affected areas. Any design changes resulting from this recent cyclone will be outlined in the PES adaptive design.

There are several companion documents that provide more detail on their respective topics:

- Standard design – documents the planned design for coverage estimation of the 2023 Census
- Linking design – documents the design of the process to link respondents of the Post-enumeration Survey to corresponding census records
- Adaptive design – due to be published late 2023, documents the statistical risks to the 2023 Census Coverage project, how they can be identified and monitored, and the planned response to each.

About the 2023 Census Coverage project

The 2023 Census Coverage project will estimate the coverage error in the census using the Post-enumeration Survey (PES), with interviews carried out at around 16,500 dwellings across the country. The 2023 Census programme will be the sixth census cycle in which Stats NZ has conducted a PES, and it remains in line with international best practice as the gold standard for measuring census coverage. Throughout this document, census coverage refers to the project as a whole, while PES is used to refer specifically to the survey.

The 2023 Census of Population and Dwellings is a substantial programme, with the goal of collecting high-quality data on every eligible person in Aotearoa New Zealand on census night, 7 March 2023. To provide the population with sufficient opportunity to participate, there are several ways in which one can fill out a census form. Despite this, it is untenable to assume that everyone will participate. The census design will account for this as much as possible through administrative sources, such as those from health, immigration, education, and tax. However, a portion of the population will still be missed, or inadvertently counted more than once (situations referred to as under-coverage and over-coverage, respectively).

A vital aspect of coverage error is that it is not homogenous. While the net under-coverage for the 2018 Census was around 2.6 percent, people of Māori descent were undercounted by more than 4 percent. Therefore, it is required that a detailed estimate of the coverage error in the 2023 Census is appropriately measured.

Expected project outcomes

The 2023 Census Coverage project has two primary goals.

- Provide the official results to assess the level of people coverage and response rates of the 2023 Census programme
- Provide a measure of coverage adjustment (census adjustment ratio (CAR)) or the scaling factor by which each stratum of the population count should be modified to account for coverage error. This is a key input to the estimated resident population (ERP).

Measuring coverage: Dual-system estimation

Census under- and over-coverage is estimated with data from the PES using an adapted dual-system estimation approach consisting of linked census and PES responses. Dual-system estimation is a method of estimating the size of a target population from two partially incomplete lists of the population. It is used frequently in population estimates across disciplines, such as in ecology and epidemiology. The 2018 Census used dual-system estimation to estimate the usually resident population in New Zealand on census night using census responses and administrative population lists. For a more detailed description of dual-system estimation and its application in this context, refer to section 1.2 of 2023 Post-enumeration Survey: Standard design for coverage estimation.

Applications of dual-system estimation commonly invoke several assumptions:

- independence between the two lists (the likelihood of being recorded on the PES list has no relationship with the likelihood of being recorded on the census list)

- a closed target population (the PES is measuring a sample of the same population the census is targeting; people cannot leave or enter the population)
- homogeneity of capture of individuals (all individuals have the same likelihood of being captured in a list)
- no undetectable erroneous inclusions in either of the two lists (no people included that are not part of the target population) or undetectable duplicates
- perfect linking between the two lists.

Dual-system estimation approaches are highly sensitive to these five assumptions being valid. If these assumptions are violated, resultant coverage estimates are at critical risk of bias. Being able to validate these assumptions is critical to the success of the project and delivering robust results to support production of Stats NZ's suite of population estimates and projections. The adaptive design details the project's approach to detecting and resolving pressure on these assumptions.

Relationship with the census programme

Independence from the census programme

The census coverage project sits within the wider 2023 Census programme and is funded out of the census multi-year appropriation. However, it maintains strict independence from census. This limits the introduction of bias in census and enhances public trust in an independent measurement of census, a large, publicly funded project. To protect the requirement for independence, it has its own governance board and management, outside of census.

Independence from the census programme is maintained throughout the project in these key areas.

1. **A separate dwelling frame**. In 2018, this was achieved by using a commercially available cadastral register (from homes.co.nz). This was not fit for purpose, so for the 2023 cycle the Statistical Location Register (Stats NZ in-house list of addresses) was used as the base. The census programme also used the Statistical Location Register as the base for their dwelling frame, but the methodology for defining dwellings was different; the PES dwelling frame was tailored towards over-coverage, and two rounds of dwelling enumeration were performed to maximise accuracy. First, a desktop canvassing operation used online and bespoke tools to canvass around 85 percent of the sample areas. Then, 100 percent of the sample areas were canvassed in the field. This provides clear separation between the PES and census dwelling frames, despite sharing an origin.
2. **Secrecy and security regarding the areas in the PES sample**. All documents and tools that contain specific sample information have been stored in secure locations separate to census and access to them is heavily restricted to approved project members.
3. **Non-overlapping field collection periods and collection staff**. A Memorandum of Understanding was agreed between the 2023 Census programme and the 2023 Census Coverage project that recognised the need for separate periods of field collection. It was agreed that all planned census field activities would be provided to the Census Coverage team in advance, as well as notification of changes to planned activities, so that the PES operations could be organised in a way to avoid them. Collection staff would not be assigned to the census operation if they have already worked on the PES operation (or vice versa). However, an exception to this would be made for staff able to work in different geographic areas for both.
4. **Secure data storage, field management, and processing environment with access restricted to approved coverage staff**. In 2018, this was achieved with separate instances of

key software. For the 2023 cycle, census coverage will share the corporate instances, but access groups will be tightly controlled and thoroughly tested prior to going into production.

The Census Coverage project has considered independence during every design decision, as violation of independence strains any estimation methodology.

