



April 2023

## Māori Linked Administrative Data: Te Hao Nui—A Novel Indigenous Data Infrastructure and Longitudinal Study

Reremoana Theodore

*University of Otago, New Zealand. moana.theodore@otago.ac.nz*

Amohia Boulton

*Whakauae Research for Māori Health and Development, New Zealand. amohia@whakauae.co.nz*

Andrew Sporle

*University of Auckland, New Zealand. a.sporle@auckland.ac.nz*

---

### Recommended Citation

Theodore, R., Boulton, A., & Sporle, A. (2023). Māori Linked Administrative Data: Te Hao Nui—A Novel Indigenous Data Infrastructure and Longitudinal Study. *The International Indigenous Policy Journal*, 14(1).  
<https://10.18584/iipj.2023.14.1.13412>

---

# Māori Linked Administrative Data: Te Hao Nui—A Novel Indigenous Data Infrastructure and Longitudinal Study

## Abstract

Worldwide, large amounts of administrative data are collected within official statistics systems on Indigenous Peoples. These data are primarily used for government and state policy purposes as opposed to by Indigenous Peoples to support Indigenous agendas (Taylor & Kukutai, 2017). In Aotearoa me Te Waipounamu New Zealand, Māori need high quality data to develop evidence-based policies and programs and to monitor government policies that impact on Māori. In this methodological paper, we describe uses of administrative data for Māori and current barriers to its use. We outline the development of a novel administrative data infrastructure and future longitudinal study. By explicating our Indigenous initiated, designed and controlled data project, we make a methodological contribution to Indigenous Data Sovereignty and Kaupapa Māori (Māori worldview) epidemiology.

## Keywords

Indigenous Data Sovereignty, Māori, wellbeing, policy, youth, administrative data

## Acknowledgments

The Te Hao Nui project is supported by the New Zealand Health Research Council (HRC) [Grant number 18/489]. Dr Reremoana Theodore is supported by a HRC Māori Health Research Emerging Leadership Fellowship [Grant number 18/664].

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

---

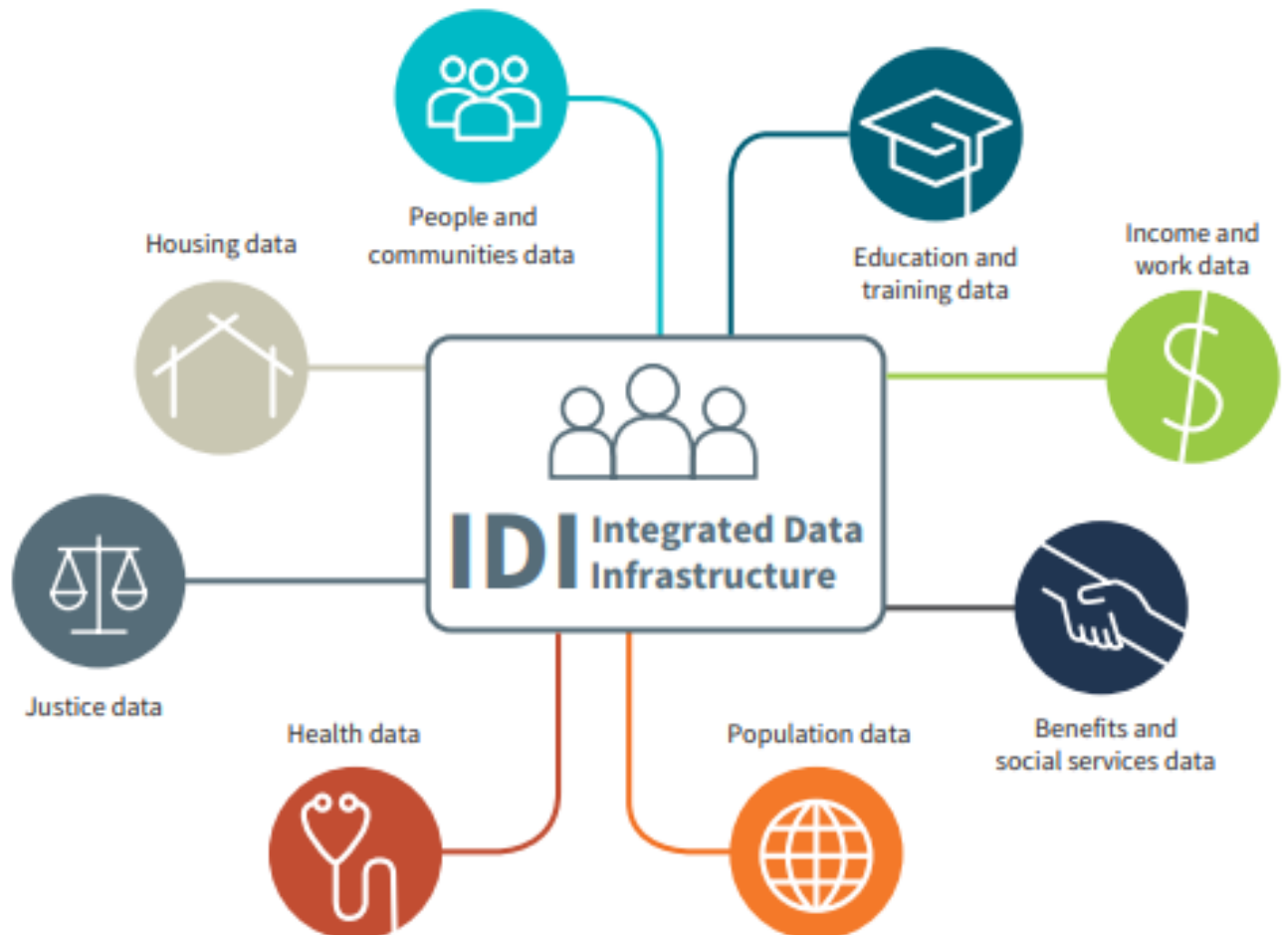
Worldwide, Indigenous Peoples are asserting their rights to fully participate in evidence-based policy-making and programs that impact their peoples and lands (Walter et al., 2020). Indigenous Peoples require good quality data to inform decision-making and program development and to monitor government and state policies and practices. In Aotearoa me Te Waipounamu New Zealand, there is an abundance of administrative data collected within the Official Statistics System on Māori individuals, whānau (families), households, Iwi (tribes) and Māori collectives. To date, these data have been primarily used for governmental purposes as opposed to supporting Māori development aspirations and agendas (Taylor & Kukutai, 2017). The New Zealand government, however, has a responsibility to ensure data resources are accessible and useful for Māori in accordance with its obligations as a partner to Te Tiriti o Waitangi (the Treaty of Waitangi: New Zealand's founding document). Moreover, greater levels of research funding and investment in training and capacity building are necessary in order for Māori to become authentic partners in producing and using official statistics (Statistics New Zealand, 2014a). In this methodological paper, we briefly describe New Zealand's administrative data and outline several challenges affecting the ability of Māori collectives to access, analyse, and use their own administrative data. We discuss the growing Indigenous Data Sovereignty movement, principles, and practices and their impact on Indigenous policy development. We then outline the creation of a novel, permanent, administrative data infrastructure and longitudinal study called Te Hao Nui. We highlight the importance of the infrastructure and future study, its alignment with Indigenous Data Sovereignty principles, and ability to monitor Māori wellbeing over time and contribute to the strategic aspirations of Iwi/Māori.

