11 November, 2024
Tāmaki Makaurau | Auckland

Indigenous Data Sovereignty Masterclass

Workbook

With thanks to our Masterclass sponsors

Social Media

Our event hashtag is #IDSov24 Please be respectful in your social media communications and seek permission from presenters and/or attendees featured in your post prior to sharing. Recording of the Masterclass sessions is not permitted.

Contents

Section 1. Foundations of Indigenous Data Sovereignty

Bobby Maher & Maggie Walter

Learning objectives:

Objective 1	Understand key Indigenous Data Sovereignty (IDSov) terms and concepts
Objective 2	Identify the longstanding deficit patterning of Indigenous statistics
Objective 3	Understand how IDSov and Indigenous Data Governance (IDGov) can be the mechanism that disrupts the traditional tropes of Indigenous data

Key Concepts and Definitions[1]:

➢ **Indigenous data**

Refers to digital or digitisable data, information or knowledge, in any format or medium, which is about, from or connected to Indigenous people and may affect Indigenous peoples both collectively and individually. It includes data about population, place, culture and environment.

➢ **Indigenous Data Sovereignty**

Refers to the right of Indigenous people to exercise ownership over Indigenous data. Ownership of data can be expressed through the creation, collection, access, analysis, interpretation, management, dissemination and reuse of Indigenous data.

IDSov derives from the unceded sovereignty of Indigenous people.

➢ **Indigenous Data Governance**

Refers to the right of Indigenous peoples to autonomously decide what, how and why Indigenous data are collected, accessed and used. It ensures that data on or about Indigenous peoples reflects our priorities, values, cultures, worldviews and diversity.

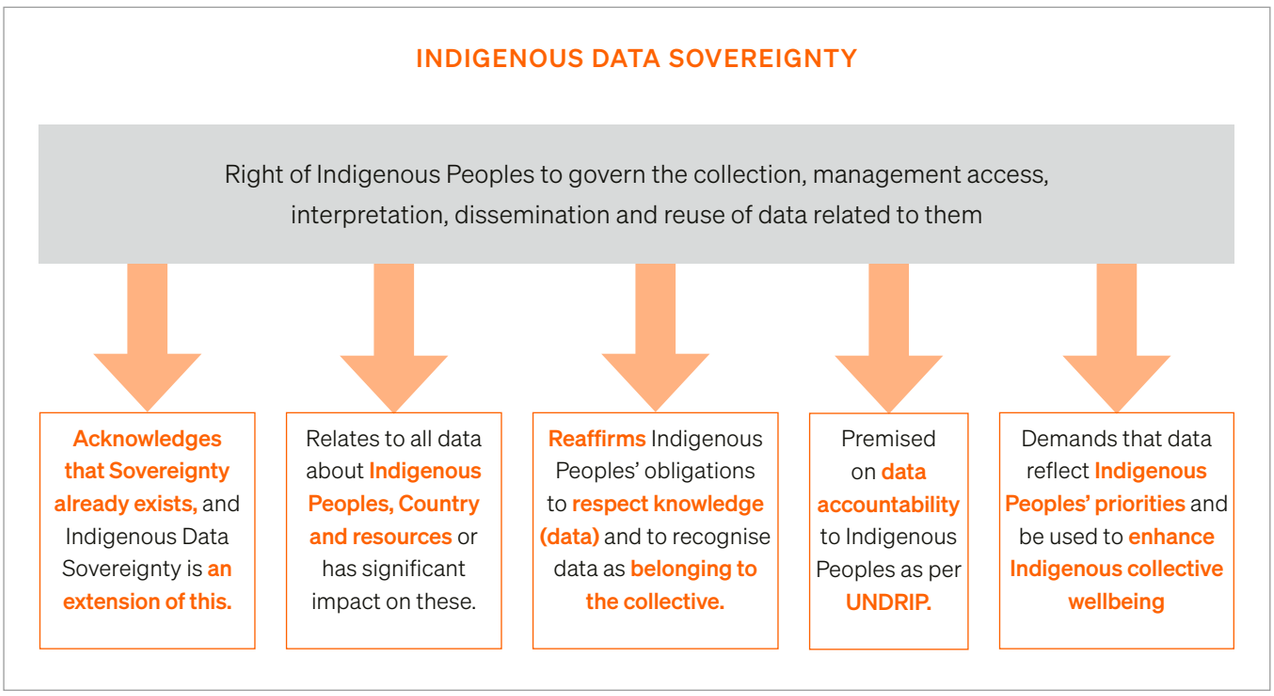
References:

Maiam nayri Wingara. (2018). *Indigenous Data Sovereignty communique.* Indigenous Data Sovereignty Summit 20th June 2018, Canberra, ACT.

Kukutai, T., Campbell-Kamariera, K., Mead, A., Mikaere, K., Moses, C., Whitehead, J. & Cormack, D. (2023). *Māori Data Governance Model*. Te Kāhui Raraunga.

1 These are widely used definitions. Some may vary slightly.

Activity 1: Identify the Indigenous data that are most precious to your community.



This slide reflects the global definition of IDSov. Indigenous peoples recognise and endorse the definition globally. It cannot be changed or co-opted.

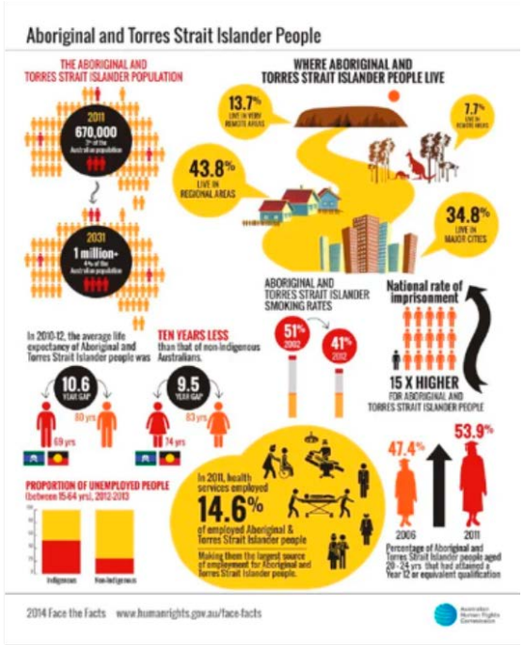
References:

Kukutai, T. & Taylor, J. (Eds.) (2016). *Indigenous data sovereignty: Toward an agenda*. ANU Press.
Snipp, M. (2016). What does data sovereignty imply? What does it look like? In T. Kukutai & J. Taylor (eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 39 – 55). ANU Press.

DOMINANT DATA PRACTICES

Colonial data practices

Difference
Disparity
Disadvantage = **5 D's of Data**
Dysfunction
Deprivation



Reference:

Walter, M. (2016). Data politics and Indigenous representation in Australian statistics. In T. Kukutai & J. Taylor (eds.), *Indigenous data sovereignty: Toward an agenda* (pp. 79 – 97). ANU Press.

BADDR DATA PARADIGM

Blaming Data
Too much data contrasts Indigenous/non-indigenous data, rating the problematic indigene against the normed Australian as the ubiquitous pejorative standard.

Aggregate Data
Too much data are aggregated at the national and/or state level implying Indigenous cultural and geographic homogeneity.

Decontextualised Data
Too much data are simplistic and decontextualised focussing on individuals and families outside of their social/cultural context.

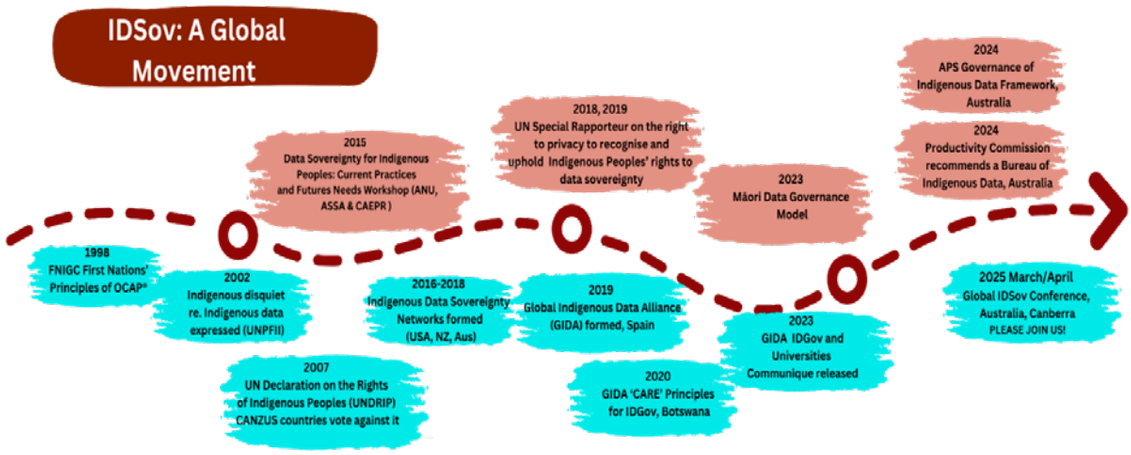
Deficit, Government Priority Data
Too much data reprises deficit linked concepts that service the priorities of Government.

Reductive Data
Too much data are reductive, treating Indigenous identity as a dichotomous, explanatory, dependent variable.

Reference:

Walter, M. (2018). 'The voice of Indigenous data: beyond the markers of disadvantage', *Griffith Review, 60*.

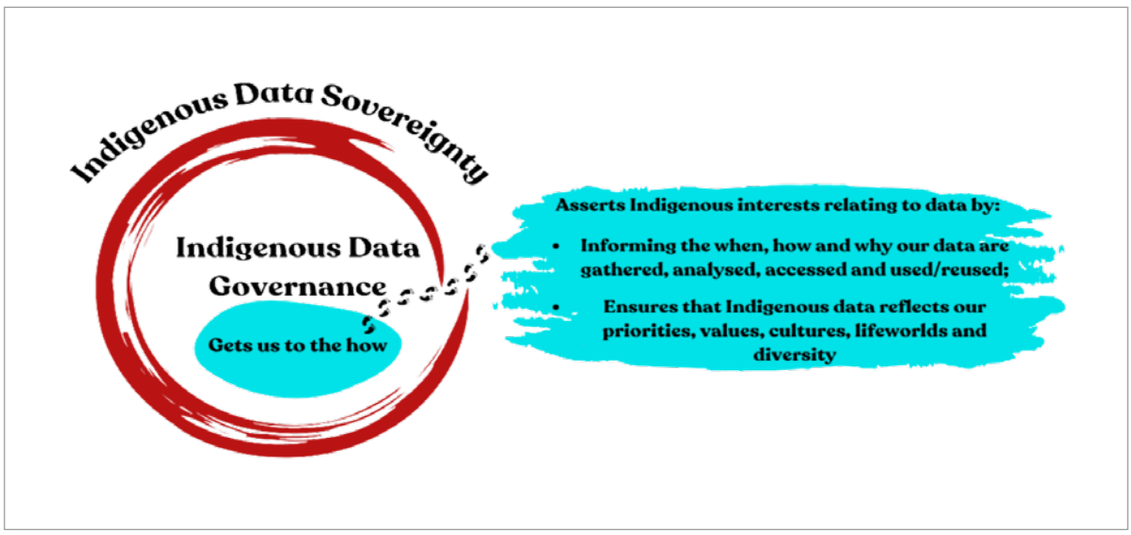
These slides reflect the dominant colonial data practices in Anglo-colonised countries and how Indigenous data are often weaponised against Indigenous peoples. Much data are collected **about us** but there is a dearth of data that are **by us, for us and with us**. Thus, the data that are collected often do not reflect our priorities, values, worldviews and culture. The BADDR data paradigm influences data-based research that excludes Indigenous data needs.



Activity 2: Choose a set of IDSov principles (Aotearoa NZ, Australia). In a small group discuss how each principle relates to the Indigenous data that are most precious to you (identified in Activity 1).

References:

Te Mana Raraunga (2018). *Māori data sovereignty principles*. Available at: https://www.temanararaunga.maori.nz/
Maiam nayri Wingara (2018). *Indigenous Data Sovereignty Principles*. Available at: https://www.maiamnayriwingara.org/mnw-principles



This slide reflects the definition for IDGov that has been developed and endorsed by Indigenous peoples globally. There is a shared understanding across national contexts that IDSov and IDGov must be **Indigenous led.** If it does not have Indigenous leadership, it **CANNOT** be IDSov or IDGov.

Indicators for Indigenous Data Governance

Indigenous Data Sovereignty Paradigm Indicators	BADDR Data Paradigm Indicators
Indigenous led, Indigenous controlled, Indigenous owned	Led, controlled and owned by the agency
Indigenous governance is 'built-in'; operates as high-level within organisational structure; is core business	Indigenous governance is an 'add-on'; exists as a limited structure (e.g. advisory body); is not core business but an after-thought
Pertains to all aspects of Indigenous data (re data life cycle)	Pertains to approve pre-determined decisions, data products and how these are disseminated
Operationalised through Indigenous Data Governance	Indigenous data are managed by the agency (as per its data management policy and procedures)

This slide reflects how IDSov is put into practice through IDGov. There are two mechanisms to IDGov that we will return to throughout the Masterclass:

- **Governance of Data** refers to decision-making about honouring, caring for, protecting and managing community and organisational data. The governance of data means Indigenous people decide what rules and processes apply to Indigenous data throughout the data lifecycle. This means Indigenous people are the decision-makers for the access, control and use of Indigenous data, particularly relating to data that are held by non-Indigenous organisations (for example, the Australian Bureau of Statistics and Australian Institute of Health and Welfare; other government agencies; non-government organisations).

- **Data for Governance** refers to Indigenous people having the required data for accurate, relevant, and timely decision-making. These data inform elements of good governance within and by Indigenous communities and First Nations, and include components like service delivery, allocation of resources, policy development, cultural protection and strategic decision-making.

