

Modernising our approach to the 2028 Census: an admin data first approach

Consultation deadline 19 June 2024

‘Statisticians play a critical role in supporting evidence-based policy and measuring civil, economic, political and social rights’ (Commissioner for Human Rights, 2014)

Dr Gerard JB Sonder, Public Health Physician, epidemiologist PhD

Dr Debbie Ryan MNZM, MPM, Principal Pacific Perspectives Limited

June 2024

Contents

1. Overview	2
2. Part 1: Historical undercount of Pacific populations	3
3. Census and surveys- what is the problem?	4
a) <i>Estimating Census undercount, Census coverage and the Resident Population</i>	4
b) <i>Census 2023 sampling frame, inaccurate address data and manual processes</i>	5
c) <i>How the suggested transition would influence Household Surveys</i>	6
4. An admin data first approach as suggested by StatsNZ is not the best way to count Pacific populations in New Zealand	6
5. Part 1 Summary	8
6. Part 2: Opportunities to create a new statistical population ‘backbone’ using IT Solutions .	9
a) <i>Examples of equity implications for public health programmes of inaccurate and incomplete address registers</i>	10
7. Improving data quality to create a new statistical population ‘backbone’ using IT solutions	11
8. Suggested system’s advantages.....	11

Modernising our approach to the 2028 Census: an admin data first approach

Consultation deadline 19 June 2024

Dr Gerard JB Sonder, Public Health Physician, epidemiologist PhD

Dr Debbie Ryan MNZM, MPM, Principal Pacific Perspectives Limited

June 2024

‘Statisticians play a critical role in supporting evidence-based policy and measuring civil, economic, political and social rights’ (Commissioner for Human Rights, 2014)

1. Overview

- 1.1. Pacific Perspectives is an independent provider of research, evaluation and workforce programmes. Our focus is on strengthening the evidence for policy and services delivery that will lead to improved health and wellbeing for Pacific communities in Aotearoa/New Zealand.
- 1.2. We commend Stats NZ on the consultation that has been undertaken with Pacific communities about the proposed changes to the population census. Whereas the ‘overall’ quality of official population data can be considered as ‘good’, for Pacific and other ethnic minorities, the devil is in the detail.
- 1.3. Our submission is presented in two parts. Part 1 includes detailed commentary about the longstanding problems with undercounting of Pacific people in official censuses, and the failed 2018 census and how this influences the quality of all other statistics based on full population enumerations. “The Census provides a skeleton; admin data is a bit like fossils – fitting fossils to a skeleton is very hard...” (Len Cook, NZ Herald, 9 June 2024). We have included examples of the problems with the statistics used to monitor health and social outcomes, and access to public services (see page 3: *Historical Undercount of Pacific Populations*).
- 1.4. Inaccurate population statistics have had significant consequences for funding, delivery of, and access to public services for the NZ Pacific population. However, the contribution of these errors to persisting inequitable access to governmental services and therefore outcomes is rarely discussed or understood (see page 4: *Census and Surveys- What is the problem?*).
- 1.5. We argue that in health for example, the consistent life expectancy gap estimated at 6 years lower for Pacific compared with the majority population, that has persisted unchanged for nearly 25 years, is in part a consequence of inaccurate population statistics. The Government’s focus on social investment cannot be equitably realised without attention to these fundamental and long-standing issues.
- 1.6. We are deeply concerned that the admin-first approach - as described in the discussion document - will contribute to a further deterioration in access to public services and therefore result in continuing inequities in outcomes for our Pacific population. We agree with the findings of the Statutory Review (April 2024)¹ that: *“Moving from a traditional census to administrative data to enumerate the population is very complicated. The decision must be made at a point that allows sufficient time for effective contingency plans, to be put in place and to be successfully deployed*

¹ <https://www.stats.govt.nz/reports/report-of-the-statutory-review-of-new-zealands-2023-census>

should an administrative data first census not be feasible for 2028. It will be extremely difficult and expensive for StatsNZ or another statistical agency to concurrently plan a traditional/hybrid census and an administrative one". The review states that objective evaluation of the readiness to proceed is critical.