Measurement of 2023 Census key performance indicators

The Census Coverage project is responsible for measuring several of the census key performance indicators (KPIs):

- KPI 1: national response rate for total population ≥90 percent
- KPI 2: national response rate for people of Māori descent ≥90 percent
- KPI 3: national net coverage rate for total population ≥98 percent
- KPI 4: national net coverage rate for people of Māori descent ≥98 percent
- KPI 8b: national response rate for Pacific ≥90 percent.

See Measuring the success of the 2023 Census – key performance indicators for the full list of KPIs.

The report of the achievement/non-achievement of these KPIs will be released alongside the coverage report, scheduled for later 2024. Interim response and coverage assessments will be produced by the 2023 Census Programme, prior to the availability of data from the PES.

Wider context and change overview

2018 Census

Following minimal-change censuses in 2006 and 2013, the 2018 Census faced a significant need for change if it were to remain sustainable over time (as explained in the 2018 Census strategy).

The drivers for change were:

- changing make-up of New Zealand resulting in challenges for achieving quality data especially for small areas and subpopulations
- increasing cost and difficulty in running a traditional field-only census, given the availability of comparatively more efficient digital options
- increasing availability of alternative administrative data sources and collection technology
- need to address and modernise census systems that were based on outdated technology.

The key changes to 2018 Census which affected the 2018 Census Coverage project were:

- address canvassing – prior to 2018, the census dwelling frame was built using a full scratch dwelling enumeration; for the 2018 cycle, census used a list-checking enumeration method, matching that traditionally used by PES
- online-first collection – instead of having collector visitation or forms delivered to every dwelling, 79.7 percent of households were able to complete the census via an online form by using an access code which was received my mail
- the introduction of supplementary data from administrative sources.

2018 Census Coverage project

The primary goals of the 2018 Census Coverage project were to:

- provide fit-for-purpose measures of undercount and overcount in the 2018 Census of Population and Dwellings
- deliver estimates of net census undercount for input to the 2018-base ERP, to agreed quality standards and within specified timeframes
- modernise, and adapt to significant change in census methods while maintaining the independence of the PES
- align census coverage practice and processes with relevant corporate systems and methods
- contribute to development of census coverage methods beyond the 2018 Census.

Key changes from previous PESs included:

- changes to the dwelling frame creation strategy – due to the change in the census approach to creating a dwelling frame, the PES pivoted to an alternative approach, beginning with a commercially available cadastral register (from homes.co.nz), supported by a field enumeration operation
- changes to the questionnaire delivery – prior to 2018, responses to the PES were collected using paper questionnaires; for the 2018 cycle, an electronic questionnaire was used to collect respondent information via a tablet.

Despite challenges during collection and processing, the 2018 Census Coverage project achieved the following outcomes:

- key quality targets were met, including an achieved survey response rate of 91 percent (exceeding the 90 percent target), and an international peer review of the PES estimation method
- coverage estimates from the PES were used as input to the ERP, and for reporting against census programme KPIs
- greater disaggregation of coverage estimates than had previously been possible, enabled by a small-domain estimation approach
- the PES was updated to use corporate tools, and modernised, including the introduction of an e-questionnaire
- independence of the PES was maintained
- research and development work undertaken to measure coverage in a combined census set was a strong platform for the 2023 PES.

Key learnings identified during the 2018 cycle include:

- importance of ensuring a high-quality dwelling frame
- investigating how various estimation methods handle high differentials in coverage and developing adaptable methodologies for a combined-census model
- underlying complexity in creating and maintaining independence across multiple tools, systems, and methods.

2023 Census plan

The high-level design for the 2023 Census programme was published in 2021, see 2023 Census: High level design.

The funded option for the 2023 Census programme is a ‘viable but minimal change’ design, making use of the successful parts of the 2018 Census, changing only what needs to be fixed and extending to deliver census to a bigger and more diverse population. It is a combined model, supporting the traditional full-field enumeration with administrative data to fill in the gaps.

Key areas of focus for 2023 Census Coverage design

The 2023 Census Coverage design is built to work with the approach described in the 2023 Census high-level design. Several assumptions are being made about the 2023 Census (and wider Stats NZ) in the design:

- the 2023 Census programme will closely follow their design described in the 2023 Census high-level design
- the 2023 Census will be held on 7 March 2023
- respondents will be able to provide census responses from mid-February 2023 until the end of April 2023
- the 2023 Census programme will provide an environment (capability and culture) to support a design-led approach
- corporate capabilities (such as tools and skills) will be used where feasible and appropriate
- data collection for the PES operation will be managed within Stats NZ
- microdata activities will be permitted pursuant to the census coverage privacy impact assessment; key actions in this space include the linking of PES results to the IDI spine to identify false-negative links to the census file and border crossing information to identify residents temporarily overseas.

Like the census programme, achievements from the 2018 cycle will be built on and reflect on lessons learned to drive methodological improvement and enhanced overall value.

The key priorities that have been identified are:

- ensuring a high-quality dwelling frame
- having an effective, assured, and error-free questionnaire and field tool
- establishing robust ways of assessing and supporting key estimation assumptions
- research and improvements on the ability of estimation methods to handle high differentials in coverage
- achieving adaptable methodologies to deal with a combined-census model.