### **Administrative Data and the Integrated Data Infrastructure (IDI)**

In this section we describe administrative data and, in particular, New Zealand's Integrated Data Infrastructure (IDI). These are powerful tools currently used by the New Zealand government to try to address "wicked" problems (problems, in planning and policy, that are difficult to solve, like reducing poverty) (Boston & Gill, 2017; Sepuloni, 2018). We also discuss how administrative data have been used by the New Zealand government to try to address Tiriti (Treaty) obligation failures, using the 2018 Census failure as an example.

Administrative data are routinely collected data that are primarily gathered as a result of operational activities rather than for a specific research purpose (e.g., health service delivery information). In Aotearoa me Te Waipounamu New Zealand, government-funded policy agencies and service providers hold their own data in secure data environments, about their own services and operational transactions. In addition, there is a research database of linked data called the Integrated Data Infrastructure (IDI) which holds linked administrative data about people and households (Statistics New Zealand, 2018b). These data are collected by a range of government agencies, Statistics New Zealand (Stats NZ) surveys, and non-government organisations (Statistics New Zealand, 2017), and collated into a single database within a secure Stats NZ data environment. Some datasets are merged directly using common unique identifiers while other datasets that do not have common unique identifiers are linked using demographic information (Statistics New Zealand, 2014b). After data from separate datasets are linked via reference to a separate individual identity dataset, information that can identify a person (e.g., names, addresses) are removed to create de-identified data. For further information on the linking methodology used in the IDI, see Statistics New Zealand (2014b). The IDI contains longitudinal data (information on the same people at different time points) from over sixty different data sources including data on justice, health, accident

compensation, education, taxes, social welfare and social services, data on people and communities, population (e.g., census), income and work, and housing (Figure 1). The IDI is regularly being refreshed with updated source data and extended with new datasets (for more information see Statistics New Zealand, 2020a).



**Figure 1: Data in Statistics New Zealand’s Integrated Data Infrastructure (IDI). Image taken from Statistics New Zealand**

The IDI is used heavily by government agencies for a range of purposes including social investment analyses. Social investment applies “rigorous and evidence-based investment practices to social services” (Treasury New Zealand, 2017). While the concept of social investment is not a new idea, over the last decade it has gained increasing traction in New Zealand, becoming a key focus of policy work (Boston & Gill, 2017; Sepuloni, 2018). More recent and intensive approaches to social investment were led by the fifth National Government (2008-2017) and drew on Big Data approaches and data analytics to inform evidence-based policy interventions (Boston & Gill, 2017; Sepuloni, 2018). In an effort to stimulate innovative solutions to enduring or “wicked” social problems, this government also made linked administrative data more open and accessible to users outside of public policy processes (e.g., academics) (Teng et al., 2017).

The IDI is considered to be a world-leading resource that can be used to analyse data about complete populations. It provides an alternative approach to researching social issues alongside the usual sample survey research which has been the more common way of conducting policy research in New Zealand. The IDI allows for the investigation of the impacts of services, interventions, and programs (Social Investment Agency, 2017). It enables researchers to examine compounding and inter-related factors that can impact what happens to people's lives over time. The longitudinal nature of the IDI also allows researchers to investigate the impacts of cumulative advantage and disadvantage. These types of data can help to identify critical points in the lifecourse to intervene with programs that support health and wellbeing. The data can therefore inform long-term policy, planning, and monitoring of services and programs by and for Māori collectives. The large number of records involved also enables research on the Māori population with enough statistical power to produce results with high levels of precision. Smaller population studies in New Zealand often have numbers of Māori participants in their sample that are inadequate for robust research that focuses on the Māori population within the research. This means that, to date, quantitative research with Māori has not had the same depth and breadth as research with non-Māori in New Zealand.

Recently the IDI and administrative data were used extensively by Stats NZ to supplement the 2018 Census of the New Zealand population (Statistics New Zealand, 2019). As described in the independent review of New Zealand's 2018 Census, this "backfilling" activity was necessary due to a significant under-count which resulted from lower-than-expected response rates to Census 2018. In particular, poor response rates disproportionately affected Māori compared to non-Māori, requiring 23% of information on the Māori ethnic group to be taken from administrative data. The comparable rate for Pākehā (New Zealander of European descent) was only 8%. Administrative data has its limitations however, and could not be used to fill the gaps in the Census for measures like Iwi affiliation and language because there were no comprehensive sources for that information at a population level (Statistics New Zealand, 2019). The independent review stated that in the future, Stats NZ needed to improve engagement with Māori organisations in the community and with key Iwi groups given the relationship with Māori as Tiriti partners (Statistics New Zealand, 2019). The review highlighted the need to invest in training, governance, and operational arrangements to facilitate better collaboration with Māori as Tiriti partners throughout the Census and more broadly in Stats NZ's broader survey work. Overall, the poorly executed and underfunded Census has been recognised as a significant failing on the part of the Government in upholding their Tiriti obligations (Statistics New Zealand, 2019).