- 1.7. We discuss the conditions that Cabinet in 2015 agreed needed to be met in order, to safely move to an administrative census and we argue that these conditions have not been met (see page 6: *An admin data first approach as suggested by StatsNZ is not the best way to count Pacific populations in New Zealand*). An admin data first approach as suggested by StatsNZ is not the best way to count Pacific populations in New Zealand and that there is ample evidence that the data quality in our admin systems is not at the required level. We strongly disagree with the statement in the Discussion Document that *"We know that the best way to count the population is through admin data"*.
- 1.8. Part 2 of our submission discusses a potential alternative.

2. Part 1: Historical undercount of Pacific populations

- 2.1. The inaccuracies in Pacific population estimates were highlighted after the failed transition to an online Census 2018, that led to unprecedented low response rates, particularly for the Pacific population. Admin data of unknown quality had to be used to plug the gaps in population enumerations. Even after admin data was used to impute population counts, Pacific people were still the most undercounted population.
- 2.2. An example of how this impacts on health service delivery became apparent when official population statistics based on Census 2018 results were used to calculate COVID-19 vaccination coverage in 2021. For Pacific people, vaccination coverage was systematically overestimated by the Ministry of Health (MOH), sometimes to greater than 100%, whereas for other ethnic groups, vaccination coverage was underestimated². To overcome this problem, the MOH used an alternative denominator - Health Service User (HSU) data. This was an arbitrary decision that introduced new biases.
- 2.3. Furthermore, although it has become known that the same flawed population statistics are also used by MOH for calculating many other health statistics, none of these statistics have been reviewed or revised by MOH – perpetuating Institutional bias. We have highlighted this in the following publication³. The publication describes how these errors lead to inaccurate policy advice to Cabinet about Pacific access to care and the extent of delivery of public health measures.
- 2.4. Another example that shows the systematic undercount of Pacific people occurred in 2021. When New Zealand's international borders were closed, and COVID-19

² Anglemeyer A, Grey C, Tukuitonga C, Sporle A, Sonder GJB. Assessment of Ethnic Inequities and Subpopulation Estimates in COVID-19 Vaccination in New Zealand. *JAMA Netw Open*. 2022 Jun 1;5(6):e2217653. doi: 10.1001/jamanetworkopen.2022.17653. Erratum in: *JAMA Netw Open*. 2022 Jul 1;5(7):e2226752. doi: 10.1001/jamanetworkopen.2022.26752. PMID: 35727584; PMCID: PMC9214581.

³ Sonder GJB, Grey C, Ryan D, Cumming J, Sporle A, Hill PC. Selective under-representation of Pacific peoples in population estimates for health indicator measurements in Aotearoa New Zealand misinforms policy making. *BMC Public Health*. 2024 Feb 22;24(1):564. doi: 10.1186/s12889-024-17984-2. PMID: 38388865; PMCID: PMC10882897.

vaccinations were freely available for the entire population, an extra (approximately) 45,000 Pacific people⁴ (11% of the estimated Pacific population) were identified through the vaccination registers that had previously not been identified or counted through Census 2018 nor through the imputed admin enumeration. There may be several reasons why these people were not known to the system, but the collected data were not of sufficient quality to explain why such large number of uncounted Pacific people were identified.

3. Census and surveys- what is the problem?

- 3.1. The Census of Population and Dwellings in combination with National Household Surveys are the two main sources upon which government policies are made.
- 3.2. Because official census population statistics are also used as the sampling frame for Household Surveys, as well as a benchmark to weight Household Survey results, inaccurate Pacific population census enumerations and population statistics directly contribute to inaccurate conclusions for Pacific people based on these surveys.
- 3.3. Official census-based population estimates are also used to estimate the response rates and coverage rates for the census itself. Consistent underestimates of the Pacific population have led to consistent over-estimation of Pacific peoples' participation in census, and in Pacific census coverage estimates.