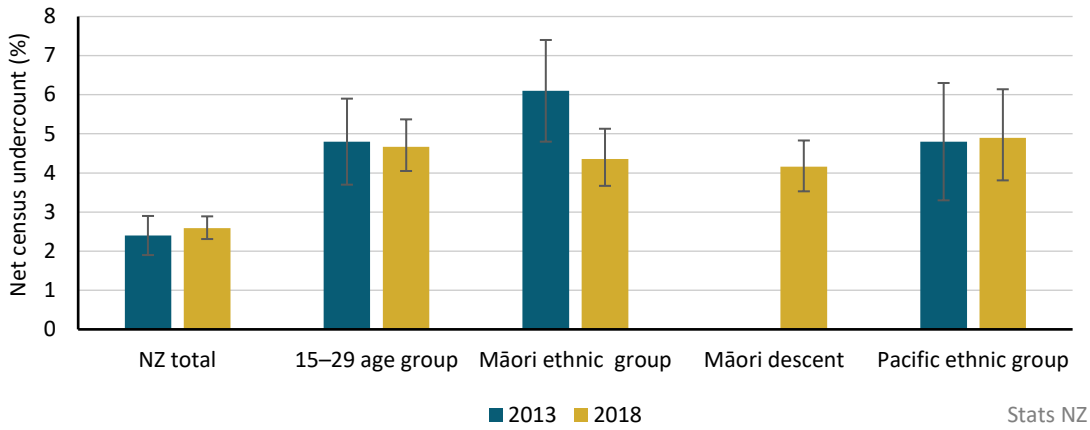
2023 Census coverage design

Delivering for and with Māori

It has been consistently shown in previous census coverage projects that coverage patterns vary across subgroups. The demographic characteristics most undercounted are Māori, Pacific peoples, and young adults (see figure 1).

Figure 1

Net census undercount for selected population groups, 2013 and 2018



In order to provide high-quality estimates, the PES has been targeted to better cover demographic characteristics historically undercounted by the census, which align with the census-defined priority response groups. This section gives an overview of some of the steps taken to provide better coverage estimates for Māori and the other census priority response groups.

Relevant design considerations

Sample design

The sample frame is the list of primary sampling units (geographic areas of 60–150 dwellings) that will be enumerated during the PES. As explained above, the PES is intentionally targeted towards census priority response groups. Therefore, the selection process monitored the inclusion rates of subpopulation indicators, such as the estimated percentage of the primary sampling unit who identify as Māori or Pacific, and presence of emergency or transitional housing facilities. The oversampling is not explicit; it is achieved implicitly through primary sampling unit selection. The selection process is based on pre-determined strata and is designed to maximise the probability of primary sampling units with high rates of Māori or Pacific being included in the sample frame.

Table 1 below shows the extent of over-sampling for the 2023 PES sample frame. The contents are explained in more detail in section 3.07 of 2023 Post-enumeration Survey: Standard design for coverage estimation.

Table 1

Oversampling in the 2023 PES sample frame			
Summary measure	**Subgroup**	**Regional council**	**TALB**
% geographic areas with evidence of over-sampling	Māori descent	88%	57%
	Māori ethnicity	81%	56%
	Pacific ethnicity	44%	45%
	Asian ethnicity	50%	45%
Average degree of over-sampling across the geographic areas	Māori descent	3.6%	2.3%
	Māori ethnicity	3.3%	1.6%
	Pacific ethnicity	-0.2%	-0.8%
	Asian ethnicity	0.6%	1.7%

Note: People of Pacific ethnicities were over-sampled in areas of high-Pacific concentrations and had lower sampling rates for the rest of the country, which explains why the average over-sampling per geographic area is less than zero.
TALB – Territorial authority / local board, a subnational geographic breakdown used by Stats NZ.

Source: Stats NZ

The PES sample design also coordinates with census on areas where collection will be led by Te Mana Whakatipu. Initially five areas were considered in the design for 2023 Census, with three areas confirmed for the final design. The 2023 PES has included this in the sample design assessments and is following the iwi-led design plans closely.

Standard design

The standard design is the written record of the statistical approach through which the coverage estimate is generated. In this section we discuss the several components that are relevant to priority response groups.

Differences between PES and census responses

Differences in response to equivalent variables in the census and PES record from the same person creates a concern for census coverage: which is the source of truth – PES, census, or a union of the two? This is a particular difficulty with variables such as ethnicity, which is self-identified, and the 'correct' answer may change over time, or there may be several 'correct' answers. Since this is a complex question that requires input from key customers and stakeholders, this is outside the timeline of the 2023 Census Coverage project. Through the development of the 2023 project, however, guidelines will be developed for how to manage conflicting and missing responses.

Alternative data sources and imputation for missing variables

The planned approach for cases where Māori descent and ethnicity variables are missing is still pending external census engagement with Māori. Under the proposed plan, these variables will be sourced from historic census data, administrative data, or statistically imputed to provide complete records in cases where they are missing. Ethnicity can be sourced from records collected by other organisations, such as Department of Internal Affairs (birth registrations), Ministry of Health (NHI), Ministry for Social Development (benefits), and Ministry of Education (tertiary and school enrolments). Similarly, Māori descent information is sourced from Department of Internal Affairs (birth registrations) where necessary and available. Since these are variables required by PES for estimation, the consequence of not sourcing this information is that the record is removed entirely. This is statistically undesirable, since it reduces the size of the sample. It is especially important if there are more missing variables in under-covered populations.

Collection of Māori descent information
The PES will collect Māori descent information in line with census and the statistical standard (that is, there are three response options: yes, don't know, and no), as it best aligns with Māori and iwi data needs and enables measurement of census KPIs. To meet the electoral and ERP needs, populations using the Māori descent electoral classification (that is, two response options: yes and no) will be estimated as part of processing the PES response data.

Linking design

Census coverage assumes linking between PES records and those in the census and administrative sources is perfect in order to generate its estimations. However, Māori, Pacific, and Asian ethnicities were shown in the 2018 cycle to have lower rates of exact linking. In the 2023 cycle, the linking process has been updated. There will be additional passes through automated and clerical linking processes, as well as a triangular linking technique to identify coverage patterns between each of the sources. These additional measures promote the stability of this key assumption for priority response groups.

Adaptive design

The adaptive design details the approach to be taken if any dual system estimation assumptions are shown to be under pressure, or if any statistical components are failing quality controls. There are two assumptions relevant to priority response groups to be considered:

1. missing entries are missing at random, that is, there is no systematic bias in the demographic characteristics of respondents from whom we get no or partial responses
2. independence of the census coverage project from the census project, that is, a record likely to be missing from census is not equally likely to be missing from PES.