### **Issues Affecting Administrative Data Use and Access for Māori**

In this section, we describe accessibility and usability issues that impact on Māori use of administrative data. Administrative data in New Zealand, discussed thus far, can be analysed through Stats NZ's standalone datasets, confidential unit record files (CURFs), and by using linked data within the IDI. In order to access the IDI, individuals and groups apply to Stats NZ for access to a secure IDI Data Lab. Stats NZ evaluates applications on a set of criteria which include whether the research is for a statistical purpose, for the public good, is conducted by a credible team, whether suitable data are available, and that Stats NZ can enforce an agreement (Social Investment Agency, 2017). Only approved researchers can access microdata in secure data labs. Microdata are anonymised unit record datasets (individual response data in surveys and censuses) (Statistics, New

Zealand, 2020b). Microdata information cannot be removed from the secure IDI data lab environment (Statistics, New Zealand, 2020b). Only aggregated data, results, tables, and graphs, that are approved by Stats NZ, can be exported from the secure data lab environment. Aggregation is a method of protecting sensitive information by collapsing categories (e.g., where the count is “too small”). Stats NZ uses a risk management framework to protect against the disclosure of confidential information (Statistics New Zealand, 2020b).

Researchers currently need high levels of technical, programming, and statistical skills to use IDI data in data labs. Data are stored in an inconsistent format (e.g., hundreds of tables) which also make the datasets difficult to use and there is a need for easier user interfaces. A person or group’s ability to use the IDI is also increasingly being affected by user demand in particular requests to add new datasets. This creates challenges to the current infrastructure because the IDI was originally designed as an experimental prototype and it now requires scaling up (e.g., increasing its processing capabilities) in order to meet current and future demand (O’Neill, 2019). This would involve changing the IDI from a prototype to a data warehouse model to enable increases in datasets (e.g., data volume) (O’Neill, 2019).

Māori collectives need analysts and researchers who are technically proficient to be able to manipulate the data via access to microdata. A medium to long-term strategy to improve administrative data use for Māori is to build statistical and research capabilities by training more Māori data scientists and statisticians (Kukutai et al., 2020). This requires a focus on increasing the number of Māori who study and use statistics. Māori graduates, however, are currently less likely to graduate with a Science degree compared to non-Māori graduates (Theodore et al., 2016) and it takes at least seven years of post-high school education to attain a PhD. At present, Māori scientists are severely under-represented, compared to non-Māori, in New Zealand universities and in Crown Research Institutes (science research businesses owned by the Crown (i.e., the New Zealand Government)) (McAllister et al., 2020).

In Aotearoa me Te Waipounamu New Zealand, the government—as a Tiriti partner—has a responsibility to correct current asymmetries in data use capabilities between Māori and non-Māori. This would help to enable partnership and more equitable benefits in line with Indigenous Data Sovereignty principles (described in the following section) and also help address societal “wicked problems” (e.g., reducing inequities, poverty).

### **Indigenous Data Sovereignty**

Despite Indigenous Peoples using data throughout time, Indigenous Data Sovereignty as a concept is relatively new (Kukutai & Taylor, 2016). Indigenous data sovereignty is defined as “the rights of Indigenous Peoples to control the collection, access, analysis, interpretation, management, dissemination, and reuse of Indigenous data” (Walter & Carroll, 2020, p. 2). Indigenous Data Sovereignty, the rights to self-determination and the governance of data about Indigenous Peoples, lands, resources, knowledges, and ways of being, is affirmed in the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) (Walter & Carroll, 2020). Walter and Carroll (2020) state that Indigenous Data Sovereignty is fundamentally about Indigenous leadership (Walter & Carroll, 2020). As such, Indigenous Data Sovereignty includes the understanding that data are subject to the laws of the nation where it is stored and collected, and not only supports tribal sovereignty but also the realisation of Indigenous aspirations (Te Mana Raraunga: Māori Data

Sovereignty Network, n.d.). Kukutai et al. (2020) have described four core principles of Indigenous Data Sovereignty which together form the CARE framework, namely (i) Collective benefit—data ecosystems are designed to enable Indigenous Peoples to benefit from data; (ii) Authority to control—Indigenous rights and interests in Indigenous data should be recognised and their authority should be empowered to control these data; (iii) Responsibility—people working with Indigenous data are responsible for sharing those data to support the self-determination and collective benefit of Indigenous Peoples; (iv) Ethics—the rights and wellbeing of Indigenous Peoples is the primary concern through data life cycles and across the data ecosystem.

In Aotearoa me Te Waipounamu New Zealand, Māori could and should be able to use their own administrative data, including their data in the IDI, as a Tiriti (Treaty) right. Yet, as described previously, Māori are less able to use their own data compared to other groups (e.g., government departments) which is unacceptable and counter to Indigenous Data Sovereignty goals (Kukutai & Cormack, 2020). These goals include ensuring data for and about Māori are protected and safeguarded, Māori involvement in the governance of data repositories, and supporting the establishment of Māori data infrastructure, including security systems. Moreover, many Iwi are moving into a post-Treaty settlement era, using governance entities to manage resources and fund programs and services. Therefore, they require good quality data to support their planning for future pathways and forward-focused development agendas. The Ngā Tikanga Paihere framework, developed in 2018, provides a guide to appropriate use of the IDI particularly data about Māori and other under-represented groups (New Zealand Government, 2018). The framework outlines the need for researchers to: have appropriate expertise, skills, and relationships with communities; be accountable and transparent to communities of interest; have good data standards and practice kaitiakitanga (data stewardship and governance); ensure communities are involved in research decisions as early as possible; make certain that community objectives align with research objectives; and ensure that benefits are balanced with risks, including the identification of sensitivities in the use of data (e.g., privacy issues for Māori collectives). Ngā Tikanga Paihere guidelines also describe the importance of development opportunities with communities (e.g., improving data literacy, capability, and resource sharing).

Overall, Iwi and other Māori collectives are taking an increasing interest in the data sovereignty space, as evidenced by the establishment of Te Mana Raraunga, the Māori Data Sovereignty Network in 2015 and the establishment by the National Iwi Chairs' Forum of a specific Data Iwi Leaders Group. Similar groups have formed in other countries. For example, the First Nations Information Governance Centre was established in 2010 in Canada. In 2016, the United States Indigenous Data Sovereignty Network was founded, followed by the Maiam nayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Collective in Australia in 2017 (Walter & Carroll, 2020). Worldwide, the Indigenous Data Sovereignty movement has become a network of Indigenous-led advocacy, education and research networks that aim to transform the data landscape to support the development, aspirations and wellbeing of Indigenous Peoples, including the Global Indigenous Data Alliance (GIDA) that was founded at a meeting in the Basque Country in 2019 (Walter & Carroll, 2020).