a) Estimating Census undercount, Census coverage and the Resident Population

- 3.4. The Usual Resident Population (URP) is the population counted on census night through standardised forms. A Post Enumeration Survey (PES) survey follows each census to estimate the undercount or overcount. Based on URP and PES results, Estimated Resident Population (ERP) is derived. ERP is considered the official population enumeration statistic for NZ. ERP is updated every 3 months to correct for natural population changes using births, deaths and migration data derived from administrative data, until the next census enumeration data becomes available.
- 3.5. The enumeration imputations to enhance 2018 population statistics had unanticipated side effects on the methodology and quality of the PES. PES is based on a Dual System Estimation (DSE) methodology (which is a capture-recapture method) and depends on two independent sampling frames with resident and dwelling addresses. The data from the two independent data sets needs to be reliably linked for comparison. Before 2018, PES sampling frames were based on existing Household Sampling Frames (2006 HLFS and 2013 GSS), but because of the data imputations in 2018, the sampling frames were no longer independent.
- 3.6. A new sampling frame and methodology for PES had to be developed after Census 2018. It proved to be unreliable: StatsNZ warns in their internal evaluation that "...the well-known difficulties with the 2018 Census had a major impact on the measurement of census coverage in 2018". Although it has not been analysed, if the failed PES had different impacts on different ethnic groups, it is likely that it most impacted the Pacific population, given that this population is a highly mobile population and linkage errors in bringing administration data together are more prevalent, further contributing to the inaccuracy of ERP for Pacific people.

⁴ <https://www.stats.govt.nz/reports/review-of-health-service-user-population-methodology/>

- 3.7. The Census 2018 issues still influence accuracy of population estimates in 2024. Preliminary Census 2023 data were published in May 2024. The preliminary census coverage for the Pacific population is reported as 98.4%. Because PES results for 2023 are not yet available, this figure is based on ERP as the denominator. The current ERP is still extrapolated from Census 2018. All biases and population underestimates for the Pacific populations described above still apply to this ERP. It is obvious and evident that Pacific people are systematically substantially undercounted. The 98.4% coverage for Pacific people is likely to be a substantial overestimate, and disproportionately so, when compared to all other ethnic groups. The 98.4% coverage gives a false sense of accuracy for Pacific population estimates.
- 3.8. For Census 2023, StatsNZ developed yet another very complicated PES 2023 methodology, which was published in 4 different documents totalling 293 pages.
- 3.9. There have been several new problems with Census 2023 and PES 2023 (see below), and official population statistics based on Census 2023 will unlikely deliver a more accurate population estimate for Pacific people, even once the PES results are made available on 9 December 2024.

b) Census 2023 sampling frame, inaccurate address data and manual processes

- 3.10. Methodologically, the census is a Household Survey like all other Household Surveys, with the difference that the Census includes the entire New Zealand population whereas Household Surveys are intended to include a random representative sample. Every survey, including the census, requires a sampling frame to invite people to participate. For the Census, an accurate, complete, up-to-date sampling frame with accurate address data is needed to include the entire population.
- 3.11. Because there is no national database with correct up-to-date address data for the entire population or of all dwellings, before each census, StatsNZ must create this sample frame, which is called the Statistical Location Register (SLR).
- 3.12. The methodology for creating the SLR 2023 was published long after census night, on 12 March 2024. The document describes in detail how a mix of independent and commercial address databases, none of which were considered accurate or up to date, were used to compare and check against each other. A team of 20 “canvassers” was needed to manually verify and correct all supposedly incorrect addresses. This process description undermines trust in the quality of the underlying data. The described method is not reproducible, which means it is neither transparent nor controllable. It is again likely that the highly mobile and deprivation-high Pacific population is disproportionately not captured in the sampling frame, which is likely to be the first step at the beginning of many inaccurate steps affecting Pacific population estimates most and resulting in systemic underrepresentation of Pacific people in NZ population statistics.
- 3.13. The Statutory Review 2023 comments on this process as: “...*Inaccuracies in the address-frame created inefficiencies and frustrations in field operations...*”, but more importantly: “...*It is noted an accurate address-frame will be more critical in any shift to an administrative census...*”
- 3.14. Incorrect address data is not only one of the most essential causes of under-representation in the Census, as it is directly related to the accuracy of the Census ‘backbone’. It is also particularly relevant at all times when population statistics are reported at the sub-national level.