The adaptive design will highlight what patterns in the data may be flags that these assumptions are at risk, and what the course of action will be to minimise the impact.

Scope

Alignment with 2023 Census scope statements

The 2023 Census Coverage project is aligning with the scope statements outlined in the census high-level design where possible. In this section, for each of the census scope statements that are relevant to census coverage, the implications of the statement are explained and highlighted where a deviation from alignment with census is made and why. 2023 Census: High level design explains the full list of census scope statements in more detail.

Table 2

Alignment with census scope statements	
Census scope statement	**Census coverage alignment**
Programme-wide scope statements	
Be smart and purposeful with what we change to achieve a better outcome	Alignment has been made with the "validate and fix" approach being taken by census.

Collaborate with Treaty Partners	A specific plan has been created for working with and for Māori, as well as ensuring te ao Māori is present within the design, governance, and quality assurance structure.
A combined census model by design	A plan for coverage measurement of a combined census model by design, has been created. This includes working in collaboration with the census programme to ensure robust coverage measurement is supported in census design decisions.
Maintain and develop a social license	Engagement with the census programme has occurred and the assumption about social and cultural licence is maintained and extends to coverage measurement.
Ethical use of data	The ethical use of the collected and administrative data we use in this project has been promoted. Consideration has been given to social licence and data practice frameworks such as Ngā Tikanga Paihere to guide the use of the data.
A respondent-centric census	Respondent experience is a high priority for us. Work with collection operations has occurred to minimise the burden to respondents while still achieving the data quality and content required.
Privacy, confidentiality, and security by design	This is of the utmost importance to the project. PES has published a privacy impact assessment (see 2023 Post-enumeration Survey: Privacy impact assessment); further details in the section below.
Robust testing	Following the census programme Dress Rehearsal in 2022, end to end testing has been undertaken. Additional testing of specific elements will be carried out as required.
Data quality by design	A focus on data quality during key stages and at key handover points throughout the project, is prioritised. Data quality across the Census Coverage project is subjected to project-level, wider Stats NZ, and external quality assurance processes.
Information and output-quality need scope statements	
New topics restricted to gender and sexual identity only	For the 2023 PES, a pivot from collecting sex to gender will be made, in accordance with the data standard, new since the 2018 Census cycle (see Data standard for gender, sex, and variations of sex characteristics).
Changes to the iwi affiliation and Māori descent topics	Two topics have been updated in the 2023 cycle: iwi affiliation has been included as a new question for the 2023 PES; and Māori descent has been updated to include 'don't know' as a top-level response option, in alignment with the census collection of this information. More information on these changes can be found in the 2023 PES: Standard design for coverage estimation.
More comprehensive and transparent evaluation of output quality	The project outputs will be subjected to thorough internal quality assurance processes prior to publication, as outlined in the outputs and dissemination section below.
Field collection scope statements	
More field staff	An increase in the number of field staff involved in the PES field operation will occur, to account for the observed increase in effort required for data collection for other social surveys plus the increase in sample size over 2018 and to meet project

	objectives. Unlike census, a significant increase in the field operation is not a focus for the Census Coverage project.
Online and paper forms available	Unlike census, the entirety of PES collection will be done via CAPI (computer-assisted personal interview). This is part of the strategy for maximising data quality. Where deemed suitable, outstanding personal questionnaires can be completed via telephone.
A mixture of field hubs and engagement centres	While census will have field hubs and engagement centres to facilitate assisted completion, PES is restricted to a face-to-face interview at the respondents' address. In some cases, a telephone follow-up will be conducted, but only after contact at the doorstep.
A fit-for-purpose address list	The PES address list is a core focus for the project. Multiple substantial operations have been conducted to ensure the quality of the address list.
A re-envisioned approach to recruitment for an operations-ready field force	A key learning from the 2018 operation was to minimise staff turnover in the 2023 operation. This will be a focus of recruitment in this cycle.
Support for New Zealand's official languages	Unlike census, PES does not have full support for all New Zealand's official languages. The PES questionnaire is available only in English, although if the interviewer is fluent and able to translate to another language, they are permitted to do so. PES responses can be collected in both te reo Māori and English, but not in New Zealand Sign Language.
Administrative data collection scope statements	
Administrative data will support data processing, analysis, and quality assurance	Administrative data will be used to support the data processing, analysis, and quality assurance. Administrative data will be linked to PES and census responses (see the 2023 PES: Linking design for more detail) to improve the quality of data imputations. Administrative data will also be used to improve the accuracy of determining which PES and census records are in scope for coverage estimation.
Statistical processing scope statements	
A statistical production environment for administrative data	We utilise the statistical production environment for administrative data to support linking PES data with the census file, and with migration data.
Analysis and dissemination scope statements	
Māori products that meet Māori needs	Delivering for Māori has been a high priority throughout the project. As described above, and in more detail in the PES standard design, PES has a specific plan for producing outputs that meet the needs of Te Tiriti partners.
Source: Stats NZ	

2023 Census Coverage project scope statements

This section provides a list of topics that have been agreed as in-scope for the 2023 Census Coverage project. They are approximately in chronological order from the start of the project to its conclusion and closure, although many will overlap and be acted on in parallel.