There is a huge cache of administrative data related to Indigenous peoples that have the potential to bring about a new era in policy development and delivery for Indigenous communities worldwide (Kukutai et al., 2020). Rapidly changing data ecosystems, Big Data, and Open Data, however, can create serious issues for Indigenous Peoples. They increase the ability for non-Indigenous users to

analyse Indigenous data at greater distances from where the data have been collected (Walter & Carroll, 2020). This means that data use, analyses, and interpretations occur away from the lived realities of Indigenous peoples, and are often informed by non-Indigenous worldviews and values (Walter & Carroll, 2020). Even worse, algorithms and Big data may result in outcomes that racialise and increase the surveillance of Indigenous communities (Kukutai et al., 2020).

### **Te Hao Nui**

One way of realising the potential of data for Indigenous Peoples that is consistent with Indigenous Data Sovereignty principles and practices is to create innovative Indigenous-led studies. In this section, we describe the development and initial stages of a novel data infrastructure and longitudinal study called Te Hao Nui. The overall aim of Te Hao Nui is to help improve policy and service interventions at local and national levels in Aotearoa me Te Waipounamu New Zealand that support wellbeing. The project will leverage off pre-existing state investment in official statistical resources. As a team, we are designing Te Hao Nui in response to the expressed need of Māori collectives for high quality information to inform and monitor programs to improve Māori wellbeing. The information produced by this research is intended to not only identify intervention targets but, through the creation of a permanent data infrastructure, act as a monitor of Māori outcomes and interventions into the future. The new Māori knowledge infrastructure within the IDI will then be made available for future research. Our work aligns with findings from a recent publication that highlighted the current limitations of the IDI for Māori data and Māori users (Greaves et al., 2023). In particular, the need for significant improvements to be made to, or rebuilding of, the IDI, to ensure that it becomes a safer and more effective tool for Māori self-determination (Greaves et al., 2023). By creating a new Māori infrastructure, we hope to transform the availability of research results, resources, and the existing Official Statistics System to Māori collectives. This will not only enable Māori-led research using current statistical resources, it will inform the improvement of those resources, as well as the creation of Māori data resources, by highlighting the limitations of existing official statistics to inform the achievement of Māori aspirations.

The Te Hao Nui project is Indigenous-led. Te Hao Nui research leaders (the authors of this paper) have extensive experience and expertise working in and with Indigenous-led organisations and undertaking Indigenous research. Moreover, one author and research-lead (Sporle) is a founding member of the Global Indigenous Data Alliance and Te Mana Raraunga—the Māori data sovereignty network. The co-design approach with key stakeholders is described in detail in the following section. Te Hao Nui research leaders are also working with a number of technical advisors and data analytics experts.

A number of features of the Te Hao Nui study require explanation, as the study takes a novel approach to combining existing government survey data with the expressed needs and aspirations of Māori communities. The long term plan for the Te Hao Nui project is to use linked data from Te Kupenga 2013 and Te Kupenga 2018 (see information below) with the IDI and the Longitudinal Census Database (LCD). This will enable researchers to create the world's largest and most comprehensive Indigenous longitudinal study using existing data, which will be capable of following individual pathways forward and backward in time (Figure 2).



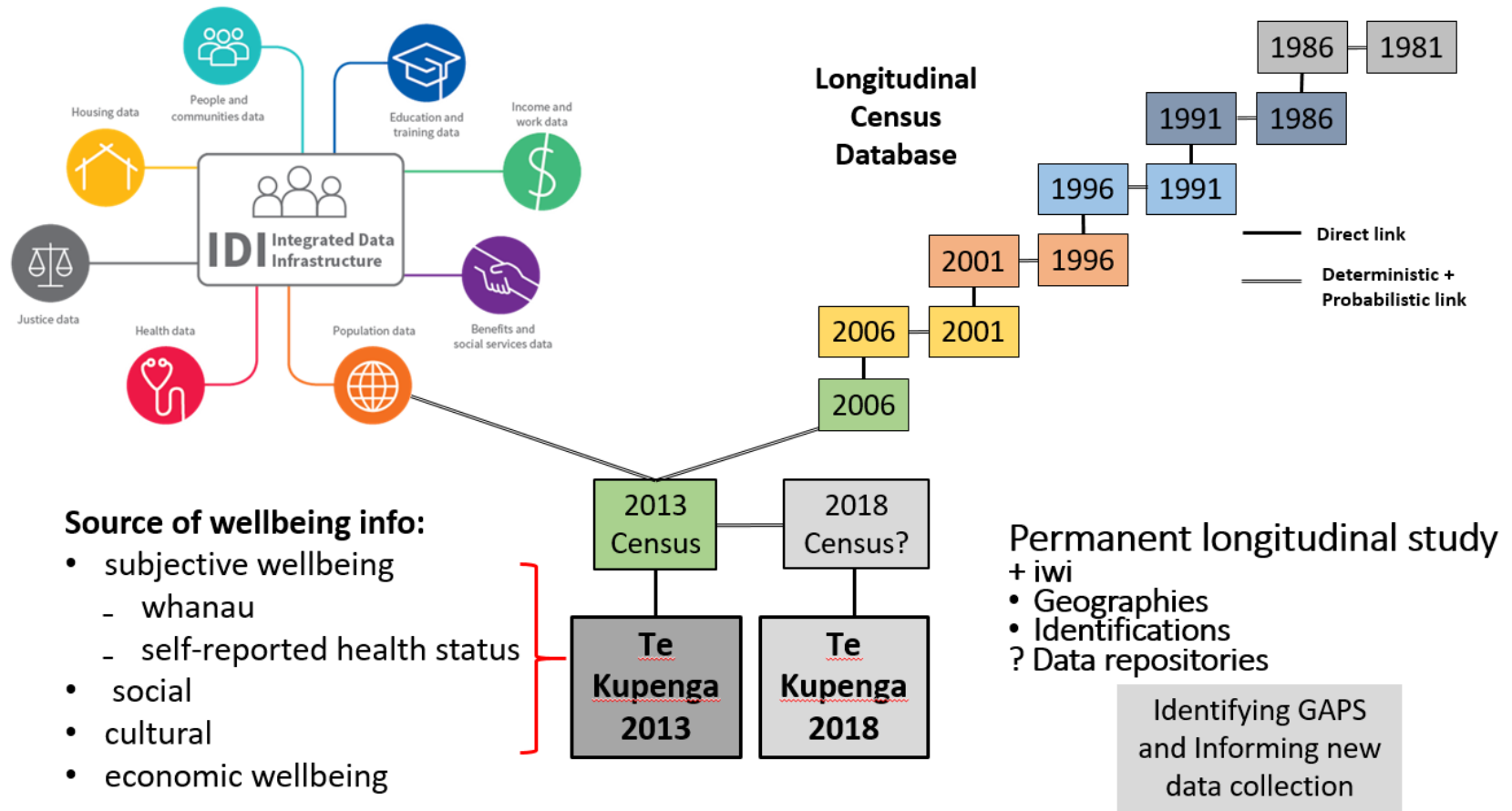


Figure 2: Overall Structure of the Linked New Zealand Data Resources for the Te Hao Nui project