References:

Maiam nayri Wingara (2023). *Taking control of our data: Discussion paper on Indigenous Data Governance for Aboriginal and Torres Strait Islander people and communities*. Melbourne: Lowitja Institute. doi: 10.48455/rtvd-7782 Available at: https://www.lowitja.org.au/wp-content/uploads/2024/01/Taking-Control-of-Our-Data-Discussion-Paper.pdf

Global Indigenous Data Alliance. (2023). *Indigenous peoples' rights in data.* doi:10.6084/m9.figshare.22138160 Available at: https://www.gida-global.org/data-rights

Resources

Watch 

Walter, M. (2018). 'Resetting Indigenous Data'. Keynote at the 5th National Aboriginal and Torres Strait Islander Health Summit. Available at: https://tinyurl.com/ms33ps6w

Read 

Maiam nayri Wingara (2023). *Taking control of our data: Discussion paper on Indigenous Data Governance for Aboriginal and Torres Strait Islander people and communities*. Lowitja Institute. doi: 10.48455/rtvd-7782. Available at: https://www.lowitja.org.au/wp-content/uploads/2024/01/Taking-Control-of-Our-Data-Discussion-Paper.pdf

Explore 

First Nations Information Governance Center - Canada (formed 1996). https://fnigc.ca/
Te Mana Raraunga Māori Data Sovereignty Network (formed 2016). https://www.temanararaunga.maori.nz/
United States Indigenous Data Sovereignty Network (formed 2016). https://usindigenousdatanetwork.org
Maiam nayri Wingara Indigenous Data Sovereignty Collective (formed 2017). https://www.maiamnayriwingara.org/
Global Indigenous Data Alliance – GIDA (formed 2019). http://www.gida-global.org

Section 2. Data for self-determination

Desi Small-Rodriguez, Vanessa Clark & Holden Hohaia

Learning objectives:

Objective 1	Understand how Indigenous self-determination over Indigenous data are inherent; however, control over Indigenous data exists in complex data ecosystems
Objective 2	Understand that accessing and building data for self-determination may come from multiple sources and require development of bespoke kaitiaki or guardianship approaches
Objective 3	Be able to identify tools, mechanisms, and/or actions to exercise data for self-determination

Key Concepts and Definitions:

- **Indigenous self-determination**
  "Something that is asserted and acted upon, not negotiated or offered freely by the state." (Corntassel and Bryce, 2011).

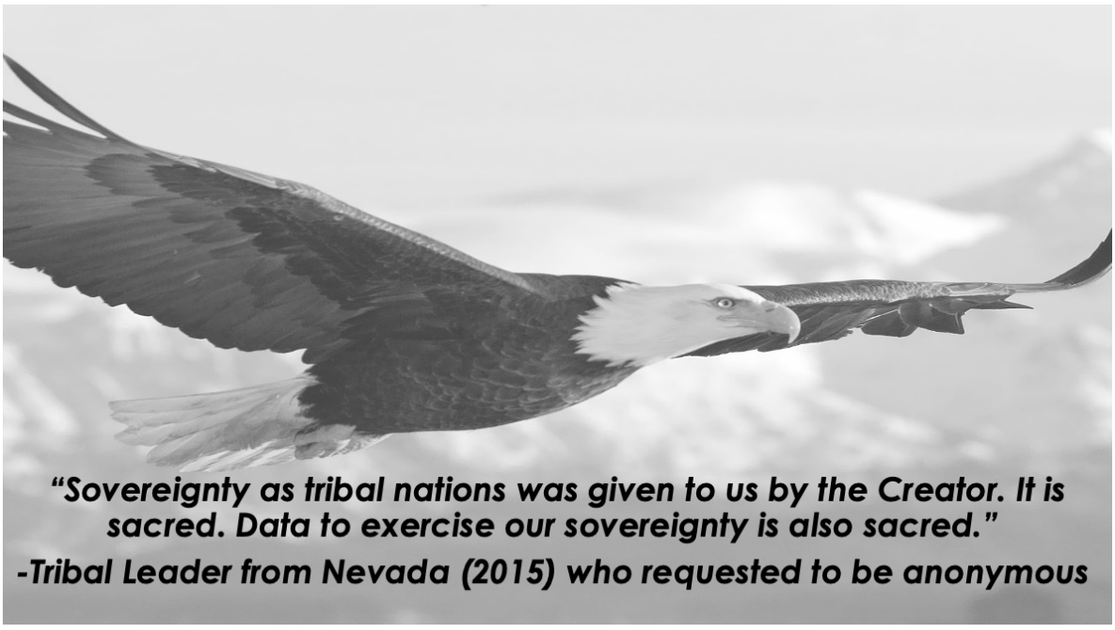
- **Data ecosystem**
  The environment in which data are collected, stored, protected, accessed, transferred, analyzed, and utilized. We note this is a simplified definition, and encourage the creation of new definitions that are specific to one's kaupapa or context.

Reference:

Corntassel, J., & Bryce, C. (2011). Practicing sustainable self-determination: Indigenous approaches to cultural restoration and revitalization. *Brown J. World Aff.*, *18*, 151.

Section 2.1. Data by us for us

Desi Small-Rodriguez



Data for self-determination refer to the data that tribal nations, iwi and Indigenous peoples need to exercise their sovereignty according to their own governance activities. These data are often described as sacred by Indigenous peoples (and as a taonga by Māori) because they are data on their most precious resources—their peoples, lands, waters, and more than human kin.

Data for self-determination are also referred to as "Data for [Indigenous] Governance" in the IDSov movement. They form part of an interrelated process whereby both **Governance of Data** and **Data for Governance** are required to move away from data dependency towards the actualization of IDSov.

This visualization can help identify where various data projects fall in terms of data for self-determination with data dependency on the bottom and data sovereignty at the top. "Us" refers to Indigenous Peoples and "them" refers to non-Indigenous people. There are many data projects across the Indigenous world whereby data are still being collected *by* non-Indigenous people *for* non-Indigenous people *about* Indigenous Peoples. Such extractive and exploitative practices must end!

Reference:

Carroll, S. R., Rodriguez-Lonebear, D. & Martinez, A. (2019). Indigenous data governance: Strategies from United States native nations. *Data Science Journal, 18*(31), 1–15. doi: https://doi.org/10.5334/dsj-2019-031 .

Data for Governance

Dimensions of Governance	Some Key Data Needs
Cultural geography and legitimacy	Social relations and territorial organisations linked to group ownership of land
Decision-making	Consensus building, free prior informed consent
Institutional bases	Standards, measures, structures, purposes, goals, policies, actions and outcomes
Strategic direction	Planning, priorities, strategies for short and long-term, risk management
Participation and voice	Group demography, participation in decision-making, elections and voting
Resource governance	Cultural, human, natural, economic, technological, financial and other resources
Governance environment	Stakeholder analyses, fiscal flows and funding
Service delivery	Group demography, needs analysis
Capacity development	Skills, expertise, knowledge, information, abilities to build governance
Governance self-evaluation	Measures of Indigenous defined governance success

Source: Smith, D (2016), Governing Data and Data for Governance: The Everyday Practice of Indigenous Sovereignty, in T. Kukutai & J. Taylor (eds), *Indigenous data sovereignty: Toward an agenda*. ANU Press

This visual identifies various types of data required by tribal nations, iwi and Indigenous peoples to drive their own governance activities.

State of Tribal Data Capacity in Indian Country

83% of the tribes who responded to the survey indicated that it is extremely important for tribes to collect or have access to data on their tribal populations for governance purposes.

Table 3: Tribal use of data on its tribal members

Response	Frequency	Percent
Complete federal grant or other required reporting	149	76%
Communicate with tribal members	135	69%
Service delivery	121	61%
Set tribal priorities and strategic goals	118	60%
Develop a budget	110	56%
Communicate with others outside the tribe	74	38%
Other	11	6%
I don't know	7	4%
Missing	5	3%

Note: number of respondents=197; number of tribes=122. Percentages are weighted for response at tribal level. Percentages total over 100 percent because respondents can select more than one response option.

National Congress of American Indians

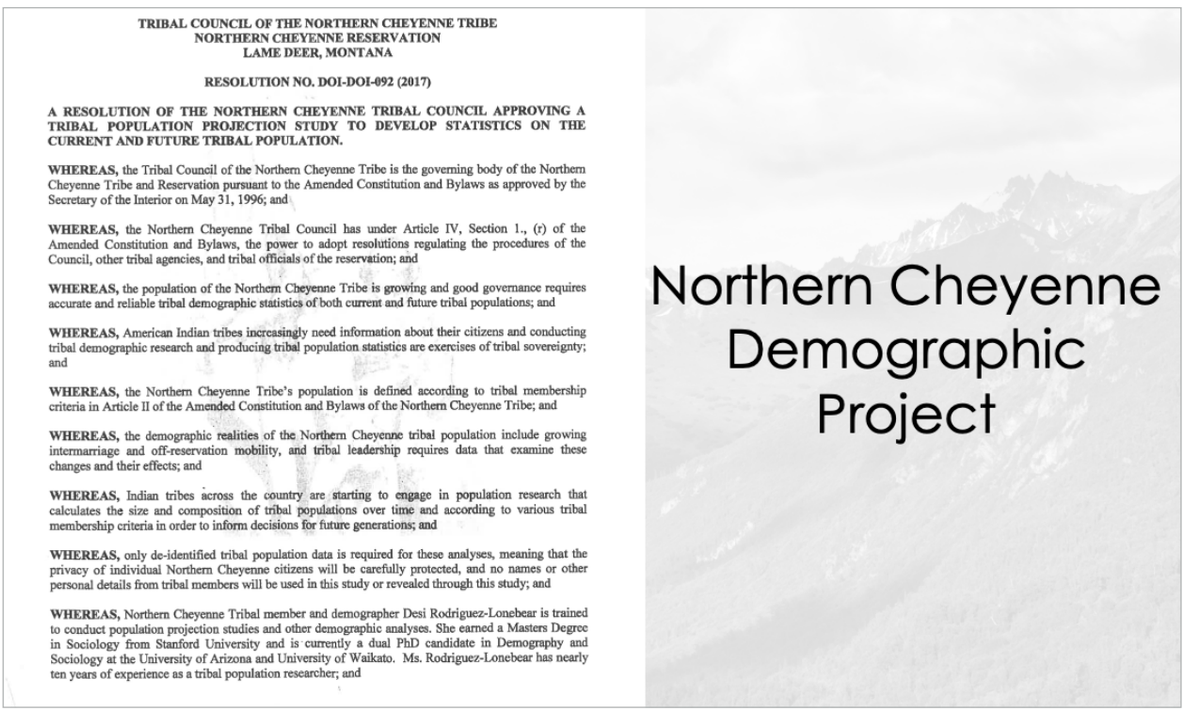
NCAI Policy Research Center (2018). The State of Tribal Data Capacity in Indian Country: Key Findings from the Survey of Tribal Data Practices. Washington, D.C.: National Congress of American Indians.

In 2018, the National Congress of American Indians (PI: Dr. Desi Small-Rodriguez) surveyed all 574 federally recognized native nations in what is currently called the USA about their tribal data practices. The survey achieved one of the highest response rates of any survey fielded to all Tribal Nations N=122 (~21%). Tribal nations were clear that data are critical to their exercise of tribal governance.

Data For Self Determination in Action: Northern Cheyenne Nation Case Study

"Tribal sovereignty is only as strong as we exercise it."
– Former Northern Cheyenne Nation President John Robinson

This next section provides examples of data for self-determination in partnership with the Northern Cheyenne Nation. The Northern Cheyenne Demographic Project provided projections for the tribal population according to various tribal enrollment changes by blood quantum. The Northern Cheyenne COVID-19 Project examines the impacts of COVID-19 on the tribal population comparing indicators of health established by the State of Montana (e.g., mortality) vs. those established by the tribe (e.g., language loss).



TRIBAL COUNCIL OF THE NORTHERN CHEYENNE TRIBE
NORTHERN CHEYENNE RESERVATION
LAME DEER, MONTANA

RESOLUTION NO. DOI-DOI-092 (2017)

A RESOLUTION OF THE NORTHERN CHEYENNE TRIBAL COUNCIL APPROVING A TRIBAL POPULATION PROJECTION STUDY TO DEVELOP STATISTICS ON THE CURRENT AND FUTURE TRIBAL POPULATION.

WHEREAS, the Tribal Council of the Northern Cheyenne Tribe is the governing body of the Northern Cheyenne Tribe and Reservation pursuant to the Amended Constitution and Bylaws as approved by the Secretary of the Interior on May 31, 1996; and

WHEREAS, the Northern Cheyenne Tribal Council has under Article IV, Section 1., (r) of the Amended Constitution and Bylaws, the power to adopt resolutions regulating the procedures of the Council, other tribal agencies, and tribal officials of the reservation; and

WHEREAS, the population of the Northern Cheyenne Tribe is growing and good governance requires accurate and reliable tribal demographic statistics of both current and future tribal populations; and

WHEREAS, American Indian tribes increasingly need information about their citizens and conducting tribal demographic research and producing tribal population statistics are exercises of tribal sovereignty; and

WHEREAS, the Northern Cheyenne Tribe's population is defined according to tribal membership criteria in Article II of the Amended Constitution and Bylaws of the Northern Cheyenne Tribe; and

WHEREAS, the demographic realities of the Northern Cheyenne tribal population include growing intermarriage and off-reservation mobility, and tribal leadership requires data that examine these changes and their effects; and

WHEREAS, Indian tribes across the country are starting to engage in population research that calculates the size and composition of tribal populations over time and according to various tribal membership criteria in order to inform decisions for future generations; and

WHEREAS, only de-identified tribal population data is required for these analyses, meaning that the privacy of individual Northern Cheyenne citizens will be carefully protected, and no names or other personal details from tribal members will be used in this study or revealed through this study; and

WHEREAS, Northern Cheyenne Tribal member and demographer Desi Rodriguez-Lonebear is trained to conduct population projection studies and other demographic analyses. She earned a Masters Degree in Sociology from Stanford University and is currently a dual PhD candidate in Demography and Sociology at the University of Arizona and University of Waikato. Ms. Rodriguez-Lonebear has nearly ten years of experience as a tribal population researcher; and