Table 3

2023 Census Coverage scope statements	
Number	**Description**
1	Upfront research on the following topics: • dwelling enumeration process • coverage for a combined census • misclassification • non-private dwellings • sample design • questionnaire content • census components
2	Standard coverage design, documented and agreed
3	Adaptive research on the following topics: • PES response rates and small sample sizes • census coverage rates and dependency • treatment of late census returns • introduction of mixture models • uncertainty measures
4	Adaptive design, documented and agreed
5	Provision of business analysis and requirements to technical teams
6	High-level enterprise architecture design
7	Solution design including dataflows
8	Questionnaire design
9	Questionnaire tool
10	Collection plan agreed with the Collection Operations team
11	Provision of casual staff to complete enumeration and interviews for the 2022 field test and the 2022/2023 field operations
12	Collection hardware
13	End-to-end technical solution
14	Field test operations: • dwelling enumeration test (2021) • interview test (2022) • processing test (2023)
15	Confidentiality and privacy impact assessment
16	Security certification and accreditation for purpose-built tools
17	Independence protection approach and strategy
18	Operational support: • collateral • processes • staff
19	Ministerial briefings
20	Detailed operational plan
21	Field operation: • dwelling enumeration (2022) • interview (2023) • processing (2023)
22	Coverage outputs
23	Communications and engagement strategy

24	Project closure
Source: Stats NZ	

Exclusions from the scope of the 2023 Census Coverage project

As the project has evolved from 2018, the following aspects are not in scope for 2023. The decision to remove them from scope was made with agreement from the 2023 Census programme.

Table 4

2023 Census Coverage project exclusions	
Number	Description
1	Coverage measurement of administrative data population
2	Dwelling coverage measurement
Source: Stats NZ	

Design assurance

This is the process and mechanisms by which the census coverage design and implementation will be assessed in its ability to meet its objectives. Alignment has been achieved with the three-pronged approach proposed by the census programme:

- assurance within the census coverage project and wider census programme
- assurance outside the program, but within Stats NZ
- independent assurance.

Assurance within the census coverage project and wider census programme

Design assurance will be coordinated and monitored by the project manager with oversight from the PES governance board. It will ensure the high-level design is being adhered to throughout the design, implementation, and operational phases of the project. A coordinated internal review and approval process is in place for key documents and decisions.

Assurance outside the programme, within Stats NZ

Design and planning will be informed and reviewed by Stats NZ subject matter experts who sit outside the census and Census Coverage programmes. For example, the Enterprise Portfolio Office (EPO) will manage some business operations, and processing specialists from other surveys will review the PES response processing system. Key design documents will be reviewed by the Technical Advisory Group (TAG), a committee of technical specialists from across the organisation.

Independent assurance

This final layer consists of independent expertise that will provide assurance to the 2023 Census Coverage design. The Quality Assurance Panel (QAP) includes international coverage experts, external expert data users, and internal specialists, all selected for the purpose of providing vigorous methodological input and review to both the standard and adaptive designs.

Privacy and security

The census coverage project is collecting data via the PES and is using it to produce statistics to benefit Aotearoa New Zealand. There is legislation in place that sets the standards for how to collect, store, and use this information. This section explains how the privacy and security of the data we collect will be ensured.

Privacy

Privacy refers to someone's ability to keep their information to themselves. Stats NZ is bound by the Data and Statistics Act 2022, which means that individuals are not identified in published statistics. Information collected in the survey will be safely linked to the Stats NZ Integrated Data Infrastructure (IDI) to enhance the quality. The IDI holds microdata about people and households, collated from other Stats NZ surveys, other government agencies, and non-government organisations.

The consideration for the use of collected and administrative data is detailed in the 2023 Post-enumeration Survey: Privacy impact assessment. It is a key part of the focus on privacy for understanding and building social licence for PES activities. The privacy impact assessment prescribes controls necessary to minimise privacy risk and to ensure we are operating within accepted social licence.

Security

Data security refers to the process by which data is protected from unauthorised access or modification. The key principles of data security for the census and census coverage projects are:

- confidentiality – access to information must be limited, and no individuals' characteristics may be deducible from published data
- integrity – the data must be consistent, accurate, and trustworthy throughout its lifecycle
- availability – the data must be accessible when it is needed to ensure the operation is able to run effectively and the statistical outputs are timely.

Confidentiality

Data confidentiality ensures how information relating to an individual is not divulged or able to be deduced from a publication. In general, Stats NZ achieves this by:

- aggregating data such that an individual or household record cannot be identified or inferred
- adding random noise to data points, such as through random rounding
- suppressing certain data points in aggregate data
- de-identifying personal data and removing correlation identifiers
- controlling access to microdata for specific, approved statistical and research purposes.

Business services

Like census, PES design-led approach includes identifying business services early in the process and having a clear understanding of the following:

- the services required by the project
- which services each team is expected to deliver and when
- the dependencies between each of the services
- with whom the responsibility for delivery sits.

The business services that have been identified for the 2023 Census Coverage project are listed below.

Population Insights

- Unit ultimately responsible for driving the delivery of the end-to-end solution for the 2023 Census Coverage project.
- Will be present at every stage of the project.
- Identifies the requirements the design needs to achieve.
- Provides the targets, sample areas, and questionnaire to Collection Operations for the field operations.
- Responsible for aspects of data collection that are unique to the PES, such as collateral.

Methods and Design

- Responsible for the design of the methodological elements of the project.
- The methodology is detailed in the standard design, linking design, and adaptive design.
- Generates the questionnaire.
- Provides input on questionnaire design.

Collection Operations

- Responsible for data collection operations.
- Creates training materials and delivers training to the field staff.
- Responsible for the recruitment, onboarding, training, and equipping of field staff.

Te Tohu Rautaki Angitu Māori

- Advice and guidance for incorporating te ao Māori perspective.
- Supports engagement activities with Māori.

2023 Census

- Responsible for providing the CURF to the Census Coverage project.
- Responsible for keeping Census Coverage project updated on any changes to the 2023 Census design that may impact upon the project's ability to accurately measure census coverage, and to consider the needs of the project when making these changes.

Data Services

- Responsible for providing administrative data to the 2023 Census Coverage project to enable accurate coverage measurement of a combined census file.
- Provides guidance on safe usage of administrative data.