The LCD contains data from six censuses (1981, 1986, 1991, 1996, 2001, and 2006) and has been described in detail previously (Statistics New Zealand, 2014b). In brief, the LCD is a longitudinal data source to enhance understandings on population change over time. The LCD is not currently within the IDI. If Stats NZ do not add the LCD to the IDI in the near future, we plan to undertake initial analyses using the two census datasets that are currently within the IDI (2013 and 2018). Future work will then use data from the 1981-2013 Censuses that are available in the LCD.

The Te Kupenga surveys have been described elsewhere in detail (Statistics New Zealand, 2014c; Statistics New Zealand, 2018). In brief, Te Kupenga 2013 was the inaugural nationally representative survey of Māori wellbeing undertaken by Stats NZ following the 2013 Census. Te Kupenga participants were a sample of the resident Māori population aged 15 years or older who identified as having Māori ethnicity or descent in the 2013 Census (n=5549). The survey involved a complex sampling strategy and used sample weights to create a nationally representative sample. Te Kupenga was the first official survey in Aotearoa New Zealand to include culturally informed variables, including information about the cultural, social, and economic wellbeing of individuals. As such, it was a significant step forward for Stats NZ in terms of responding to Māori requests for high-quality data (Kukutai & Walter, 2015). Following the 2018 Census, Te Kupenga was repeated with some question changes, (Statistics New Zealand, 2018a). A key difference between the 2013 and 2018 surveys was an increased sample size to approximately n=8,500. Te Kupenga 2018 has suffered from delays, particularly with regard to the dissemination of results, due to the issues with the 2018 Census. Provisional information was released in April 2020 (Statistics New Zealand, 2020c).

Initial Te Hao Nui analyses will be undertaken using Te Kupenga 2013 data. When Te Kupenga 2018 data are added to the IDI and quality assessed, we will then link and use the 2018 data. To date, members of the Te Hao Nui team and technical advisors have been working with Stats NZ to support this work. Te Kupenga are post-census surveys and the censuses are the sampling frame meaning that the surveys are automatically linked to census data. Te Kupenga 2013 is an official statistics survey, therefore its administration is determined by legislation covered by the Statistics Act 1975 including limiting data access to public good projects (New Zealand Government, 1975).

The key focus of initial Te Hao Nui analyses and research projects will be on the wellbeing of rangatahi or young people aged between 15 to 24 years. Māori make up 16.5% of the New Zealand population (n=775,836) according to the 2018 census. Māori are a youthful population, with more than half the population aged 25 years or younger. Research has demonstrated that adolescence is a life stage where many health-determining behaviours are established (Clark et al., 2013). Currently rangatahi outcomes are measured periodically and in a cross sectional (measured at one time point) manner. This prevents the identification of the determinants of wellbeing and the timely assessment of changes in the population level outcomes—including the impact of interventions.

The Te Hao Nui research team will initially examine the wellbeing of rangatahi based on the information that they provided when they took part in the Te Kupenga 2013 survey. Exposure measures/predictors in the Censuses and the IDI will be associated with various forms of self-reported wellbeing outcomes in the Te Kupenga surveys. Initial analyses will focus on the following variables in the Te Kupenga dataset: housing issues; experienced discrimination; employment; knowing one's Iwi; knowing one's ancestral marae (the courtyard of a Māori meeting house used for social or ceremonial purposes and often includes the buildings around the marae); Iwi registration; having been to ancestral marae; speaking te reo; importance of culture; and whānau wellbeing. Data

on key influencing factors from the Census will likely include household income, NZ Deprivation (an area-based socioeconomic measure), occupational socio-economic position, education, housing tenure, and household overcrowding. Data from the IDI will likely include education qualifications, years at school, number of schools, enrolled in a Primary Health (primary care) Organisation, and household mobility. We will calculate the prevalence of key influencing factors for the whole country and then for specific regions based on stakeholder needs, which we describe in the next section.

Given the unique nature of the Te Hao Nui project, including the need to develop the data infrastructure to make it fit for purpose, it is difficult to outline exact timelines for the project. Moreover, issues with the availability of 2018 Census data and related administrative datasets have meant that it has taken substantial time to fix these infrastructure issues and undertake quality assessments. Our research team had planned to have all major infrastructure issues addressed by the end of 2021 but the extensive COVID-19 lockdowns in Auckland prevented access to the Stats NZ data access portals for 18 months. This work will be completed by early 2023. Planned training workshops with key stakeholders will take place throughout 2023, as well as the documentation of the research processes. Analyses and report-writing on research examining the wellbeing of rangatahi (described above) will also take place in 2023. Future topics of interest will be identified with key stakeholders in 2024.

### **Designing a Project to Align with Indigenous Data Sovereignty Principles and Practices**

The design of the Te Hao Nui project can illustrate how to initiate research using administrative data by Māori for Māori in a way that upholds Indigenous Data Sovereignty principles and practices. Key aspects of the project include working with Māori collectives from the initiation of the project to enable them to access their own administrative data, the design of Māori data governance models, and the building of Māori community research and analytic capability. As described previously the concept of Indigenous data sovereignty is relatively new. The Global Indigenous Data Alliance (GIDA), formed in 2019, endorsed and host the CARE Framework for Indigenous Data Governance (described previously). In addition to international best practice, GIDA note that national Indigenous data sovereignty groups are best placed to respond to the needs of their communities. As described previously, the Ngā Tikanga Paihere framework, developed in 2018, provides a guide to appropriate use of the IDI particularly data about Māori (New Zealand Government, 2018). In this section, we will describe how key aspects of the study align with the CARE and Ngā Tikanga Paihere frameworks.