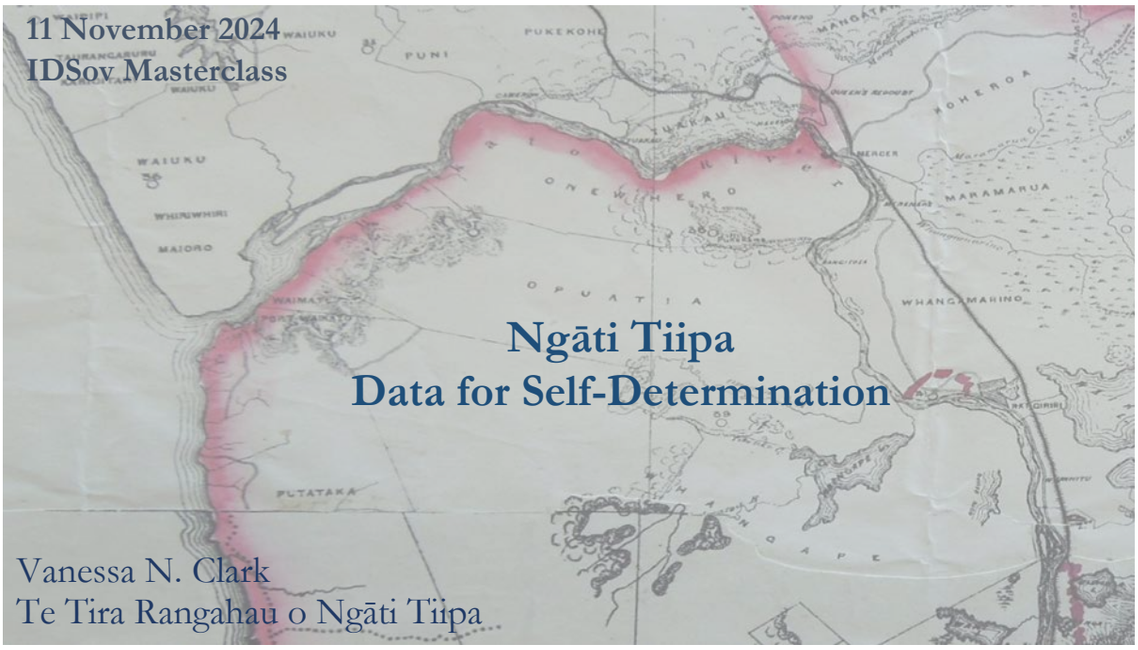
- COVID-19- associated mortality was leading cause of death among AI Montana residents in 2020.
- Between March-October 2020, AIs 19% of Montana's COVID-19 cases and 32% of the state's COVID-19 deaths, but ~7% of pop.

State Data

Tribal Data

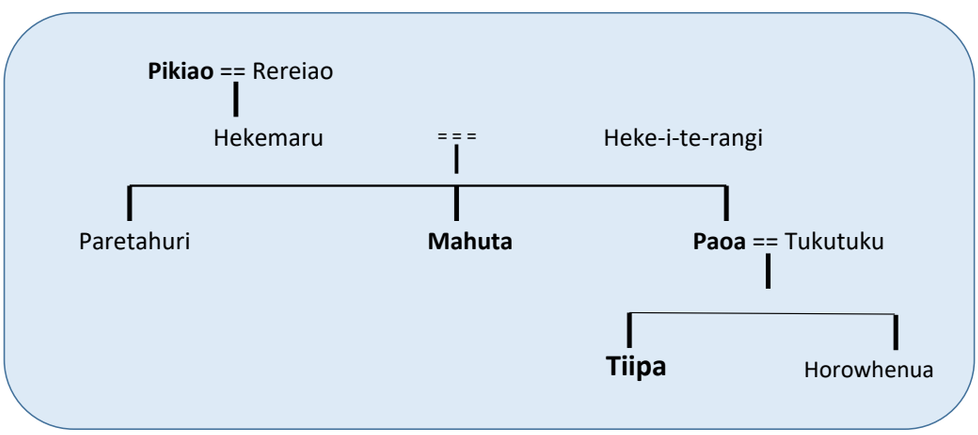
Section 2.2. Ngāti Tiipa data for self-determination

Vanessa Clark



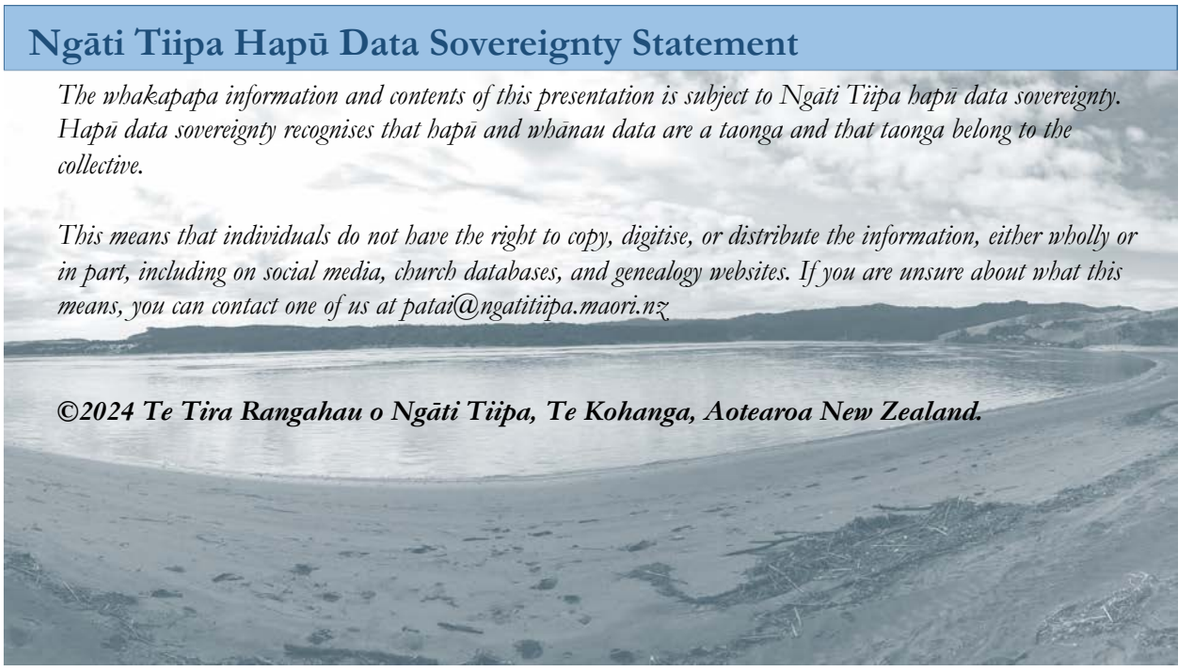
This case study explores the seven-year journey of Ngāti Tiipa – one of the 33 hapū (sub-tribes) and iwi (tribes) of the Waikato-Tainui confederation – who have been on a mission to reclaim sovereignty over our most precious data—that is, data relating to our whakapapa (ancestral genealogies) and whenua (territories).

Eponymous Ancestor Tiipa



This slide shows some of the whakapapa that connects Ngāti Tiipa to other iwi and hapū from the Tainui and Te Arawa waka.

Ngāti Tiipa Hapū Data Sovereignty Statement



The whakapapa information and contents of this presentation is subject to Ngāti Tiipa hapū data sovereignty. Hapū data sovereignty recognises that hapū and whānau data are a taonga and that taonga belong to the collective.

This means that individuals do not have the right to copy, digitise, or distribute the information, either wholly or in part, including on social media, church databases, and genealogy websites. If you are unsure about what this means, you can contact one of us at patai@ngatitiipa.maori.nz

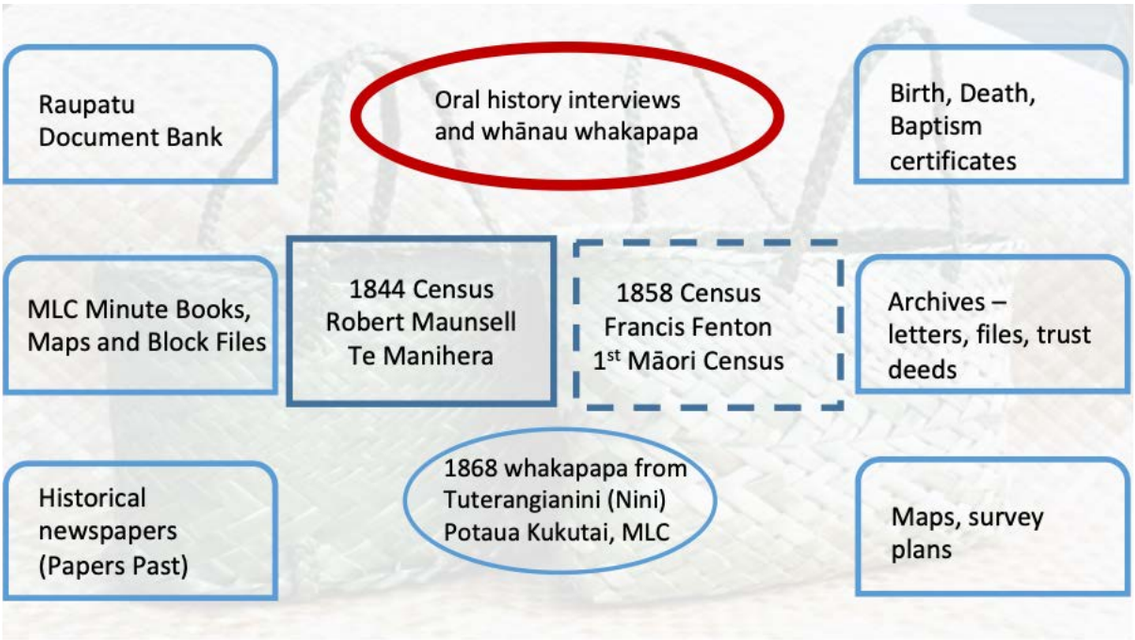
Ngāti Tiipa have developed our own hapū data sovereignty statement that accompanies the dissemination of any of our whakapapa and whenua information that we hold.

Ngāti Tiipa - Context Setting for the Rangahau

- Started out as 3-year research project (Counting our Tūpuna) funded by Marsden Fund in 2018.
- Bringing together whakapapa (genealogical) information on our Ngāti Tiipa tūpuna from c.1800 onwards and data on our whenua (lands) to create Ngāti Tiipa whakapapa and whenua database(s)
- The databases are owned and controlled by Ngāti Tiipa as a taonga for current and future generations – tikanga for access and use have been developed by Ngāti Tiipa
- We are not an entity; this is by whānau for our whānau, marae and hapū
- Capacity building is a key part of the project – Ngāti Tiipa kaiarahi, haukāinga kaitiaki, Masters student and research assistants at Te Whare Wānanga o Waikato

All of our ranghau is led by and for the whānau (families) of Ngāti Tiipa. To date we have undertaken three research projects. The first was funded by the Marsden Fund (PI Tahu Kukutai, 2018-2021) and which lay the foundations for developing our own hapū research team Te Tira Rangahau o Ngāti Tiipa. Subsequent projects/case studies were funded by A Better Start National Science Challenge and the Ministry of Business Innovation and Employment (MBIE) Endeavour Fund ('Tikanga in Technology').

The key focus of the initial Marsden project was to identify all Ngāti Tiipa tūpuna (ancestors) who lived between 1750 and 1900. It was a mixed methods project with archival, oral history, statistical and technological components.

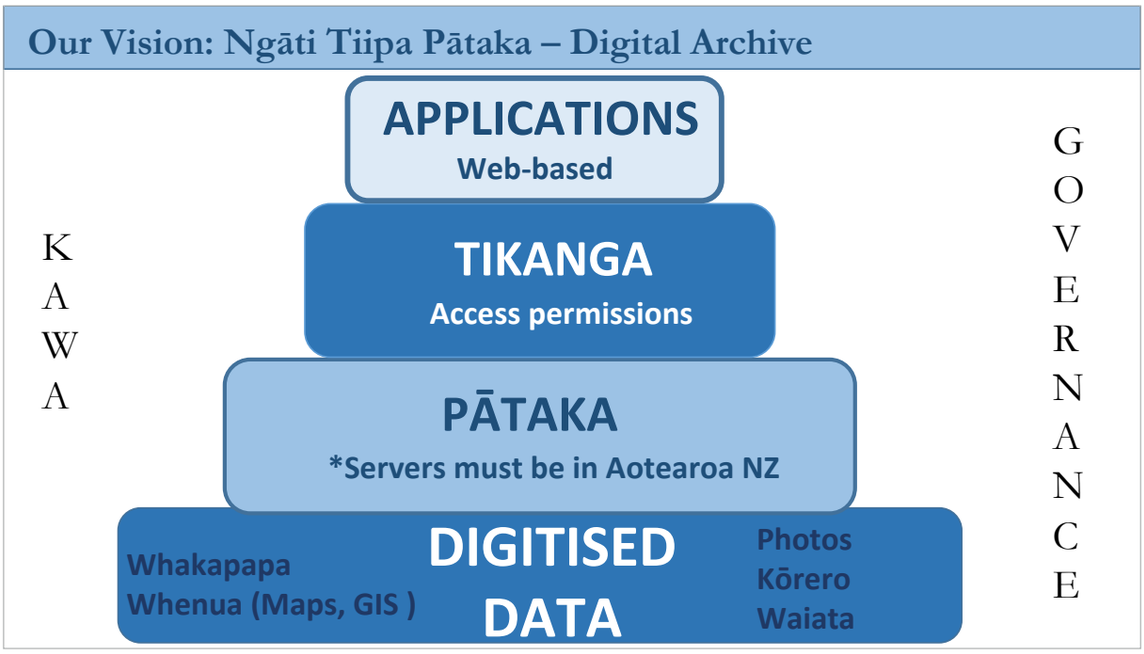


Many different kinds of information sources were used to identify our tūpuna over the 150 year focal period. This slide shows some of the main sources.

Data repatriation, integration and repurposing

1. Primary source material in public repositories
2. Located and digitised it (e.g., lists of owners in the Māori Land Court) OR if the material was already digitised, we obtained a copy
3. Labelled all files to retain provenance information
4. Transcribed or indexed key terms. Still much to do here…
5. Where the source contained direct whakapapa information (e.g., name of tūpuna, residence, date of death) we added this to the whakapapa database – initially using an off-the-shelf package (roots magic) but developed our own database based on open source (webtrees) adhering to Ngāti Tiipa tikanga

Our team of whānau researchers spent many hours identifying and digitising information in public repositories. The information that we gathered was shared with whānau at a series of marae-based wānanga (gatherings) and pānui (notices) were sent out well in advance so that whānau could plan to attend. Anyone with a whakapapa connection to Ngāti Tiipa was welcome.