3.15. Inconsistent address data is not the only data quality issue that is directly related to accuracy of population enumeration. Consistency of all data that is used in the complicated data linking processes (name, date of birth, gender, address) has an influence on enumeration accuracy. This is discussed in Part 2 of this submission.

c) How the suggested transition would influence Household Surveys.

- 3.16. The quality of Household Surveys depends largely on census enumerations. The Household Survey Sampling Frame is created every 5 years based on the most recent census enumerations. The sample size is based on the census enumerations, the actual samples used are stratified based on data collected in census.
- 3.17. StatsNZ reported to Cabinet on 21 October 2015 that StatsNZ would work actively with agencies exploring new data-sources including scoping the redevelopment of the StatsNZ Household Survey Programme.
- 3.18. The currently publicly available information on methodology for household surveys do not imply any changes in methodology have occurred since the publication of the Sampling Frame 2009 by Clark et al.
- 3.19. The long-standing systematic undercount of Pacific people in the census has never been addressed in the household sampling frame. Likely, the current Household Surveys are least accurate of all ethnic group for Pacific people, given the systematic inaccuracies of population enumerations for Pacific people.
- 3.20. When StatsNZ reported the scoping of redevelopment of the Household Surveys, it was not then known that this would be followed by two failed censuses, impacting the reliability of the census backbone, which is also the sampling frame for household surveys.
- 3.21. Following a census transition, as proposed, the quality and accuracy of Household Survey results will most likely further deteriorate for Pacific people.
- 3.22. Before any decision on census transition is being made, this problem needs to be reviewed and addressed.

4. An admin data first approach as suggested by StatsNZ is not the best way to count Pacific populations in New Zealand

- 4.1. The decision to work towards a modernised approach to the Census was made by Cabinet in 2015.
- 4.2. Although it is possible to enumerate populations through administrative data, this cannot be done reliably whilst the conditions reported to Cabinet in 2015 are not met.
- 4.3. One important condition was to run an administrative census parallel to a full traditional census twice (2018 and 2023) to compare administrative data quality against the backbone of the traditional census.
- 4.4. Time has caught up with this plan: a disastrously failed Census 2018 created the need to enhance our population statistics with methods not previously tested, leaving us without a 'backbone' of known quality to test the admin census data quality against. The same applies to Census 2023.
- 4.5. The need to improve data quality as one of the essential conditions to safely move to an administrative Census was first communicated by StatsNZ to Cabinet on 27 October 2015. The paper acknowledges that ethnic breakdowns of populations and communities are a fundamental aspect of census, for example for population-based

health funding. The paper also acknowledges that collection of ethnicity across different admin sources lacks consistency and that this is *particularly problematic when people have multiple ethnicities*. This means that it particularly concerns the Pacific population. At a recent consultation meeting, StatsNZ admitted that this problem has not been solved and also that they don't yet know how to solve this problem.

- 4.6. A Census Transformation Progress Update paper to Cabinet dated 7 November 2016 reported that Stats NZ had made good progress in improving data quality:

Table 1: Extent to which agencies currently meet the 2015 directives

	Name	Address	Date of Birth	Sex	Ethnicity	Iwi	Māori descent	Data Documentation	Address Validation	Time Referencing											
Inland Revenue	Green	Green	Green	Green	Green	Green	Green	Red	Red	Red	<table border="1"> <thead> <tr> <th colspan="2">Key</th> </tr> </thead> <tbody> <tr> <td>Green</td> <td>Fully meets</td> </tr> <tr> <td>Orange</td> <td>Partially meets</td> </tr> <tr> <td>Red</td> <td>Does not meet</td> </tr> <tr> <td>Grey</td> <td>Not collected</td> </tr> </tbody> </table>	Key		Green	Fully meets	Orange	Partially meets	Red	Does not meet	Grey	Not collected
Key																					
Green	Fully meets																				
Orange	Partially meets																				
Red	Does not meet																				
Grey	Not collected																				
Ministry of Health	Green	Green	Green	Green	Green	Green	Green	Red	Red	Red											
Ministry of Education	Green	Green	Green	Green	Green	Green	Green	Red	Red	Red											
Ministry of Social Development	Green	Green	Green	Green	Green	Green	Green	Red	Red	Red											
Ministry of Business, Innovation and Employment	Green	Green	Green	Green	Green	Green	Green	Red	Red	Red											
Department of Internal Affairs	Green	Green	Green	Green	Green	Green	Green	Red	Red	Red											