A number of Māori collectives will be key stakeholders in the Te Hao Nui project. Since the conception of the project, the research team have worked with Māori collectives who were interested in being named key stakeholders and with whom they have strong existing working relationships. Discussions to date have informed the initial selection of Te Hao Nui research questions. In particular, the focus on rangatahi wellbeing pathways and the identification of modifiable risk and protective factors. This includes the need to identify variables that can inform locally-based interventions and resource allocation, are amenable to policy or program intervention, allow stakeholders to monitor change over time, and enable them to lobby for resources. Other stakeholders will include a range of service providers such as Iwi-mandated social services, primary health service providers and other health entities. Having a diverse range of stakeholders means that the project is designed to meet the needs of urban, regional and rural stakeholders/providers. Information on Te Hao Nui key stakeholders is not publicly available yet. In 2023, memorandums of

understanding (MOUs) will be signed. Signing was delayed due to the unplanned need for detailed data quality assessments of Census data and related datasets, then by the extensive COVID-19 related lockdowns and travel restrictions in 2020-21.

The Te Hao Nui co-design approach aligns with the Ngā Tikanga Paihere framework of ensuring communities are involved in research as early as possible and in accordance with CARE principles of being responsible for sharing data to support the self-determination and collective benefit of Indigenous Peoples. Previous research has highlighted the importance of working with and understanding the needs of stakeholders to improve health and wellbeing services for Māori (e.g., Port et al., 2008). Studies have also shown the importance of co-design processes for Māori given that programs and interventions designed for the general population tend to be less effective for Māori and may even increase inequities (Te Morenga et al., 2018).

An Indigenous quantitative methodological advisory group (members still to be named) will guide the Te Hao Nui project. This advisory group will include leading international Indigenous researchers who are members of GIDA. With support from international advisors, we are developing a Māori data governance model. Governance processes must be in place to ensure that Iwi have control over their identified data within the official data system – consistent with Indigenous Data Sovereignty practice. In particular, Māori stakeholders will have control over their data that they aggregate (i.e., a data resource), from the design to the dissemination stages. As already negotiated with Stats NZ, these stakeholders will also determine the protocol and processes regarding who can then use that aggregated data (data resource). For example, if a Māori collective are interested in aggregating data about a particular geographical area that is not a standard Stats NZ aggregation, then that collective will be able to determine who can access and re-use that information. The foundation of the Te Hao Nui project is to create new information processes that enhance rangatiratanga (sovereignty) over data resources. Our Māori stakeholders will therefore maintain control over their intellectual property. This work aligns with the Ngā Tikanga Paihere framework in relation to the need for good data standards and practice kaitiakitanga (data stewardship and governance) and CARE principles, including having data ecosystems designed to enable Indigenous Peoples to benefit from data.

The Ngā Tikanga Paihere framework outlines the importance of development opportunities with communities (e.g., building capability, improving data literacy). Building Māori capability to access and apply data resources is an important goal of the Te Hao Nui project. This goal will be achieved through the combination of a more accessible data infrastructure and workforce development initiatives. Workshops will be run with key stakeholders and tailored to meet local needs. These workshops will cover issues such as finding data, data quality, data ownership and governance, using data to monitor change and creating graphs using iNZight (described below). These workshops will be a “stepping stone” for non-statistically trained stakeholder members toward further training, should they wish. The creation of a new data infrastructure with standardised methods of reporting and access will lower the currently very high technical barrier to accessing official data. Stakeholders will have training on how to access the IDI system to view results. IDI portals (secure access networked computers) will be set up with our regional stakeholders so there will be a permanent and local resource for further training and research projects (all current data portals are currently in main centres). The research team will produce a technical users’ guide to assist new users to use the longitudinal data infrastructure. All code other than that governed by Iwi/Māori organisations will be put in the public domain via Stats NZ’s metadata repository for others to use and adapt for future

research. There will also be post-graduate opportunities for tertiary students to work in the Te Hao Nui project. To date, three summer students (students who undertake a small research project over the university break) and one honours student have worked on the project. More opportunities, including PhD and post-doctoral projects will become available in 2023 and beyond.

The research team will create new data visualisation tools to produce national level data and data for regions associated with our key stakeholders. This aligns with CARE principles, for example, ensuring data ecosystems are designed to enable Indigenous Peoples to benefit from data. The new tools will be adapted from current visualisation tools and geographical information system mapping (GIS) add-ons in the free iNZight software (insight.nz). The iNZight software is a point and click version of R (one of the world's leading statistical software packages). iNZight is easy to learn and was initiated in collaboration with the University of Auckland and Stats NZ with support from New Zealand's Ministry of Education. The use of iNZight has been taught in the New Zealand secondary school maths curriculum to children as young as year 9 (approximately 13 years of age) making the software familiar to most New Zealand secondary school students who were at school in the last decade (The University of Auckland, n.d.). iNZight will be used to visualise and analyse data. Datasets for stakeholders will be stored in the IDI. iNZight will then be used to visualise and analyse those data. iNZight is currently being piloted within the secure Stats NZ data environment, with the intention that it will be made available to IDI users in 2022. Aggregated data, however, will still need to go through the Stats NZ protocol to be released, as described previously. An Iwi-specific capability using rohe (region) maps will also be created if requested by Iwi stakeholders and there are strong data governance processes over the reuse of any mapping resources provided by Iwi.

Over time, for variables sourced from the IDI, regional prevalence rates will be updated and compared with baseline prevalence, to identify change over time. These methods can then be reused as an outcome monitoring tool. Importantly, the co-design process will inform research design, analyses and dissemination strategies over time. This is important for the translation of research findings and resources into Māori health and wellbeing gains at a local and national level. This aligns with the Ngā Tikanga Paihere Framework including ensuring community objectives align with research objectives. This work also aligns with the CARE principles including supporting the self-determination and collective benefit of Indigenous Peoples. Previous research has shown the importance of having knowledge users working alongside researchers to lead research in order to make research relevant and useful (Canadian Institutes of Health Research, 2012; Oetzel et al., 2017). Importantly, Te Hao Nui stakeholder partners will be resourced from the project's budget to enable stakeholder staff to work directly on the project aligning with the Ngā Tikanga Paihere framework of resource sharing with communities.

Like all studies, there are a number of strengths and limitations of our proposed study. As described previously, using administrative data enables our team to undertake research at a population level and we can provide information to inform long-term policy, planning and monitoring of services and programs by and for Māori collectives. Moreover, using data in the IDI allows us to examine information collected across a wide range of sectors (e.g., health, education). Working with Māori collectives we are able to focus the research on areas of importance to Māori communities as well as support them to have control over the analysis, interpretation, management and use of their own data for their own benefit. Conversely, there have been and will be a number of challenges to undertaking this project. As a team we have already experienced delays due to the poorly executed 2018 Census. Moreover, we are restricted to using administrative data collected to date. Fortunately,

the Te Kupenga surveys does enable us to examine culturally informed variables such as cultural and social wellbeing. Moreover, as variables of interest are identified by Māori collectives, we will be able to inform future official surveys.