This technology 'keke' (cake) was developed by our hapū technologist, Vanessa Clark. It sets out the key elements of our technology approach in a way that is intuitive and easy to understand. Wānanga were also held with whānau to identify the kawa and tikanga (customary protocols) requirements.

Te Tira Rangahau o Ngāti Tiipa

Although the first project was mainly focused on data repatriation, digitisation and integration, our whānau told us that they wanted something tangible to hold in their hands. So we worked together to publish our Ngāti Tiipa taonga book, *Te Pūahatanga o Ngāti Tiipa.* The book is closed access and limited edition. It contains biographies of prominent tūpuna written by their direct descendants, as well as whakapapa, kōrero tuku iho (oral traditions), photographs and other information specific to Ngāti Tiipa. There are also blank sections for whānau to record their own whakapapa and moemoeā (dreams and aspirations) for their mokopuna (grandchildren). Only 500 copies of the book were published and each book is individually numbered, with the whānau kaitiaki (guardian) recorded in a register.

This picture of Te Tira Rangahau o Ngāti Tiipa was taken at the launch of the book at Te Kumi pā (Te Kotahitanga marae) in 2022.

Tiipa tangata, Tiipa whenua, Tiipa te iwi.

Puea te pūtaiao, te katonga whenua, te weherua atu o rātou te āka te rēinga.

Hīria he oranga hei painga mo te iwi
mo ake tonu atu.

Tiipa the person, the land, the people.
Let our science and ways of knowing flourish, about our lands that were severed
from us, and our ancestors who have departed.
So that our collective wellbeing may likewise flourish
and endure.

Nā Ngāti Tiipa koroheke Karu Kukutai

References:

Clark, V., Newbold, E., Kani, H., & Kukutai, T. (2024). *E Kore au E Ngaro: How tikanga, kawa and kōrero tuku iho inform kaitiaki approaches to whakapapa and whenua data of Ngāti Tiipa.* Report commissioned by A Better Start National Science Challenge. (Available upon request).

Newbold, E., Clark, V., Kani, H., & Kukutai, T. (2022). *E kore au e ngaro: A Ngāti Tiipa framework for hapū and whānau whakapapa research.* Report commissioned by A Better Start National Science Challenge. (Available upon request)

Kukutai, T., Whitehead, J., & Kani, H. (2022). Tracing Opuatia: Repatriating and repurposing colonial land data. *New Zealand Geographer, 78*(2), 134-146. https://doi.org/10.1111/nzg.12344

Kukutai, T., Mahuika, N., Kani, H., Ewe, D., & Kukutai, K. (2021). Survivance as narrative identity: Voices from a Ngāti Tiipa oral history project. *MAI Journal.* doi 10.20507/MAIJournal.2020.9.3.11

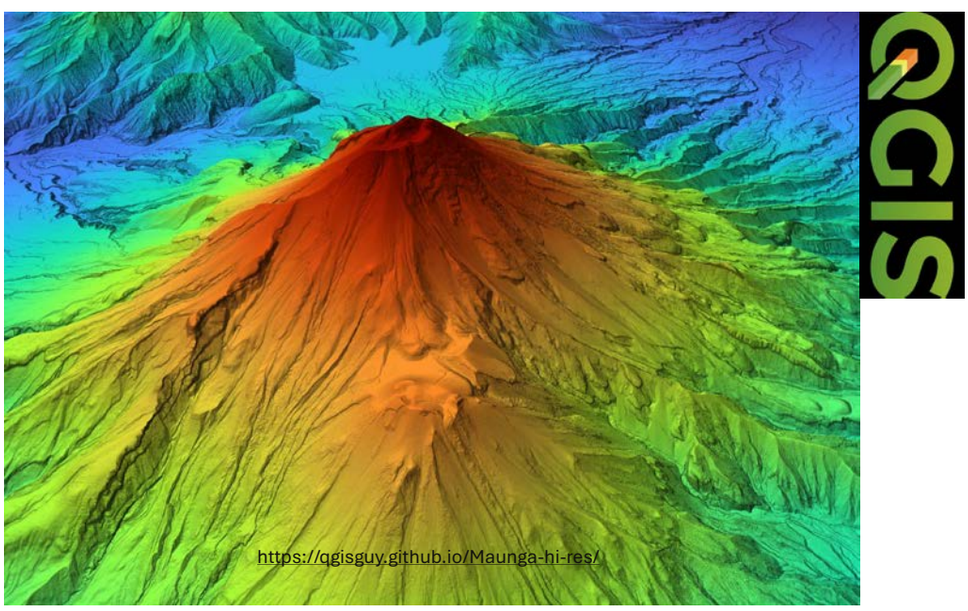
Section 2.3. QGIS as a tool for data for self-determination

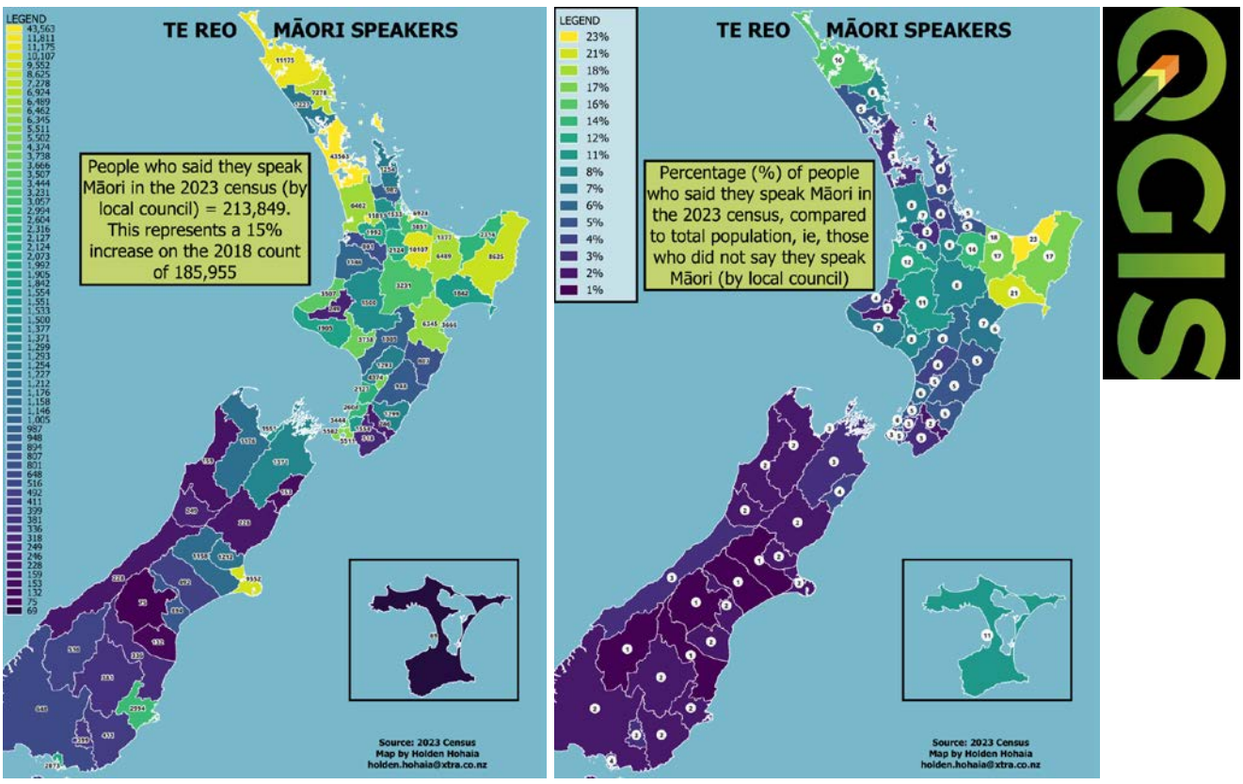
Holden Hohaia

This case study looks at how QGIS, an open-source mapping software, can be used to create maps that help hapū and iwi be better kaitiaki, and to tell their own Indigenous stories through data.

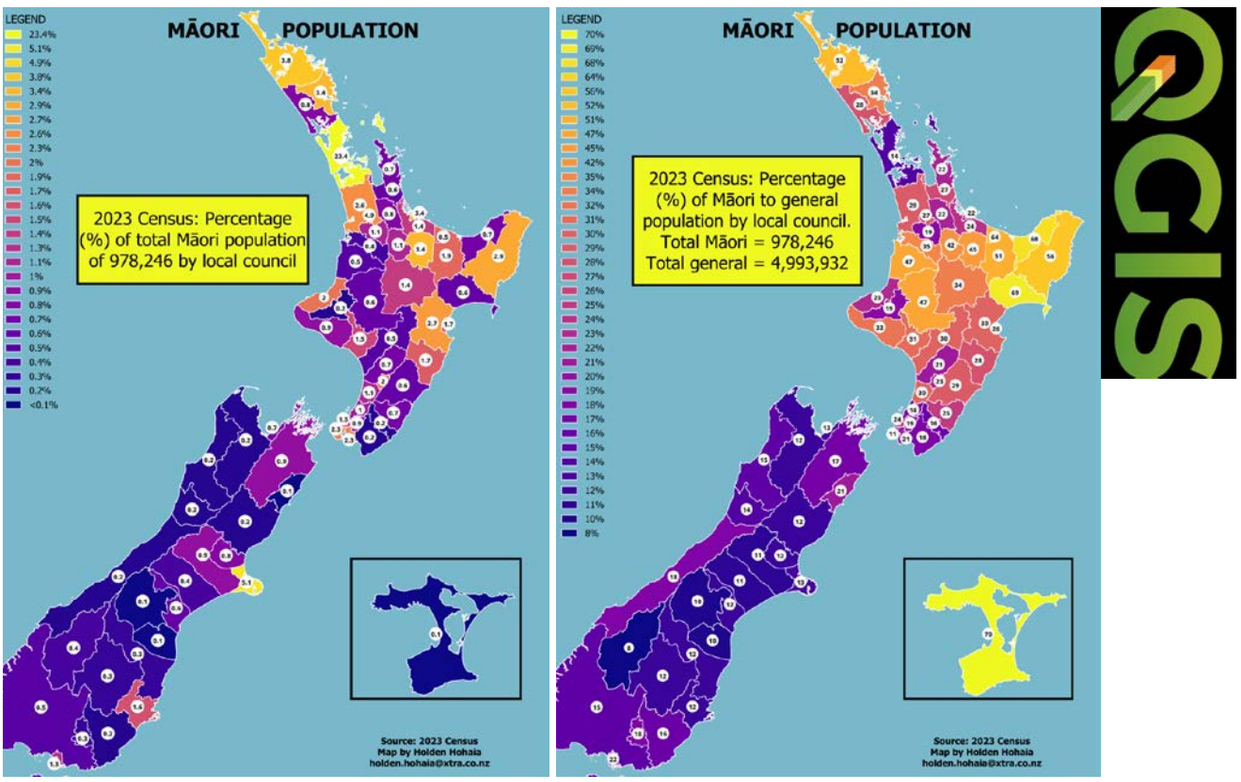
A practical, free & open-source tool that gives you control over your own data

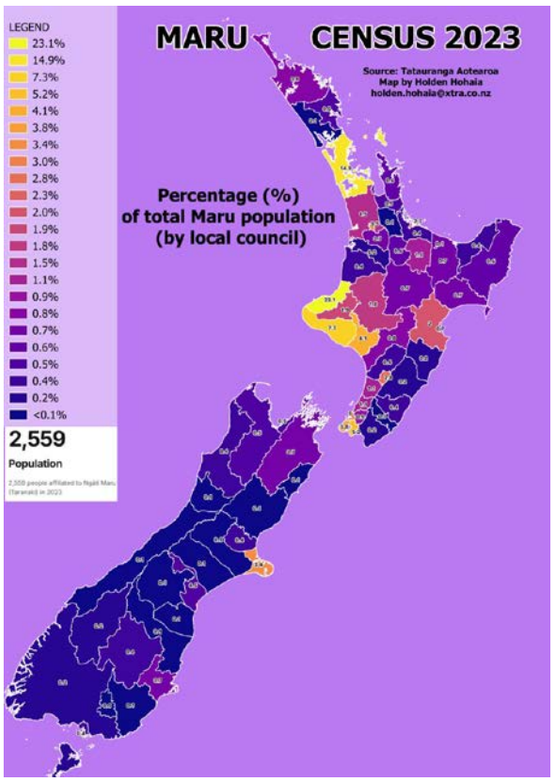
Some examples:

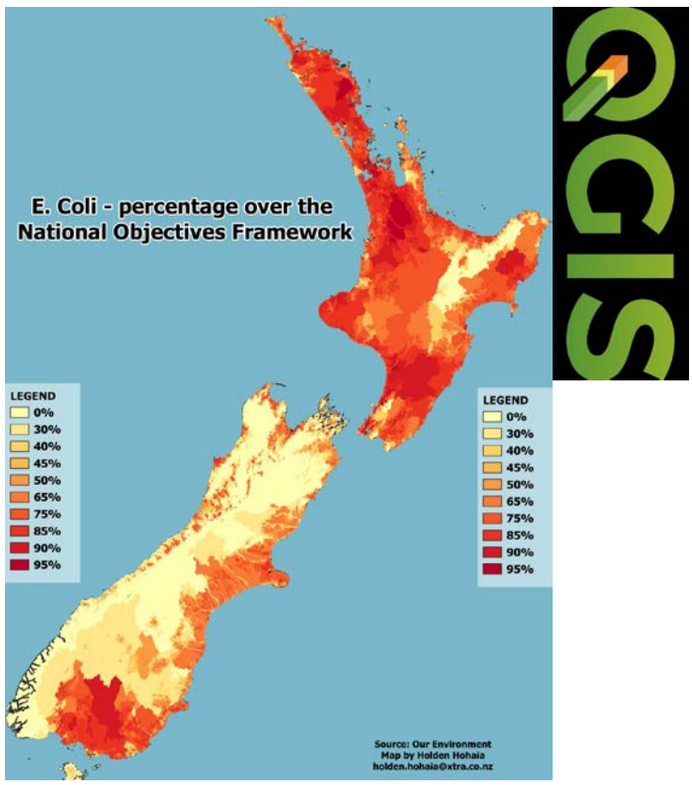




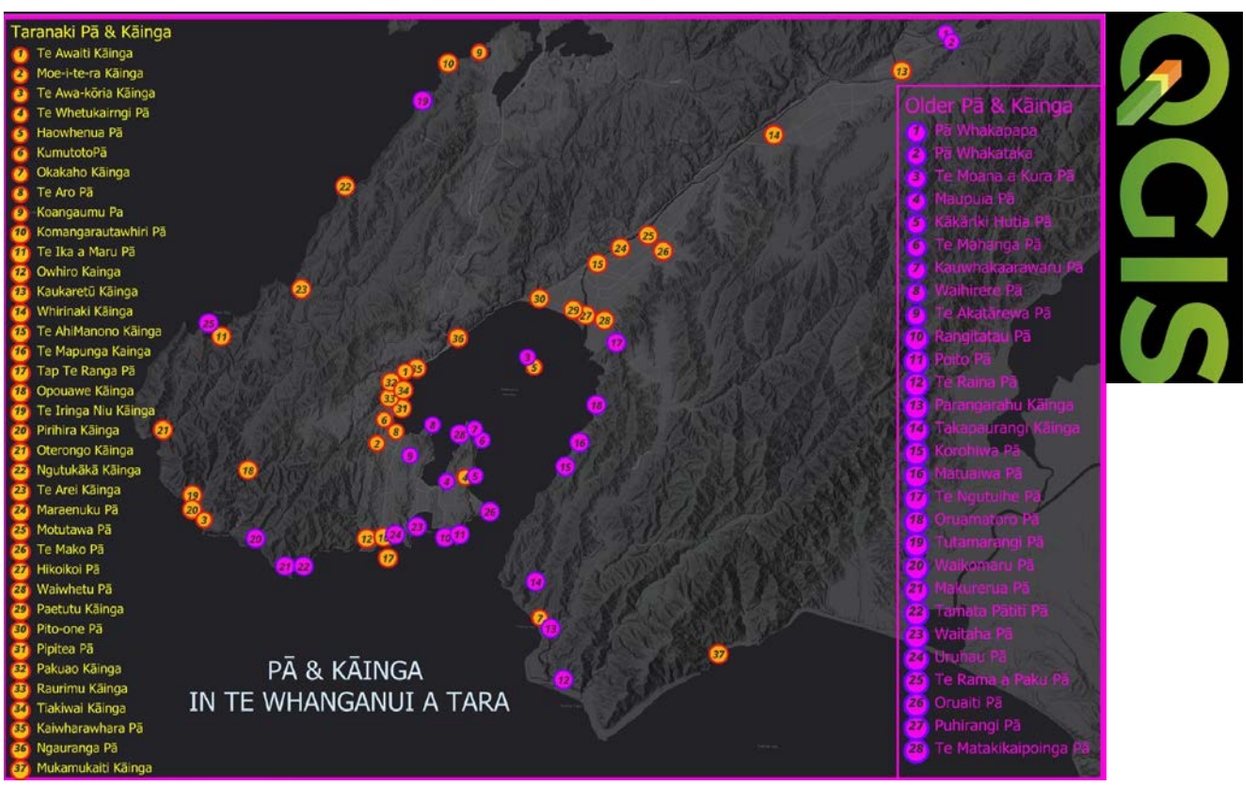
These maps capture the spatial variation in te reo Māori speakers and Māori population (Māori descent) across Aotearoa. The different segments represent Territorial Authorities.

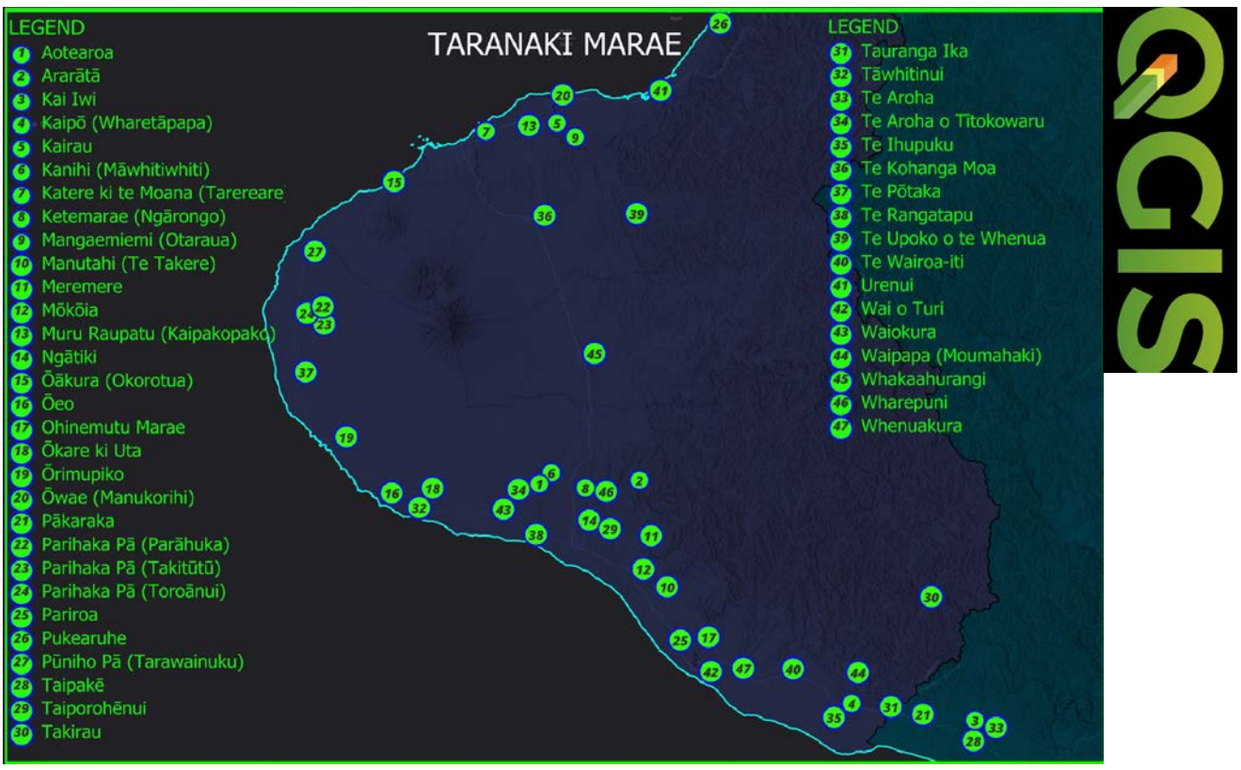


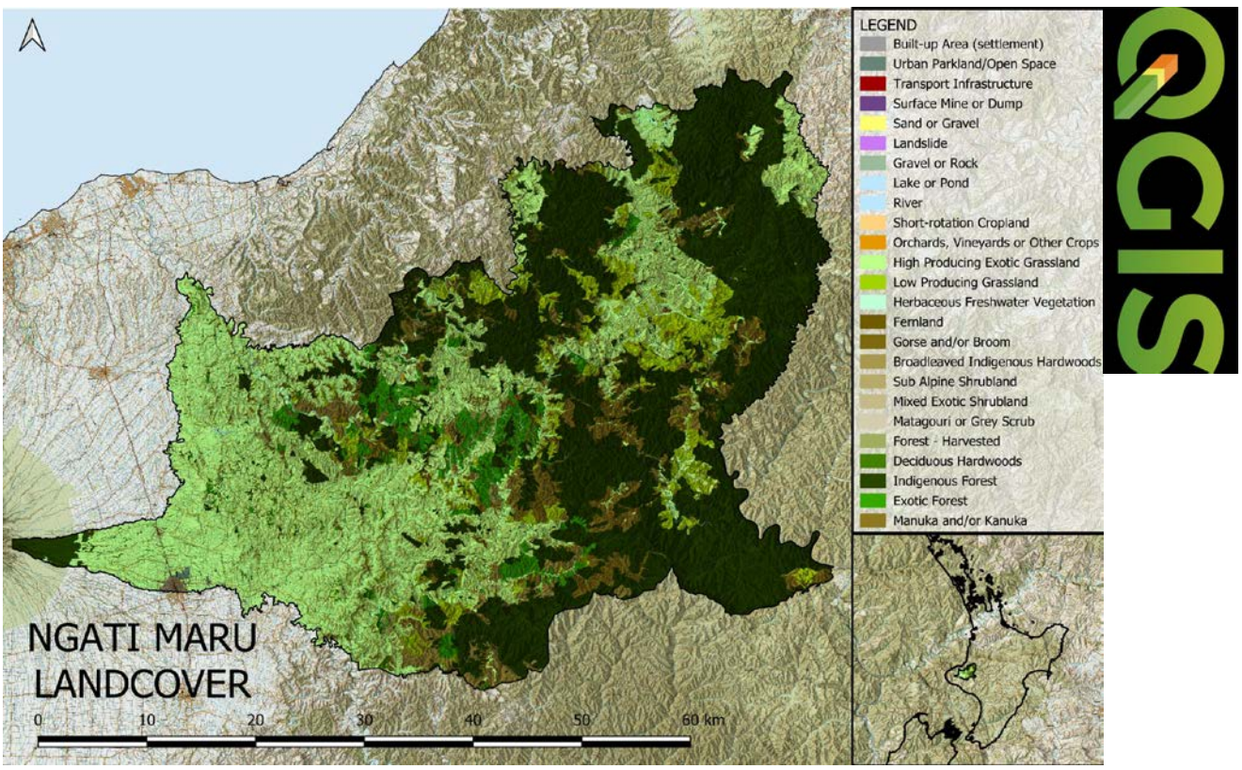




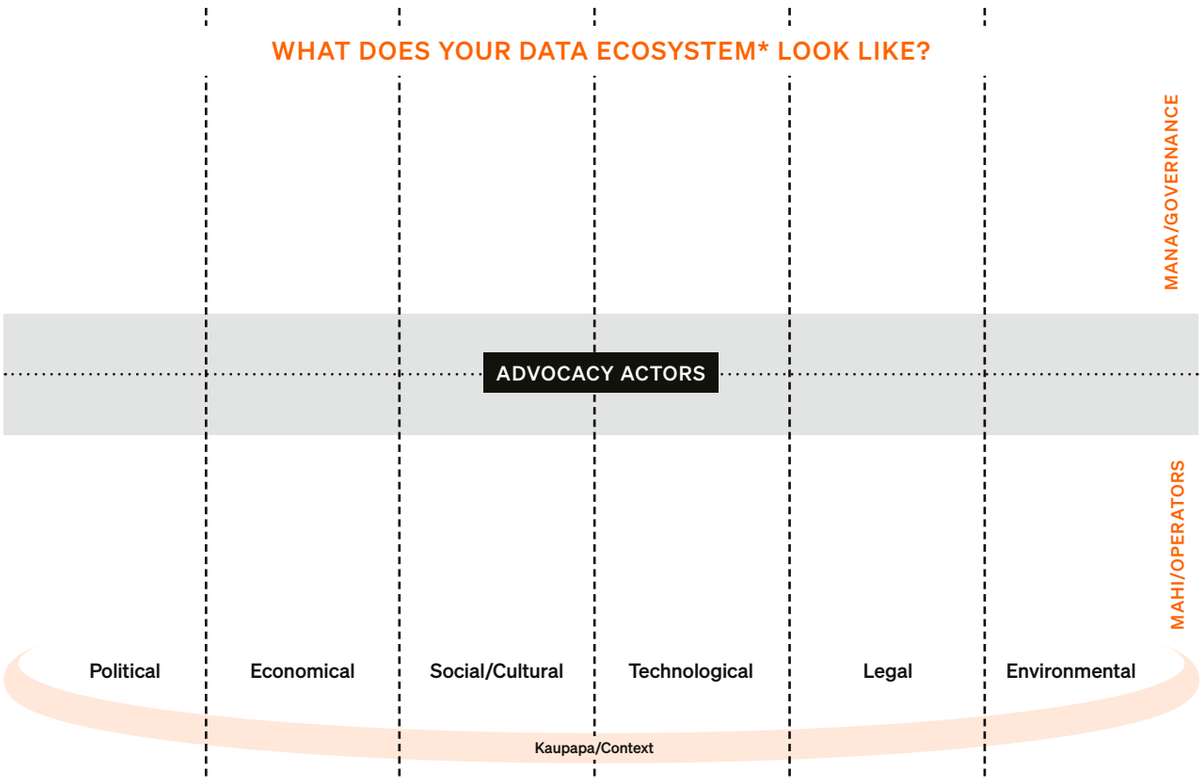
These maps are specific to Ngāti Maru (https://maru.nz/)







Activity 1: Use this chart to plot the different data actors for your kaupapa or context. Identify data actors across the political, economic, sociological, technological, legal and environmental dimensions that hold Indigenous data. Locate data actors based on the level of control over the data, i.e., who has mana or governance authority over the data? Identify advocacy data actors that might assist in navigating the spaces. What insights or 'aha' moments does this reveal?



Resources

Watch 

Small-Rodriguez, D. (2019). Growing data warriors – by us, for us. REDTalks. https://www.youtube.com/watch?v=GtPb3pFOqwA

Small-Rodriguez, D. (2024). Indigenous data science. https://www.youtube.com/watch?v=Tq4U4RzYSIo

Russo Carroll, S. (2024). Indigenous peoples breathing data back. TEDx Talks. https://www.youtube.com/watch?v=jPS_3mZXWXw

Read 

Carroll, S. R., Rodriguez-Lonebear, D. & Martinez, A. (2019). Indigenous data governance: Strategies from United States native nations. *Data Science Journal, 18*(31), 1–15. doi: https://doi.org/10.5334/dsj-2019-031

Kukutai, T., Mahuika, N., Kani, H., Ewe, D., & Kukutai, K. 2021. Survivance as narrative identity: Voices from a Ngāti Tiipa oral history project. *MAI Journal.* doi 10.20507/MAIJournal.2020.9.3.11

National Congress of American Indians (2018). The state of tribal data capacity in Indian Country: Key findings from the survey of tribal data practices. Available at: https://nativephilanthropy.issuelab.org/resources/34730/34730.pdf

National Congress of American Indians (2018). Support of US Indigenous Data Sovereignty and inclusion of tribes in the development of tribal data governance principles. Available at: https://archive.ncai.org/resources/resolutions/support-of-us-indigenous-data-sovereignty-and-inclusion-of-tribes-in-the-development-of-tribal-data

Explore 

Native Biodata Consortium - the first 501(c)(3) nonprofit research institute led by Indigenous scientists and tribal members in the United States. https://nativebio.org/

Section 3. Data governance models: From local to global

Tahu Kukutai, Jacob Prehn & Cheryl Ellenwood

Learning objectives:

Objective 1	Understand how IDGov is being applied in different contexts
Objective 2	Be able to distinguish between 'governance of data' and 'data for governance'
Objective 3	Have tools to assess IDGov implementation and governance

Key Concepts and Definitions:

- **Māori data**
  Maori data refers broadly to digital or digitisable data, information or knowledge (including mātauranga Māori) that is about, from or connected to Māori. It includes data about population, place, culture and environment.