Statistical Infrastructure

- Supports the processing environments.
- Provides guidance on end user computing.

Technology Services

- Provides setup and configuration of off-the-shelf tools and services required throughout the project, such as:
    - field management tool
    - field workload allocation tool
    - questionnaire delivery tool
    - processing environment.
- Supports security (internal and external) across the project.
- Responsible for the admin data processing environment.

End-to-end design

The end-to-end design has been partitioned into phases, broadly aligning with the Generic Statistical Business Process Model (GSBPM – refer to Appendix for further details), that also aligns with the language used by census. Each phase culminates in deliverables which feed into the next phase. The phases are grouped together to enable the coverage project to identify clear points of demarcation in time.

Principles

Table 5 provides a list of the key information principles and how they relate to the end-to-end design. All principles apply but only the key ones have been highlighted here. These principles align to the New Zealand data and information management principles.

Table 5

Information principles		
Principle	Rationale	Implications
Information is managed as an asset throughout its lifecycle	Data is a system asset for all Aotearoa New Zealand.	Data should meet corporate policies for management.
Information is accessible and open by design	Data should be accessible, so users know what is available and how to access it.	As much as possible, metadata should be made available to internal and external users.
Data is described to enable reuse	• Data should have enough metadata to support its management and use. • Data needs to be understood so it can be used appropriately. • It should be easy to determine what is the same and what is different.	• Metadata should be described at the lowest level possible, using common models and standards to maximise reuse and interoperability. • Metadata should be machine readable to support process automation. • As much quality information as possible should be published to ensure the data limitations are well understood. • Classifications used to produce published data (even non-standard) should be published so users can understand the data.
Use an authoritative source	Having an authoritative source makes it easier to find information and have confidence in its integrity.	• Metadata should use corporate repositories and enterprise services where possible • Data production pipelines should provide a single source of truth from which products and services are generated
Use agreed models and standards	Key information should be described using common models and standards, agreed by the	• New projects or components should align to agreed models and standards before designing new ones • Projects or components should align to the steady-state framework

	organisation, to facilitate sharing and reuse.	
Source: Stats NZ		

Phase 1: Information need and design

These phases have been grouped together in acknowledgement that their functions can happen concurrently across different aspects of the project, and that there might be some iteration between need and design as both the priorities, and technical demands from different approaches, are understood.

Independent information need

This phase entails consulting with customers and stakeholders regarding what they need or expect from outputs of the project. There are two dimensions to this:

- **Information need**, which identifies the content of the statistical product required. In the coverage context, this relates to the required output demographic categories, but also dimensions of coverage measurement, such as net and gross coverage measures.
- **Output data quality**, which identifies the quality standards that the data needs to meet to be useful for the customer.

Engagement with customers is staged over several iterations as common ground is sought between user need and what the project can commit to delivering with high confidence.

Design

This phase entails determination of what is needed to build to deliver the project and its outputs. While there are several aspects of the design that will closely resemble that of the 2018 Census Coverage project (such as the estimation approach), part of the design work will address aspects of implementation that are updated for this cycle.

The major pieces of design work prioritised for 2023 PES were:

- a questionnaire re-design to optimise data quality and buildability in an electronic solution
- a robust dwelling frame to support person estimation using a dual-system estimation approach
- an adaptive design
- robust processing infrastructure.

Phase 2: Build and testing

This phase entails the building and testing of methodology and associated solutions required in advance of the live field operation. As with phase 1, the build and testing phases have been grouped together to highlight the cyclic nature of development.

Build

While some technological services are standard corporate tools and platforms, many methodology solutions will be developed in-house by the project team and delivery partners. There are several components the project will prepare in this phase, including:

- methods for producing dwelling and interview samples
- electronic questionnaire solution
- data collection solution and hardware (including Salesforce, Blaise 5.0, EPIC Response Store and integration)
- field progress reporting solution
- post-collection data processing
- record linking solution
- imputation solution
- estimation solution.

Testing

Included in this phase is the programme of testing to provide confidence in the solutions and quality assurance.

Software developed for each component will be version controlled and packaged where applicable, with unit tests developed for each module. Appropriate code promotion practises will be followed. For practical components, 2023 Census' field testing activities will be leveraged, and small tests run within the project for smaller components.

All components from data collection to data ingest and processing will be subject to an end-to-end test prior to the field operation. As part of the quality assurance, ongoing analyses of results will be run, along with comparison against the expected results for each stage.

Phase 3: Field operation

This phase describes the live process of data collection for the 2023 PES, and the support systems and quality assurance processes required for it to run effectively. It closely resembles the approach used in the 2018 cycle.

The 2023 PES will consist of a face-to-face interview of approximately 16,500 households. The interview consists of two questionnaires:

- Household questionnaire (HQ): questions about the address, the nature of the dwelling, and an overview of the people currently occupying it.
- Personal questionnaire (PQ): one PQ will be filled in for each person identified in the dwelling in the HQ. Questions in the PQ are designed to determine whether the respondent is in-scope for the PES, and if so, gather information on their interaction with census, their demographic characteristics, and addresses at which their records may be found.

The total PES interview consists of approximately 40 questions and will take 8–10 minutes per respondent.

The PES field operation is due to begin on 1 May 2023 (approximately eight weeks after the census on 7 March) and will continue for eight weeks. The households selected for the PES interview are designed to be a representative sample from across the country. Approximately 200 field staff will be required to complete the field operation; 60 percent of the field staff will be hired on temporary contracts specifically for the duration of the operation, and supported by existing, more experienced, permanent Stats NZ data collection specialists (40 percent, part-time).

Phase 4: Processing, linking, derivation, imputation, and estimation

These phases have been grouped together to reflect that the processing and estimation may be iterative. The estimation design needs to be adaptive so that plans for a range of data patterns can be made. This manages the delivery risk.