As further lessons are learned within the Te Hao Nui project, we will share these in future research papers with other Indigenous Peoples to help inform best practice in regards to administrative data use by Indigenous Peoples for Indigenous Peoples in line with Indigenous Data Sovereignty.

### **Policy Relevance and Lessons Learned for Indigenous Peoples Internationally**

Indigenous Peoples globally have experienced colonisation and repeated breaches of local Treaties that resulted in widespread land confiscations, loss of Indigenous Knowleges and resources, the destabilisation of Indigenous socio-political organisations, racism, and discrimination. Colonisation led to high levels of poverty and poor health, education, and social outcomes affecting generations of Indigenous Peoples (Anderson et al., 2006; Reid & Robson, 1998). To date, data have been used to define and make sense of Indigenous Peoples and their circumstances, often describing their differences to non-Indigenous People and their levels of disadvantage (Walter & Carroll, 2020). A lack of connection between often well-meaning primarily non-Indigenous policy makers and Indigenous Peoples has resulted in the imposition of policies to “fix” these issues (Walter & Carroll, 2020). That is, policy makers who diagnose Indigenous problems and create policy solutions have often created policies that fail across Nation states. Walter and Carroll (2020) note that these policy failures that result in poor Indigenous policy outcomes have often come to be viewed as inevitable. Or these policy failures are explained as only perceived failures by Indigenous Peoples. That is, Indigenous Peoples are perceived as not taking advantage of the opportunities presented within policy programs. Therefore, the outcomes may be viewed as poor behaviour or choices made by Indigenous Peoples.

Data informs policy from the problem through to strategies to resolve the policy problem (Walter & Carroll, 2020). Traditionally, Indigenous data are used by primarily non-Indigenous policy makers to make sense of Indigenous Peoples. Walter and Carroll (2020) describe Indigenous Data Sovereignty as having the ability to invert the standard Indigenous data/policy nexus with Indigenous Peoples’ own “governance of data and data for governance” (Walter & Carroll, 2020, p. 11). This means Indigenous Peoples govern and use their own data to inform their own policies and programs, and to monitor government and state policies that affect their communities. It therefore focuses the policy question(s) and subsequent analyses on the aspirations and goals of Indigenous Peoples. This requires Indigenous leadership. In this paper we have described the development of a project, Te Hao Nui, that is undertaken by Māori for Māori, that has relevance for Indigenous Peoples internationally. In order to support the work of other Indigenous communities, our research team is committed to sharing information about the development, design and the inevitable results of this Indigenous-led data project.

In this methodological paper, we have discussed the potential uses of linked administrative data in Aotearoa me Te Waipounamu New Zealand for Māori and some current issues that affect its use. We have described the growing Indigenous Data Sovereignty movement. We then outlined the development phase of a novel Indigenous-led and designed permanent data infrastructure and longitudinal study—Te Hao Nui—which is being created using existing administrative data. The inevitable goal of Te Hao Nui is to improve local and national level policies and services to support

long-term Indigenous wellbeing and development. Using Indigenous Data Sovereignty principles and practices, other outcomes include an increased Indigenous stakeholder involvement with official statistical resources, including access and application of data resources, but extending to governance and co-design. Also, enabling the application of high quality statistical information to inform and monitor policy and services that support Indigenous collectives, so that the use of these data becomes standard practice.

## References

- Anderson, I., Crengle, S., Kamaka, M. L., Chen, T. H., Palafox, N., & Jackson-Pulver, L. (2006). Indigenous health in Australia, New Zealand, and the Pacific. *Lancet*, 367(9524), 1775-1785. [https://doi.org/10.1016/s0140-6736\(06\)68773-4](https://doi.org/10.1016/s0140-6736(06)68773-4)
- Boston, J., & Gill, D. (2017). *Social investment: A New Zealand policy experiment*. New Zealand: Bridget Williams Books. <https://doi.org/10.7810/9781988533582>
- Canadian Institutes of Health Research. (2012). *Guide to knowledge translation at CIHR: integrated and end-of-grant approaches*. Canadian Institutes of Health Research. <https://cihr-irsc.gc.ca/e/45321.html>
- Clark, T., Fleming, T., Bullen, P., Crengle, S., Denny, S., Dyson, B., Peiris-John, R., Robinson, E., Rossen, F., Sheridan, J., Teevale, T., Utter, J., & Lewycka, S. (2013). Health and well-being of secondary school students in New Zealand: Trends between 2001, 2007 and 2012. *Journal of Paediatrics and Child Health*, 49(11), 925-934. <https://doi.org/10.1111/jpc.12427>
- Greaves, LM., Latimer, CL., Muriwai, E., Moore, C., Li, E., Sporle, A., Clark, TC., & Milne, B. (2023). Māori and the Integrated Data Infrastructure: An assessment of the data system and suggestions to realise Māori data aspirations. *Journal of the Royal Society of NZ*. <https://doi.org/10.1080/03036758.2022.2154368>
- Kukutai, T., Carroll, S. R., & Walter, M. (2020). *Indigenous World 2020: Indigenous Data Sovereignty*. [Report]. <https://www.iwgia.org/en/ip-i-iw/3652-iw-2020-indigenous-data-sovereignty.html>
- Kukutai, T., & Cormack, D. (2020). "Pushing the space": Data sovereignty and self-determination in Aotearoa NZ. In M. Walter, T. Kukutai, S. R. Carroll, & D. Rodriguez-Lonebear (Eds.), *Indigenous data sovereignty and policy* (pp. 21-35). Routledge. <https://www.taylorfrancis.com/books/e/9780429273957/chapters/10.4324/9780429273957-2>
- Kukutai, T., & Taylor, J. (2016). *Indigenous data sovereignty. Towards an agenda*. ANU Press.
- Kukutai, T., & Walter, M. (2015). Recognition and indigenizing official statistics: Reflections from Aotearoa New Zealand and Australia. *Statistical Journal of the IAOS*, 31(2), 317-336. <https://doi.org/10.3233/sji-150896>