- **Māori data governance**
  The principes, structures, accountability mechanisms, legal instruments and policies through which Māori exercise control over Māori data.

- **Australian Public Service – Governance of Indigenous Data Framework**
  A Framework 'co-designed' by the Australian Federal Government and Aboriginal and Torres Strait Islander people. This is not IDSov or IDGov but framed as a step towards more equitable data relations.

- **Maturity model**
  A maturity model is a tool to assess, measure, and improve organisational systems and processes.

- **FAIR Data Maturity Model**
  FAIR stands for Findability, Accessibility, Interoperability and Reusability. The FAIR data maturity model can be used by stakeholders, including researchers, data stewards, policy makers and funding agencies, to assess the FAIRness of data and identify aspects that can be improved upon to increase the potential for reuse of research data.

➢ CARE Principles for Indigenous Data Governance

The CARE principles for Indigenous data governance stand for Collective benefit, Authority to control, Responsibility and Ethics.

➢ CARE Data Maturity Model

The CARE maturity model is a tool that can be used to assess, measure and improve upon the implementation of the CARE Principles.

References:

Kukutai, T., Campbell-Kamariera, K., Mead, A., Mikaere, K., Moses, C., Whitehead, J. & Cormack, D. (2023). *Māori Data Governance Model*. Te Kāhui Raraunga.

Carroll, S., Garba, I., Figueroa-Rodríguez, O., Holbrook, J., Lovett, R., Materechera, S., ... & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data science journal*, *19*. Available at: https://www.research.ed.ac.uk/en/publications/the-care-principles-for-indigenous-data-governance

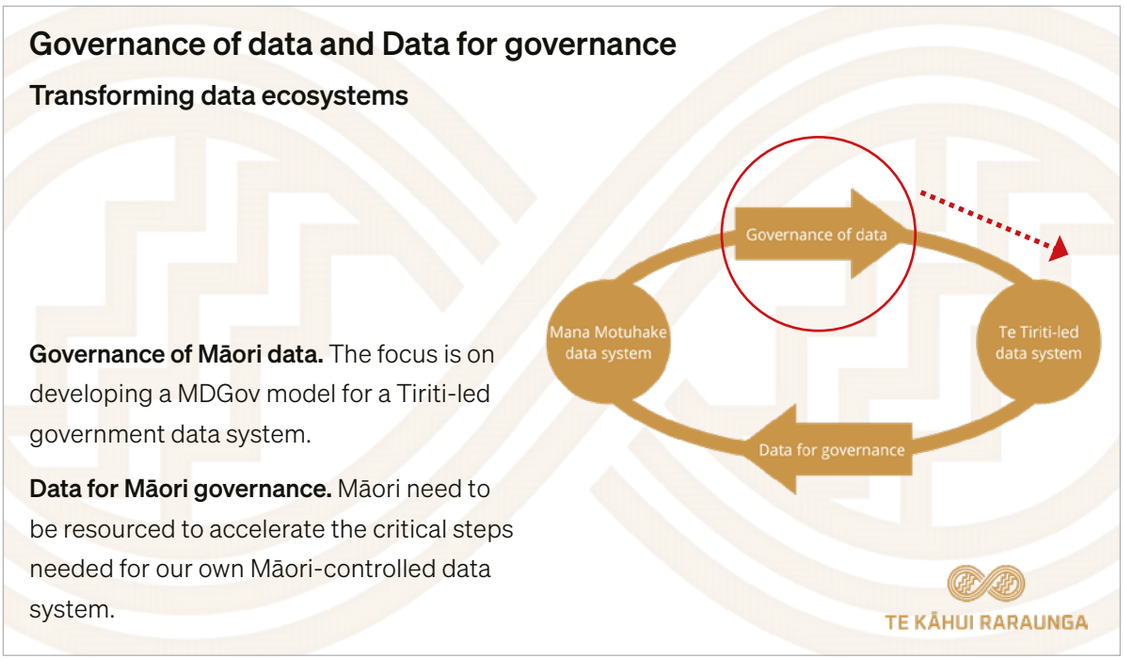
Commonwealth of Australia (2024). *Governance of Indigenous data framework*. Australian Public Service. Available at: https://www.niaa.gov.au/sites/default/files/documents/2024-05/framework-governance-indigenous-data.pdf

Section 3.1. Māori Data Governance Model

Tahu Kukutai

Aotearoa NZ context

- 1840 Te Tiriti o Waitangi: Art. 2 – '... te tino rangatiratanga o o ratou ... taonga katoa.' [Māori retain chiefly authority over all their taonga or treasured possessions]. Data is a taonga.
- Māori comprise 20% of the total Aotearoa (New Zealand) population.
- Ethnicity data is collected across most government data collections; some also collect Māori descent/ancestry and iwi (tribe).
- Māori are disproportionately impacted by ethnic/racial inequities that are systemic, embedded and intergenerational.
- Aotearoa is one of the most advanced digital nations in the world, especially with regard to statistical data linkage.
- The National Iwi Chairs Forum (tribal leaders) has a dedicated data group called the Data Iwi Leaders Group (Data ILG) operating as Te Kāhui Raraunga Charitable Trust.

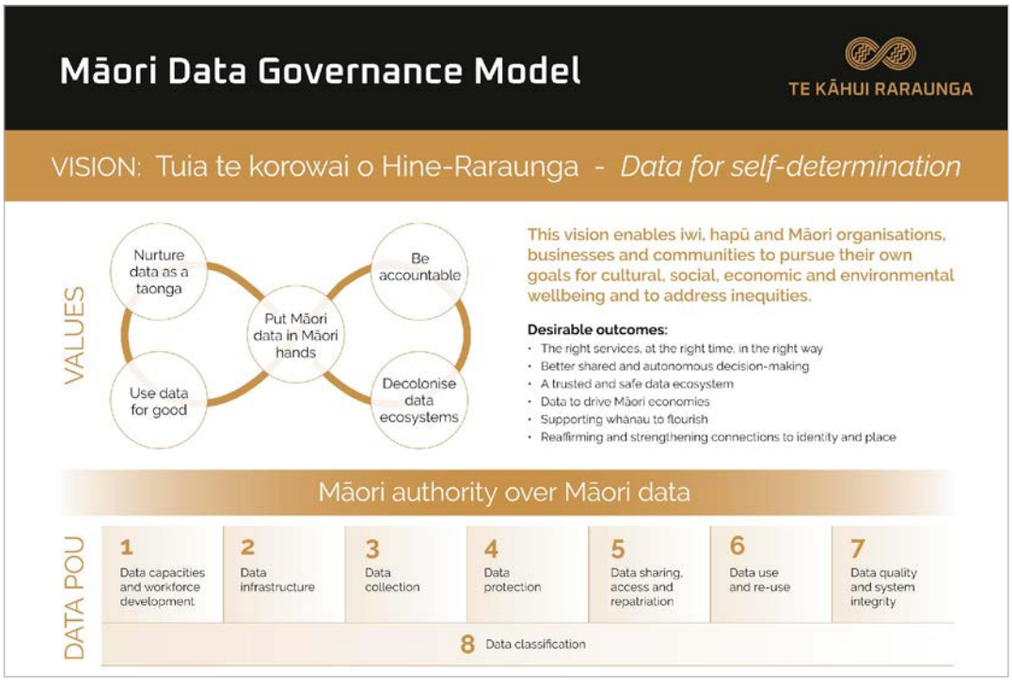


Source: Kukutai et al., 2023. Used with the permission of Te Kāhui Raraunga/Data ILG

Several of the talks in this Masterclass draw a distinction between the **Governance of data** and **Data for governance.** The Māori Data Governance Model (the Model) belongs in the former category – its focus is on the governance of Māori data held by public service agencies. The Model was undertaken as part of the 'Mana Ōrite' Tiriti relationship agreement between the Data ILG and Stats NZ. Māori data experts designed the Model with the expectation that experts from within the public service would design their own implementation plan. The Model was informed by extensive engagement with Te Ao Māori in a series of wānanga held by Te Kāhui Raraunga. The First Nations Data Governance Strategy team in Canada also generously shared their insights and experiences with the team developing the Model.

Other sections in this workbook discuss Data for governance initiatives that are directly under mana Māori (Māori authority).

Trustees and kaimahi (staff) of Te Kāhui Raraunga, the kaitiaki (guardian) of the Model. Used with the permission of Te Kāhui Raraunga/Data ILG.



Source: Kukutai et al., 2023. Used with the permission of Te Kāhui Raraunga/Data ILG

This is a high-level snapshot of the Model, without all the detail (the full report is 73 pages). The overarching vision 'Tuia Te Korowai o Hine-Raraunga' was gifted by iwi leaders who also provided an accompanying cultural narrative and whakapapa.

The **Desirable Outcomes** focus on collective benefits and system impacts that are meaningful and achievable including supporting whānau to flourish and reaffirming and strengthening connections to identity, place and te reo Māori.

On the left five **Values** that are embedded throughout the Model. The values are intended to work together to guide data decision making and practices.

The **Data Pou** are the building blocks that represent priority areas of action and against which government agencies can assess their Maori data governance maturity level. Each pou defines a critical area of data governance and specifies a set of directives about actions, processes and activities that should be undertaken by agencies – which is the practical 'how to' guidance. The pou include:

- data capacities and workforce development
- data infrastructure
- data collection
- data protection
- data sharing, access and repatriation
- data use and reuse
- data quality and system integrity
- data classification.

Pou 2: Data Infrastructure

Enabling data consumption, storage, and sharing

To serve communities, data infrastructure needs to be flexible, scalable, and interoperable, and offer technology options that power choices close to where decisions are made, including outside of government.

Key priorities and actions:

- **2.1 Data infrastructure that works for Māori** – Share decision-making on data infrastructure with Māori. This includes shared decision-making at a policy-setting level and a system-level.
- **2.2 Go DaD (Distributed and Decentralised)** – Provide resources, equitable funding, and support for the development of Mana Motuhake systems of distributed and decentralised data infrastructure for Māori.
- **2.3 Sustainable and future focused** – Invest in green cloud architecture and Māori-owned cloud providers and data warehouses.

Source: Kukutai et al., 2023. Used with the permission of Te Kāhui Raraunga/Data ILG

This slide summarises the purpose, priorities and actions for Pou 2 Data Infrastructure (for a deeper dive into sovereign iwi data infrastructure, see Section 4 of the workbook). In the Model each priority has a dedicated section, with real life examples of poor and good Māori data governance practices.

References:

Kukutai, T., Campbell-Kamariera, K., Mead, A., Mikaere, K., Moses, C., Whitehead, J. & Cormack, D. (2023). *Māori Data Governance Model*. Te Kāhui Raraunga. https://www.kahuiraraunga.io/maoridatagovernance

Te Kāhui Raraunga. (2021b). *Tawhiti nuku: Māori data governance co-design outcomes report*. https://www.kahuiraraunga.io/maoridatagovernance

Applying the Model – NZCOVID Tracer App

- As a 'next step' we developed a tool to assess the implementation of the Model The project was funded by the Ministry of Health.
- We used the tool to do a retrospective assessment of the NZ Covid Tracer App.
- The tool has 24 criteria drawn from the 8 data pou of the Model. Each criteria was assessed as 'met', 'partially met' or 'not met'.
- We found the NZCTA met 7 criteria, partially met 7 and failed 9. (1 was N/A)
- The tool can be used across a wide range of data contexts.

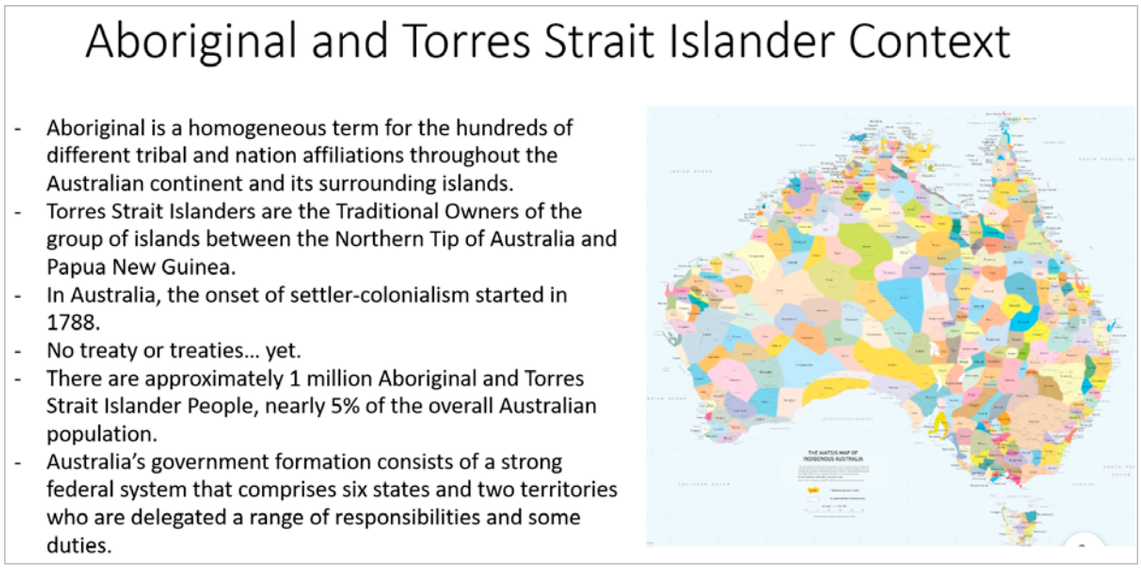
To provide an example of how to use and assess the Model, a team of researchers developed a Māori data governance assessment tool and used it to retrospectively assess the NZ Covid Tracer App. The tool can be found in the appendix of the open access paper, details below.