- 4.7. This progress report may concern the overall population. Several issues concerning data quality however, even for the fields that are green in Table 1, still exist in 2024. Statistical Location Register 2023 showed that there is no uniform national address database that can serve as the sample frame for the Census and of which the Statutory Review 2023 writes that will be “even more critical in any shift to an administrative census”.
- 4.8. Furthermore, there are well-known inconsistencies in quality of data collected on ethnicity: after Census 2018, when Māori ethnicity data was compared between data registered in NHI (health) and data collected on individual Census 2018 forms (which is the gold standard), it appeared 16% of Māori were missing as such in health data⁵. When the same comparison was made for Pacific people, there was a 20% discrepancy between the datasets. Incorrect registration of ethnicity data in an admin census would mean that people are not counted as part of that population: if NHI data would have been used for an administrative enumeration census, Pacific people would have been undercounted by another 20%.
- 4.9. For the health system and the health of individuals, this means that 20% of these target populations are not included in all the policies and monitoring designed to improve equitable access and outcome of care. This leads to systemic racism and contributes to the continuing gap in life expectancy.

⁵ Harris R, Paine SJ, Atkinson J, Robson B, King PT, Randle J, Mizdrak A, McLeod M. We still don't count: the under-counting and under-representation of Māori in health and disability sector data. N Z Med J. 2022 Dec 16;135(1567):54-78. PMID: 36521086.

5. Part 1 Summary

- 5.1. We are concerned that the admin-first approach will contribute to a further deterioration in access to public services and therefore result in continuing inequities in outcomes for our Pacific population.
- 5.2. Pacific populations in Aotearoa New Zealand have been systematically undercounted in official statistics for decades.
- 5.3. Inaccurate population statistics have had significant consequences for funding, delivery of, and access to, public services for the NZ Pacific population and contribute thereby to the long-standing inequities in health outcomes.
- 5.4. The Census transition to make better use of data was planned, and communicated to Cabinet in 2015, including the conditions that had to be met to make the transition possible.
- 5.5. Some of these conditions have not been met, particularly the comparison of admin data against two more full enumeration censuses and the standardisation of the administrative data quality essential for accurate admin census enumeration. The failure of two consecutive censuses was not anticipated.
- 5.6. Census failures have not only led to inaccurate enumerations, the imputation of admin data from 2018 onward also impacted on the Post Enumeration Survey. Based on this, StatsNZ concluded that “...*the well-known difficulties with the 2018 Census had major impact on the measurement of census coverage in 2018*”. Lack of insight in census coverage means lack of insight in the actual size of the population.
- 5.7. We believe that the roadmap towards an administrative census, as communicated to Cabinet in 2015, has been compromised. This cannot just be ignored but requires a robust alternative. We have not seen any alternative roadmaps that address these issues.
- 5.8. We discuss in Part 2 our recommendation to consistently separate ‘Census’ into two parts: one part concerns the population enumerations, and the other part the collection of detailed population data that is now partly collected on census forms and partly through household surveys.
- 5.9. A full accurate population enumeration is essential, not only for accurate population statistics, but also for the context of all other statistics, including population surveys.