- McAllister, T. G., Naepi, S., Wilson, E., Hikuroa, D., & Walker, L. A. (2020). Under-represented and overlooked: Māori and Pasifika scientists in Aotearoa New Zealand's universities and crown-research institutes. *Journal of the Royal Society of New Zealand*, 52(1), 38-53. <https://doi.org/10.1080/03036758.2020.1796103>
- New Zealand Government (1975). *Statistics Act 1975*. Wellington: New Zealand Government. <https://www.legislation.govt.nz/act/public/1975/0001/latest/DLM430705.html>
- New Zealand Government. (2018). *Ngā Tikanga Paihere*. [Guidelines]. <https://data.govt.nz/use-data/data-ethics/nga-tikanga-paihere/>
- O'Neill, R. (2019 April 15). Stats NZ faces a further challenge: shorting up its Integrated Data Infrastructure. *Reseller News*. <https://www.reseller.co.nz/article/660010/stats-nz-faces-further-challenge-shoring-up-its-integrated-data-infrastructure/>
- Oetzel, J., Scott, N., Hudson, M., Masters-Awatere, B., Rarere, M., Foote, J., Beaton, A., & Ehau, T. (2017). Implementation framework for chronic disease intervention effectiveness in Māori and other indigenous communities. *Globalization and Health*, 13(69). <https://doi.org/10.1186/s12992-017-0295-8>
- Port, R. V., Arnold, J., Kerr, D., & Gravish, N. (2008). Cultural enhancement of a clinical service to meet the needs of indigenous people; Genetic service development in response to issues for New Zealand Māori. *Clinical Genetics*, 73(2), 132-138. <https://doi.org/10.1111/j.1399-0004.2007.00943.x>
- Reid, P., & Robson, B. (1998). Dying to be counted. In *Proceedings of Te Oru Rangahau Māori Research and Development Conference* (pp. 267-271). Palmerston North: Massey University.
- Sepuloni, C. (2018). *The launch of 'Social investment - A New Zealand policy experiment*. [Government Speech]. <https://www.beehive.govt.nz/speech/launch-%E2%80%98social-investment-new-zealand-policy-experiment%E2%80%99>
- Social Investment Agency. (2017). *Social Investment Agency's beginner's guide to the Integrated Data Infrastructure*. [Guidelines]. New Zealand Government <https://swa.govt.nz/assets/Documents/Beginners-Guide-To-The-IDI-December-2017.pdf>
- Statistics New Zealand. (2014a). *Guide to using Te Arotahi Tatauranga*. [Guidelines] Wellington: New Zealand Government. <https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Methods/He-Arotahi-Tatauranga/guide-to-using-he-arotahi-tatauranga.pdf>
- Statistics New Zealand (2014b). Linking methodology used by Statistics New Zealand in the Integrated Data Infrastructure project. New Zealand Government. <https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Methods/Linking-methodology-used-by-Statistics-New-Zealand-in-the-Integrated-Data-Infrastructure-project/linking-methodology-IDI-project.pdf>



- Statistics New Zealand. (2014c). *Te Kupenga 2013 (English)*.  
<https://www.stats.govt.nz/information-releases/te-kupenga-2013-english/>
- Statistics New Zealand. (2017). *Integrated Data Infrastructure*. [Report].  
<https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>
- Statistics New Zealand. (2018a). *Differences between Te Kupenga 2013 and 2018 surveys*. [Report]. <https://www.stats.govt.nz/methods/differences-between-te-kupenga-2013-and-2018-surveys>
- Statistics New Zealand. (2018b). *Integrated Data Infrastructure*. [Report].  
<https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>
- Statistics New Zealand. (2019). *Report of the Independent Review of New Zealand's 2018 Census*. [Report]. New Zealand Government. <https://www.stats.govt.nz/reports/report-of-the-independent-review-of-new-zealands-2018-census>
- Statistics New Zealand (2020a). *Data in the IDI*. <https://stats.govt.nz/integrated-data/integrated-data-infrastructure/data-in-the-idi/>
- Statistics New Zealand (2020b). *Microdata output guide (fifth edition)*.  
<https://www.stats.govt.nz/assets/Methods/Microdata-Output-Guide-2020-v5-Sept22update.pdf>
- Statistics New Zealand. (2020c). *Te Kupenga: 2018 (provisional) – English*. [Report].  
<https://www.stats.govt.nz/information-releases/te-kupenga-2018-provisional-english>
- Taylor, J., & Kukutai, T. (2017). *Indigenous data sovereignty: Toward an agenda*. ANU Press.
- Te Mana Raraunga, Māori Data Sovereignty Network. (n.d.). *Our data, our sovereignty, out future*. [Report]. <https://www.temanararaunga.Māori.nz/>
- Te Morenga, L., Pekepo, C., Corrigan, C., Matoe, L., Mules, R., Goodwin, D., . . . Ni Mhurchu, C. (2018). Co-designing an mHealth tool in the New Zealand Māori community with a “Kaupapa Māori” approach. *AlterNative: An International Journal of Indigenous Peoples*, 14(1), 90-99. <https://doi.org/10.1177/1177180117753169>
- Teng, A. M., Blakely, T., Ivory, V. C., Kingham, S., & Cameron, V. (2017). Living in areas with different levels of earthquake damage and association with risk of cardiovascular disease: a cohort-linkage study. *The Lancet Planetary Health*, 1(6), e242-253.  
[https://doi.org/10.1016/S2542-5196\(17\)30101-8](https://doi.org/10.1016/S2542-5196(17)30101-8)
- Theodore, R. F., Tustin, K., Ciro, K., Gollop, M., Taumoepeau, M., Taylor, N., Chee, K-S., Hunter, J., & Poulton, R. (2016). Māori university graduates: Indigenous participation in higher education. *Higher Education Research and Development*, 35(3), 604-618.  
<https://doi.org/10.1080/07294360.2015.1107883>
- The University of Auckland. (n.d.). *Census at School New Zealand*.  
<https://new.censusatschool.org.nz/resource/using-inzight-for-the-first-time/>

Treasury New Zealand. (2017). *Social Investment*. [Report].

<http://www.treasury.govt.nz/statesector/socialinvestment>

Walter, M., & Carroll, S. R. (2020). Indigenous data sovereignty, governance and the link to Indigenous policy. In M. Walter, T. Kukutai, S. R. Carroll, & D. Rodriguez-Lonebear (Eds.), *Indigenous data sovereignty and policy* (pp. 1-21). Routledge.

Walter, M., Kukutai, T., Carroll, S. R., & Rodriguez-Lonebear, D. (2020). *Indigenous data sovereignty and policy*. Routledge.