Processing

Processing refers to the system by which data is taken from the field collection storage and prepared for later stages. There are several components of this, including:

- aligning the data metadata to Stats NZ conventions
- automatic and clerical (human) categorisation of non-categorical variables (such as ethnicity)
- filtering out facetious responses.

Linking

The linking design document for the 2023 Census Coverage project explains the plan for linking PES responses to census records and IDI records in detail. This is an important stage; perfect linking between PES and census records is a key assumption of dual-system estimation and without it, no

estimation can take place. As part of this piece of work, contractors will be hired to carry out clerical linking in Q1 2024.

Imputation

For coverage rates of different demographic characteristics (such as age, gender, ethnicity) to be modelled, there is a set of variables to which there must have complete responses. Once linking has been completed, incomplete PES records are able to be filled in, using census data, IDI sources, or statistical strategies – a process referred to as imputation. For the 2023 imputation design, the variables are:

- age/date of birth
- gender
- ethnicity
- Māori descent
- country of birth
- geographic location at the territorial authority and local board (TALB) level.

The imputation design for the 2023 project is explained in more detail in 2023 Post-enumeration Survey: Standard design for coverage estimation.

Derivation

Derivation is the process of categorising individuals according to their classifications of interest for coverage measurement (for example, under-coverage, over-coverage, and key demographic information). There are a variety of derivation processes, all of which use strict rule-based assessments of the individuals.

One set of derivations establishes whether an individual is in scope to be included in the coverage analysis, and if so, the type of coverage they can be classified as (for example, under-coverage, correct inclusion).

A second type of derivations classifies the record as being a part of a demographic group for an overall coverage adjustment of that group (for example, age group, gender, geographic location, ethnicity). For a full explanation of the rules applied during derivation, refer to the standard design document.

Estimation

Estimation refers to the process of generating the following outputs for all population groups of interest:

- PES estimated population
- estimated gross overcount in the census dataset
- estimated gross undercount in the census dataset
- estimated net undercount achieved by census.

These outputs are achieved through computational modelling of the data and the census data. The estimation design is described in more detail in the standard design document.

Phase 5: Outputs and dissemination

The census outputs produced by Stats NZ are high priority, and so clear, prompt, and accurate dissemination of the methodology and results of the Census Coverage project is vital.

Output design priorities

The Census Coverage project outputs have several key priorities. The first is to raise awareness of the PES and its outcomes, to bolster public confidence in such projects.

Another priority is to work with our Te Tiriti partners, and engage with customers, stakeholders and internal collaborators to promote participation in the design throughout this and future cycles.

Finally, the PES concept is not unique to New Zealand; most countries that perform a survey census of their population also execute a post-censal coverage survey. It is therefore important to collaborate with other national statistics organisations (NSOs) to keep up to date with, and contribute to, international best practice.

The PES outputs have a varied audience, ranging from experts to the general public. The outputs therefore need to be written clearly, concisely, and targeted to the specific audience for each piece. To maximise confidence and minimise confusion, consistency between outputs will also be an important factor. In order to best serve the customers and stakeholders, it is essential that they are all timely, factual, and accurate.

Planned outputs

Table 6 lists the documents and data planned for release during the 2023 Census Coverage cycle.

Table 6

Planned outputs		
Release title	**Description**	**Estimated publication date**
Field test report: key findings	High-level key findings and learnings from the 2022 PES field test.	Published on 11 October 2022
Privacy impact assessment	The report describing controls necessary to minimise privacy risk and to ensure PES are operating within social licence.	March 2023
High level design	Describes the key design elements needed to complete the 2023 PES project and the strategy for implementing them.	March 2023
Standard design	The planned design for coverage estimation of the 2023 Census using the 2023 PES and supporting information.	March 2023

Linking design	Overview of the research and process undertaken to design robust methodology to link PES responses to their corresponding census records.	March 2023
Adaptive design	Description of the statistical risks to the project, how they can be identified and monitored, and the planned response to each if they materialise.	Late 2023
News story media releases	Non-technical releases of key facts and findings.	December 2024
Coverage report	Detailed explanation of the key features of the 2023 project, including detail on scope and methodology; detailed presentation of results. As part of delivering for and with Māori, iwi, and hapū, this will include an additional targeted explanation of results of specific interest to Māori.	December 2024
Data dictionary	A document defining key concepts and terminology used in other releases related to the coverage report	December 2024
Information release: Post-enumeration Survey	Brief, high-level overview of the key features of the 2023 project.	December 2024
Source: Stats NZ		

Glossary

CAR – census adjustment ratio. The stratified rebasing weights for population groups, estimated using PES data. The CAR is applied to the census population as a component of calculating the ERP.

clerical linking – process of manually reviewing records not able to be linked by automatic processes, and searching for missed links.

combined census – a census in which some information on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes, but where other information that is not available from such sources is collected directly from individual persons and households by means of full or partial field enumeration or from other sample surveys. The 2018 Census used a combined census model after the survey period to address data quality issues from low response, whereas the 2023 Census has been designed to use a combined census model from the outset.

computational modelling – using computers to simulate and study complex systems by applying mathematics, physics, and computer science.

confidence interval – probability that a population parameter in PES will fall between a set of values for a certain proportion of times. Used in the PES prior to 2018, with a 95 percent confidence level.

correlation identifier – indicates whether a piece of data is about a specific real-world object.

credible interval – interval within which an unobserved parameter value falls with a particular probability. Commonly used in Bayesian statistics.

design-led – co-ordinated approach to design with practical processes and assurance mechanisms.

desktop canvassing – enumeration activity whereby the cleaning and updating of an address list is through desk-based investigation instead of field activity. Desktop staff are instructed to use a range of desk-based resources (such as satellite imagery, rates databases, internet searching) to classify addresses and record relevant information. Can be used in conjunction with field enumeration.

dwelling frame – list of in-scope addresses from which the PES sample will be drawn.