6. Part 2: Opportunities to create a new statistical population ‘backbone’ using IT Solutions

- 6.1. A full enumeration census provides the ‘backbone’ or ‘skeleton’ for all other statistics. It is the basis for the most accurate population enumeration data but only if response rates are very high. A full enumeration census also serves as the sampling frame for all other (household) surveys.
- 6.2. Pacific populations have been systematically undercounted for decades, even when overall census response rates were much higher than in the last two censuses.
- 6.3. Because of the very low response rates in two consecutive censuses (2018 and 2023), the accuracy of the overall population ‘skeleton’ has declined. This decline is most pronounced for Pacific and Māori populations.
- 6.4. In 2015 it was anticipated there would have been two full enumeration censuses against which the newly built administrative population census could be tested. However, from 2018 onwards, administrative enumeration data has had to be imputed in the census to enhance population estimates. The low census response rate in 2018 and the unanticipated need to impute administrative data in census enumerations, required changes in the Post Enumeration Survey methodology as well, resulting in a census ‘backbone’ of unknown accuracy since 2018. As StatsNZ noted: “...the well-known difficulties with the 2018 Census had major impact on the measurement of census coverage in 2018...” that is inaccuracy of the estimated population size.
- 6.5. This has affected the accuracy of Pacific population estimates most. Accurate insight in the Pacific population size has long been problematic but has further deteriorated since 2018.
- 6.6. There hasn’t been a reliable ‘gold-standard’ for population estimates, or an accurate ‘statistical backbone’, since 2018.
- 6.7. We do not believe another traditional full-enumeration census 2028 would be more successful or feasible, given the experiences with the last two censuses.
- 6.8. As the roadmap towards an administrative census reported to Cabinet in 2015, has failed, an alternative strategy is needed to obtain more accurate population estimates.
- 6.9. Given their different purposes, it would be beneficial to consistently - - separate two components of ‘Census’: the *population enumerations*, and the collection of *detailed statistical population data* (currently partly collected on census forms and partly through household surveys).
- 6.10. Improvement of each census component part requires a different approach. Before any data can be reliably collected through surveys, a full, equitable, accurate population enumeration that can serve as the ‘backbone’ will be essential.
- 6.11. A major – and long known - gap that influences the quality of population enumeration statistics is the absence of an accurate and up-to-date address register of the entire population in New Zealand. As described in the methodology of the creation of the sampling frame for Census 2023⁶, and commented on in the Statutory Review 2023⁷,

⁶ <https://www.stats.govt.nz/methods/creating-the-census-dwelling-frame-for-the-2023-census/>

⁷ <https://www.stats.govt.nz/reports/report-of-the-statutory-review-of-new-zealands-2023-census/#:~:text=2024%2C%2012%3A00pm-Report%20of%20the%20Statutory%20Review%20of%20New%20Zealand's%202023%20Census,makes%20recommendations%20for%20future%20censuses.>

this problem has not been resolved and will be even more critical in the move away from a full enumeration towards an administrative census.

a) Examples of equity implications for public health programmes of inaccurate and incomplete address registers

An accurate address register is not only an essential part of the sampling frame for a census, inaccurate addresses also play an important role in the (equitable) delivery of public services. Addresses are essential for the electorate roll, but also for cancer screening, and delivery of vaccinations under the National Immunisation Programme and during pandemics. Inaccurate addresses are one of the main reasons why people can't be recalled for cancer screening or vaccinations. The lack of such an accurate address register was a major problem during the COVID-19 pandemic when vaccinations were delivered, and coverage reported in small geographic areas.

The National Screening Unit in Te Whatu Ora delivers cervical cancer screening, breast cancer screening and bowel cancer screening. To identify and recall people who are eligible, they maintain 3 different population registers, all using similar, partially manual, methods as StatsNZ describes to create their Statistical Location Register (SLR): a team collects several quite random databases, compares addresses, guesses what the most likely accurate address is, and recalls people for screening.

Mobility of populations is related to deprivation. In New Zealand, Māori and Pacific populations are more mobile than the majority population. Inaccuracy of address data is a major factor in the inequitable delivery of these public health measures.

Currently, SLR only requires accurate address data once every 5 years before the census, whereas public health programme delivery would be optimised if address data was available real-time.

It would be a major improvement in accuracy and (cost)effectiveness of public health programmes if there was one national accurate address database, based upon which all public services could be delivered.

A consistent accurate address database would also be a major step towards more equitable delivery of care, and therefore eventually contribute to reducing the life expectancy gap.

- 6.12. Consistency of address data is also an essential condition to safely move to an administrative census. StatsNZ (Cabinet paper on 27 October 2015) noted that standardisation of a small selection of demographic data: 1) name, 2) sex, 3) date of birth, 4) address, 5) ethnicity and 6) Māori descent and iwi; was needed, together with sufficient data description and time reference to enable efficient and effective data re-use.
- 6.13. Standardisation of this data has not progressed sufficiently since 2015 to be considered consistently reliable for accurate admin data linkage.