Reference:

Sterling, R., Kukutai, T., Chambers, T., & Chen, A. 2024. A Māori data governance assessment of the NZ Covid Tracer App. *Discover Social Science and Health.* https://doi.org/10.1007/s44155-024-00092-2

Section 3.2. Framework for Governance of Indigenous Data (Australia)

Jacob Prehn



Aboriginal is a homogeneous term for the hundreds of different tribal and nation affiliations throughout the Australian continent and its surrounding islands.

Torres Strait Islanders are the Traditional Owners of the group of islands between the Northern Tip of Australia and Papua New Guinea.

In Australia, the onset of settler-colonialism started in 1788.

The continent of Australia was claimed using the legal doctrine of terra nullius, meaning 'no man's land or land belonging to no one'.

No historical or contemporary treaties have been signed with Aboriginal and/or Torres Strait Islander people and the Crown, the former colonies of Australia, or the modern Australian nation-state.

No treaty or treaties... yet.

There are approximately 1 million Aboriginal and Torres Strait Islander People, nearly 5% of the Australian population.

Australia's government is based on a strong federal system, which comprises six states and two territories, each delegated a range of responsibilities and some duties.

Governance of Indigenous Data Framework (2024)

- Federal government's first major step towards IDSov and IDGov.
- This Framework will change how the federal government uses Indigenous data and goes across all 15/16 federal departments, and agencies like ARC, NHMRC, AIHW, etc.
- The framework was published in May 2024.
- Started in August 2021, led by the National Indigenous Australians Agency (NIAA).
- Used a co-design process between 2022 and 2023.
- I co-chaired the co-design working group.

Guideline 1: Partner with Aboriginal and Torres Strait Islander people

The Australian Federal Government's Framework for Governance of Indigenous Data (the Framework) is positioned as a first step towards greater data equity and the concepts of IDSov and IDGov.

This Framework will change how the Federal government uses Indigenous data. It applies to all 15/16 Federal government departments and agencies, including key entities like the Australian Research Council, the National Health and Medical Research Council, the Australian Institute of Health and Welfare, the Federal Department of Health, etc.

The Framework was published in May 2024. Conversations about it started as early as 2020, but the National Indigenous Australians Agency (NIAA) officially led the work in August 2021.

The Framework used a 'co-design' approach between the 2022 and 2023. There was roughly an equal amount of non-Indigenous employees from the Australian Public Service.

Dr Jacob Prehn co-chaired the co-design working group with Dr Jess Hartmann, a non-Indigenous colleague from the National Indigenous Australians Agency.

Governance of Indigenous Data Framework - 'Co-design' process

Group Numbers:

Group	Total	Non-Indigenous	Indigenous
DSDG Sub-Committee (APS Members)	16	15	1
DSDG Sub-Committee (Non-APS/Indigenous Members)	8	0	8
DSDG Sub-Committee (Non-APS/ Non-Indigenous Members)	1	1	0
DCN Working Group (APS Members)	17	15	2
DNC Working Group (Non-APS/Indigenous Members)	14	0	14
	56	**31**	**25**

The Australian Federal Government positions itself as using a 'co-design' method.

As the name suggests, co-design involves a product co-designed by consumers and an entity, such as the government or universities.

It should be noted that 'co-design' receives mixed reviews from Aboriginal and Torres Strait Islander people. It is not sovereignty because power and the final decision ultimately rests with those with power. This is generally not Indigenous people.

Some Aboriginal and Torres Strait Islander people have been as critical as to say it is window dressing on a government product.

- "They say they want to co-design something, but at the end of the day, they are the ones who are making the final decisions, not us".
- "Co-design can be good. The government listens to us; we create something together, and both own it."

Co-design is more complex than a binary good or bad; various components contribute to its meaningfulness. However, it is essential to underline that co-design is not IDSov or IDGov as defined by Aboriginal and Torres Strait Islander people.

APS Framework Guidelines

- **Built upon the International Data Principles of CARE & FAIR – not the national IDSov Principles (Taitingfong et al. 2024).**
  - CARE stands for Collective benefit, Authority to control, Responsibility, and Ethics.
  - FAIR stands for Findability, Accessibility, Interoperability, and Reusability.
- **Guideline 1: Partner with Aboriginal and Torres Strait Islander people**
  - Partner with Aboriginal and Torres Strait Islander people at all stages of the data lifecycle (see Appendix C) to ensure their priorities are reflected in data about their communities.
- **Guideline 2: Build data-related capabilities**
  - Improve the capabilities of APS staff and First Nations partners relating to Indigenous data across the data lifecycle.
- **Guideline 3: Provide knowledge of data assets**
  - Develop straightforward methods for Aboriginal and Torres Strait Islander people to know what data are held relating to their interests, its use, and how it can be accessed.
- **Guideline 4: Build an inclusive data system**
  - Build towards organisational and cultural change to support the inclusion of Aboriginal and Torres Strait Islander people in data governance.
- **Each guideline has 6 to 14 action items, a total of 38 action items.**

The Australian Public Service Framework is built upon the Global Indigenous Data Alliance (GIDA) Principles of CARE, along with the principles of FAIR.

It is important to recognise that while the CARE principles are important, they are not the Maiam nayri Wingara IDSov principles.

The Framework is built upon four guidelines aimed at achieving greater data equity.

Each of the guidelines is a series of action items. The action items have three stages that build upon the previous. These are:

- **Prepare**
- **Grow**
- **Extend**

Next steps for the APS Framework

- Each department and/or agency is developing implementation plans to progress action items, which need to be completed by the end of 2024.
- Ongoing capacity building within the government and within Indigenous communities.
- Finalising mechanism to publicly show how each department is progressing.
- We have called for the appointment of an Indigenous Data Commissioner.
- Suggested one stop shop for Aboriginal and Torres Strait Islander people for access to their government held data.
- State and Territory Governments need to enact similar Frameworks as parties to the National Agreement on Closing the Gap Strategy under priority reform 4.
  - NSW are using the APS Framework to be bolder in their equivalent.

For the rest of 2024, each Federal Government department and/or agency is developing implementation plans to progress each of the 38 action items.

These action plans need to be completed by the end of 2024. Ring up your relevant department or agency from 2nd January 2025 and ask them to see their Framework Implementation Plan.

There needs to be ongoing capacity building within the government. There is a culture regarding how Indigenous data is done. This generally consists of comparing Indigenous and non-Indigenous peoples across most measures and reporting on it. Aboriginal and Torres Strait Islander people don't want to know how every aspect of our lives compares to the non-Indigenous population on every government output.

The Federal Government committed to having a publicly showing mechanism on how each department and/or agency is progressing with the Framework. Too often, these documents disappear into the federal government, and we never know how they are going. Often they don't progress anywhere and they get rolled into a new document.

In the foreword, we have called for the appointment of an Indigenous Data Commissioner.

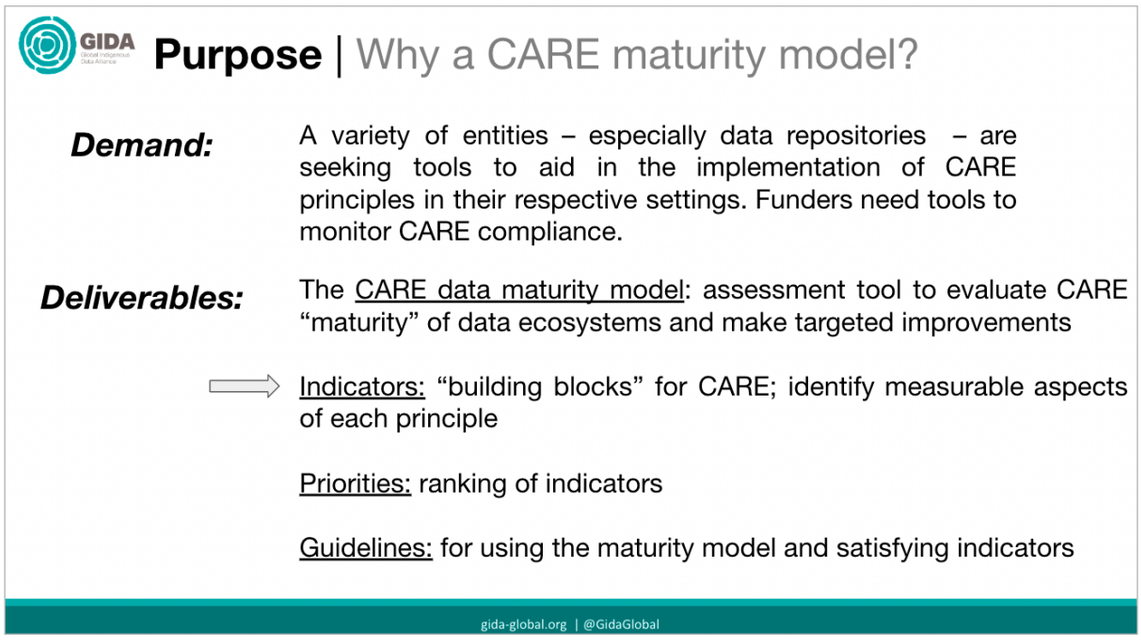
Because of the structure of the Australian government system and the Closing the Gap Strategy, each state and territory needs to enact a similar Framework under Priority Reform 4, which concerns data.

The state of New South Wales (NSW, where Sydney is located) is using the Framework to be bolder in its equivalent document.

Activity 1: In small groups, discuss the key similarities and differences between the Māori Data Governance Model (Aotearoa) and the Framework for Governance of Indigenous Data (Australia)

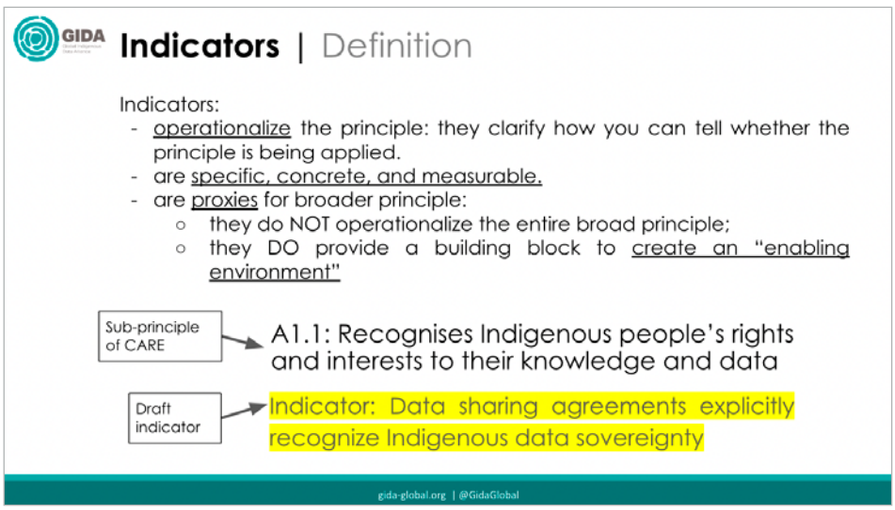
Section 3.3. The CARE Data Maturity Model

Cheryl Ellenwood

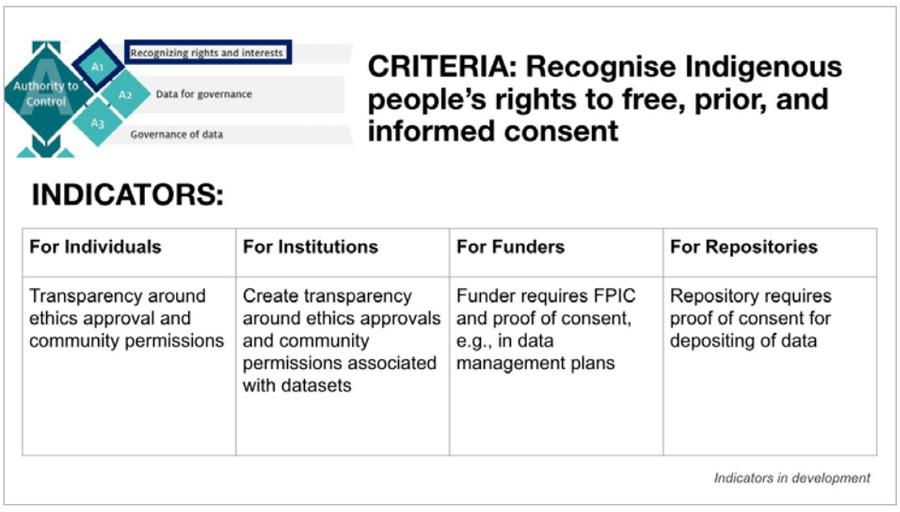


The CARE Principles for Indigenous Data Governance (Principles) were created by Indigenous researchers in 2018 – you can read more about them on the Global Indigenous Data Alliance (GIDA) website – https://www.gida-global.org/care. The Principles are people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self-determination. The Principles have been adopted or endorsed by a wide range of international bodies and instruments including the Research Data Alliance and the UNESCO Recommendation on Open Science.