ERP – estimated resident population of New Zealand on 30 June the year of the census. The ERP is produced by Stats NZ using the census count, the census adjustment ratio, border crossing data, and birth and death registrations.

facetious response – providing a false answer to a question and which is intended to be humorous.

field enumeration – physical address-checking exercise completed prior to sample selection, to determine the number of eligible dwellings contained within a geographic area of interest, and their address. Can be undertaken via different approaches including scratch enumeration, where a complete list of dwellings is created from scratch without a starting address list, and list-based enumeration, where updates/corrections are made to a starting (internal or external to Stats NZ) address list.

instance – new context in an existing computer system where values are stored separately, but the data structure remains common.

microdata – individual response data in collections, including census, administrative, and survey data.

PES – 2023 Post-enumeration Survey. A predominantly face-to-face interview and questionnaire carried out after the 2023 Census, the data from which provides the official coverage and response measures of the census.

post-censal – something that happens after the census and which uses census data as an input, for example post-censal survey, post-censal estimates.

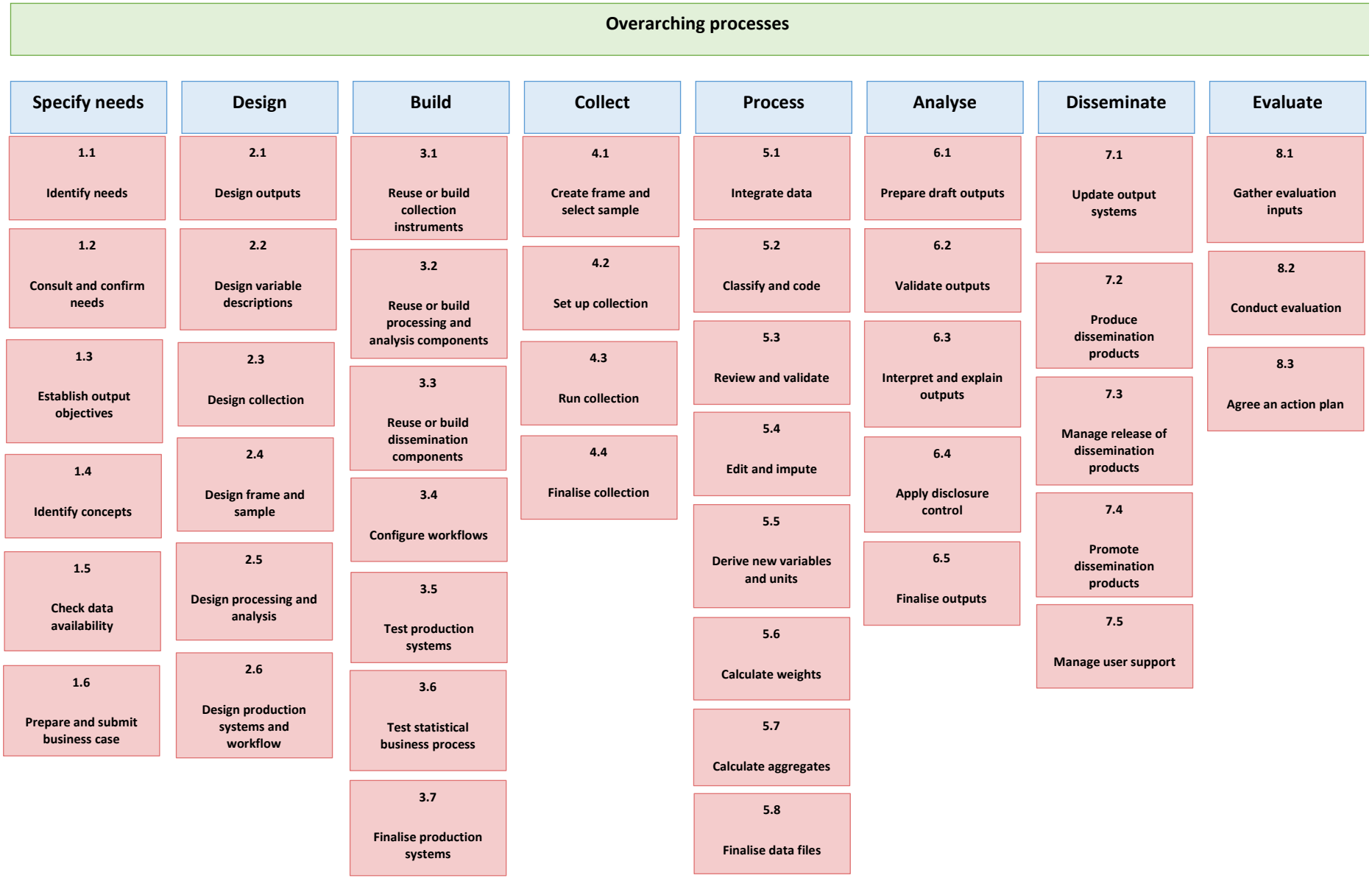
random noise – introducing uncertainty in output data to protect confidentiality, such as by random rounding.

social licence – having social licence to operate is the ability of an organisation to carry out its business because of the confidence society has that it will behave legitimately, with accountability and in a socially and ethically responsible way.

strata – defined groups based on shared characteristics, for example socio-economic characteristics. Used in stratified sampling technique.

triangular linking – the process linking three datasets together by using three sets of pairwise links. The pairwise links can be used to create closed and broken triangles. Broken triangles can help detect linkage error in one or more of the pairwise links.

Appendix: Phases and sub-processes of the generic statistical business process model



Source: UNECE https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2019/mtg2/MWW2019_GSBPM_Munoz_Presentation.pdf