- 6.14. In the absence of unique national identifiers, administrative data will be linked based on these demographic data. Therefore, the standardisation of this data is essential to avoid data linkage inaccuracies, that can go unnoticed. Automatic linkage based on algorithms that cannot be checked is a risk and widespread cause of systemic ethnic bias.
- 6.15. If linked admin data were to be used as a census 'backbone', inaccuracies and inequities in the linking process could have a major impact on the accuracy of the census backbone.
- 6.16. Since the standardisation of basic administrative data, has not been achieved, we suggest an alternative method to standardise this data.

7. Improving data quality to create a new statistical population 'backbone' using IT solutions

- 7.1. We recommend using an innovative IT solution to create a new, reliable, timely and equitable, population backbone in a separate IT application.
- 7.2. This backbone will serve as the new 'gold standard' for population enumeration statistics and serve as the 'backbone' to give context to other admin data, as well as the sampling frame for future population surveys.
- 7.3. The IT solution focusses on the standardisation of a minimal dataset only, which contains: 1) name, 2) gender/sex, 3) DoB, 4) address, 5) ethnicity, and 6) Māori descent and iwi (if desired).
- 7.4. This data is currently collected by many government agencies, although it varies in consistency and quality. This means no additional data needs to be collected by agencies.
- 7.5. The new IT application would centrally store and monitor the quality of this small dataset on behalf of all government agencies.
- 7.6. Collecting data through a shared application enforces and guarantees data collection consistency and quality.
- 7.7. The IT application could be linked to other IT systems by means of 'chain automation'.
- 7.8. Each time a citizen interacts with any agency, their key data is automatically retrieved from the central application, shown to the client, checked for accuracy, updated if necessary and send back to be centrally stored, saved and logged.
- 7.9. For transparency citizens can be given access to their data stored in the application through RealMe. They can also be given the option to update their information if needed

8. Suggested system's advantages

- 8.1. This system would align with the recent change in the Data and Statistics Bill which proposes the Government Statistician be given more power to collect data through public sector agencies
- 8.2. It will solve all current linkage problems in the IDI, that are an important cause of inequity for Māori and Pacific populations.
- 8.3. A more accurate and timely statistical population backbone would enhance all statistics and surveys. It would also make comparisons between studies and surveys more reliable and therefore decrease unintended institutional bias. It is only a minor change from the current suggested method by StatsNZ in creating a new census

backbone, in the sense that in the current StatsNZ proposal, data is linked retrospectively based on assumptions and statistical methods, whereas in our suggested method, data would be linked in real-time and would likely deliver population statistics of better quality than we have had in the past decade(s).

- 8.4. The application with basic information can be stored in New Zealand to meet Māori data sovereignty requirements. It will also allow Māori and Pacific communities to collect their own (local) data and link these datasets to the wider backbone to compare outcomes with other ethnic groups when needed and agreed.
- 8.5. Privacy protection is much more controllable. For example, sensitive medical data can still be securely stored by the health sector, but still be purposefully linked (anonymously) to the population backbone to obtain reliable population-based health statistics.
- 8.6. It will also facilitate and enhance the production and quality of interagency statistics with data no longer having to be linked based on name, date of birth and address.
- 8.7. It will also provide the opportunity to create a new individual sampling frame to replace the current Household Sampling Frame. Ideally, individual population sampling frames are used for population surveys because they allow for better targeted sampling (more targeted stratification, more targeted selection of the sample population), population sampling frames allow for individual response rate calculations, and they allow for more accurate weighting of the survey results and therefore improved reliability of conclusions drawn.
- 8.8. Not only is the current household sampling frame – based on full enumeration census - most inaccurate for Pacific populations, the use of household instead of population sampling makes it impossible to correct for other biases such as non-response. Given the low census response rates, and the low inclusion of Pacific populations in all non-governmental surveys – non-response is likely to be yet another bias that impacts the Pacific population most.
- 8.9. The establishment of a more accurate statistical population backbone would provide for the opportunity to transition from a Household Sampling Frame to a more accurate population sampling frame with smaller sample sizes needed to obtain reliable survey results and improving the quality of the survey results and equity for Pacific populations.