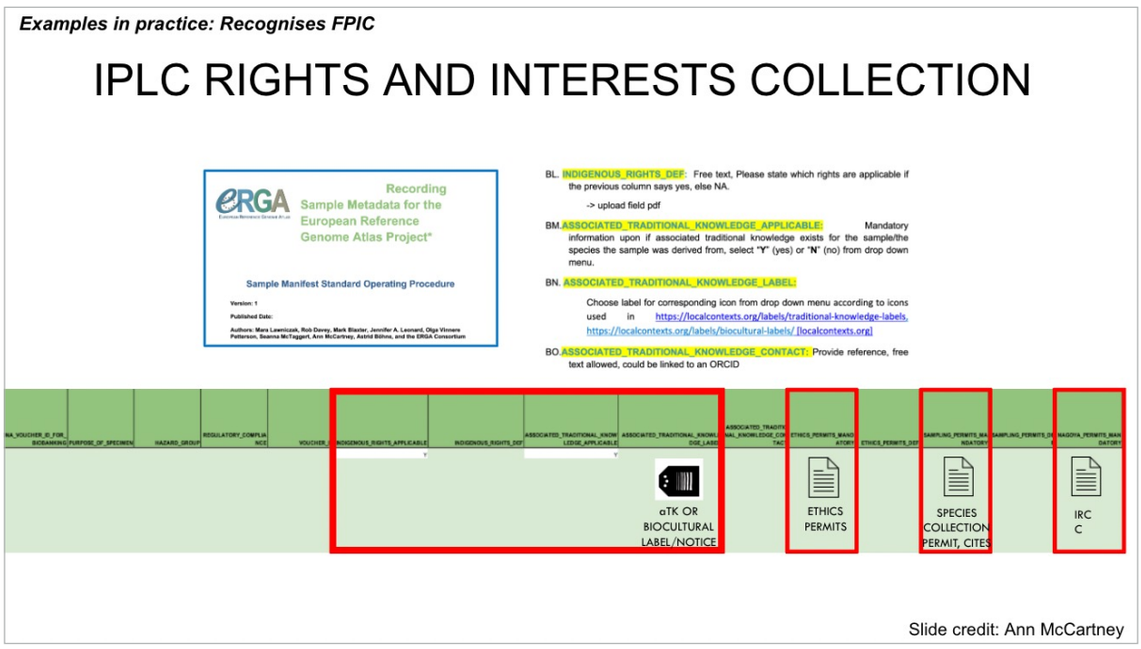
This slide identifies the key elements of the CARE Data Maturity Model (CDM Model). The CDM Model is a tool that can be used to assess, measure and improve upon the implementation of the CARE principles. The **indicators** are the building blocks for the implementation of CARE and allows for the identification of measurable aspects of each principle. **Priorities** allow for a consideration of the different level of readiness across communities. **Guidelines** help explain in detail what each indicator is about and how it can be satisfied.



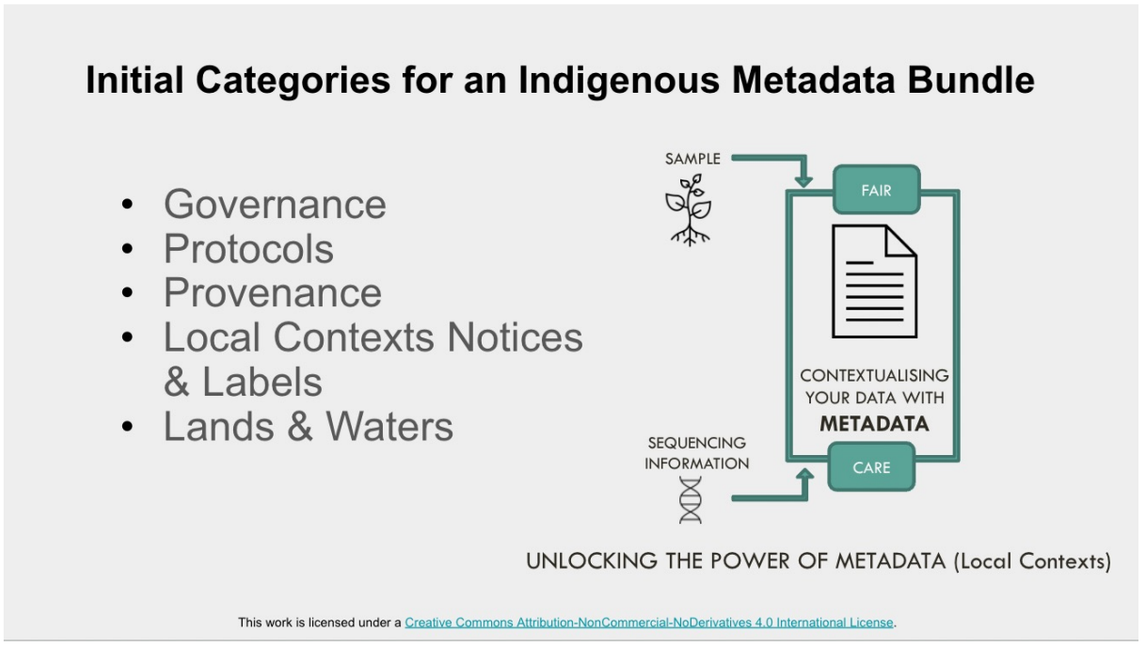
Indicator expectations are defined as something that is concrete and measurable. Ideally, indicators help to focus on and target specific areas where decisions get made. Foundational or initial steps are necessary to pursuing the broader principle.



This slide provides consideration and examples of criteria for different entities such as individuals, institutions, funders and repositories. Additional considerations for other actors will likely continue.



This slide demonstrates that there are fields pertaining to the rights of Indigenous peoples and local communities including traditional knowledge (TK) or biocultural (BC) labels.



This slide further demonstrates usage in the context of the Indigenous Metadata Bundle. Indigenous metadata provides critical organization and structure for Indigenous peoples' data to be findable, accessible, interoperable and with proper attribution, which enables governance, decision-making and cultural authority by Indigenous peoples. Indigenous metadata guides the inclusion of cultural context, data protections, and ownership for Indigenous rights holders. This helps to recognize, refine, and define the relationships that Indigenous peoples have with their data.

References:

Carroll, S.R., Herczog, E., Hudson, M. et al. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Sci Data 8, 108*. https://doi.org/10.1038/s41597-021-00892-0

Riley Taitingfong, R., Martinez, A., Russo Carroll, S., Hudson, M., & Anderson, J. (2023). *Indigenous metadata bundle communiqué.* Collaboratory for Indigenous Data Governance, ENRICH: Equity for Indigenous Research and Innovation Coordinating Hub, and Tikanga in Technology. doi: 10.6084/m9.figshare.24353743

Jennings, L., Anderson, T., Martinez, A. et al. Applying the CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. *Nat Ecol Evol 7*, 1547–1551 (2023). https://doi.org/10.1038/s41559-023-02161-2

Russo Carroll, S., Taitingfong, R., Jennings, L., Prehn, J., Small-Rodriguez, S., Walker, J. D., Sedran-Price, C., Kukutai, T. & Franks, A. (2024). "CARE Directs Us Home: Prioritizing Indigenous Peoples' Community Standards Communiqué." Global Indigenous Data Alliance.

Overview of Draft CARE Indicators (in development)

COLLECTIVE BENEFIT	AUTHORITY TO CONTROL	RESPONSIBILITY	ETHICS
Conduct data needs assessment Utilise Indigenous identifiers Supporting Indigenous use Alignment of permissions for data access and re/use to Indigenous frameworks Indigenous approval of outputs from research projects Ensure Indigenous Peoples determine benefits Develop benefit sharing plans Fund training and education	Recognition of Indigenous data sovereignty Recognition of Indigenous Peoples' rights to FPIC Transparent ethics approval processes Transparent community permissions processes Enable audit of Indigenous data Make disclosures to Indigenous communities about Indigenous data	Build relationships with Indigenous Peoples Support community capacity-building Promote equitable attribution including acknowledgment and authorship Collect data relevant to Indigenous languages and worldviews Ensure data of interest are findable by communities Enable Indigenous metadata fields	Support use of Indigenous ethical frameworks Promote Indigenous interpretation and presentation of findings Share data of interest with Indigenous organizations Reflect Indigenous knowledge systems in agreements Compensate research participation Share copyright Agreements reflect Indigenous methods for dispute resolution Administrative mechanisms for rights violations in research

Contact: usindigenousdatasovnetwork@gmail.com

Activity 2: Answer the following questions for your own community/project/context.

1. Which activities apply to your context?
2. Which activities are most feasible to implement now?
3. Which activities seem more challenging to implement?
4. What's missing?

Resources

Watch 

Series of short videos (2 – 10 mins) on Māori data and Māori data governance. Te Kāhui Raraunga. https://www.kahuiraraunga.io/

First Nations Information Governance Centre. Understanding the First Nations Principles of OCAP. https://fnigc.ca/ocap-training/

Walter, Maggie (2018) Keynote Address – Resetting Indigenous data. https://www.youtube.com/watch?v=P9f86EAHaNY

Read 

Sterling, R., Kukutai, T., Chambers, T., & Chen, A. 2024. A Māori data governance assessment of the NZ Covid Tracer App. *Discover Social Science and Health.* https://doi.org/10.1007/s44155-024-00092-2

Carroll, S. R., Garba, I., Plevel, R., Small-Rodriguez, D., Hiratsuka, V. Y., Hudson, M., & Garrison, N. A. (2022). Using Indigenous standards to implement the CARE principles: Setting expectations through tribal research codes. *Frontiers in Genetics, 13,* 823309. https://doi.org/10.3389/fgene.2022.823309

Commonwealth of Australia (2024). Framework for Governance of Indigenous Data. Australian Public Service. Available at: https://www.niaa.gov.au/sites/default/files/documents/2024-05/framework-governance-indigenous-data.pdf

Explore 

Te Mana Raraunga (2018). *Māori Data Sovereignty Principles*. Available at: https://www.temanararaunga.maori.nz/

CARE Principles for Indigenous Data Governance. Available at https://www.gida-global.org/care

Maiam nayri Wingara (2024). Maiam Nayri Wingara – About Us. Available at: https://www.maiamnayriwingara.org/

Section 4. Keeping our data sovereign

Chris Cormack, Peter Lucas-Jones & Lynell Tuffery Huria

Learning objectives:

Objective 1	Understand the risks and challenges associated with processing and storing data offshore
Objective 2	Gain insights into IDSov pathways and solutions
Objective 3	Understand how Intellectual Property Rights relates to Indigenous data and knowledges

Key Concepts and Definitions:

- **Data residency**
  The physical or geographic location where data are stored (e.g., servers, databases, data centers) and processed.

- **Data localisation**
  The practice of storing or processing data within a specific geographic location.

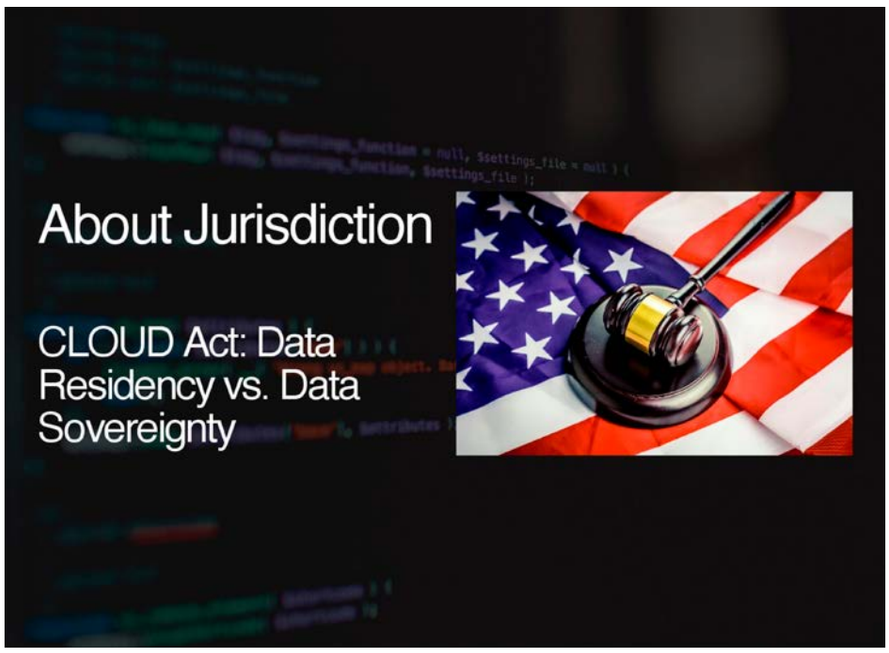
- **Decentralised data system**
  A data storage and processing system where data are spread across multiple locations – there is no centralised control.

- **Distributed data system**
  A data storage and processing system where data are spread across multiple locations but there may still be a centralised authority.

- **Large Language Models (LLM)**
  A type of Artificial Intelligence model that uses deep learning techniques and massive amounts of written human language and textual data to understand, summarize and generate new content.

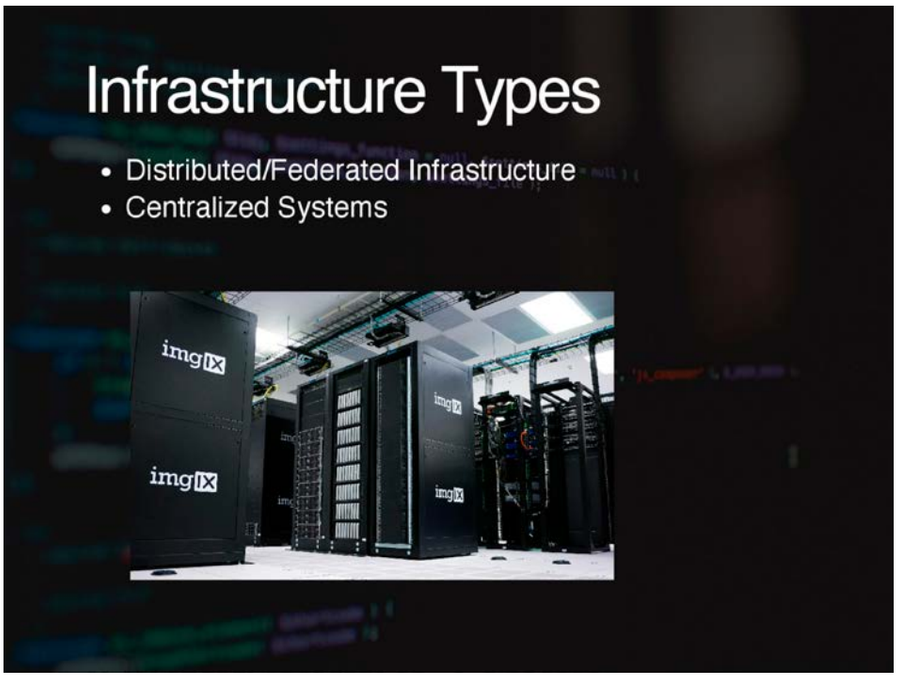
Section 4.1. Sovereign data infrastructure

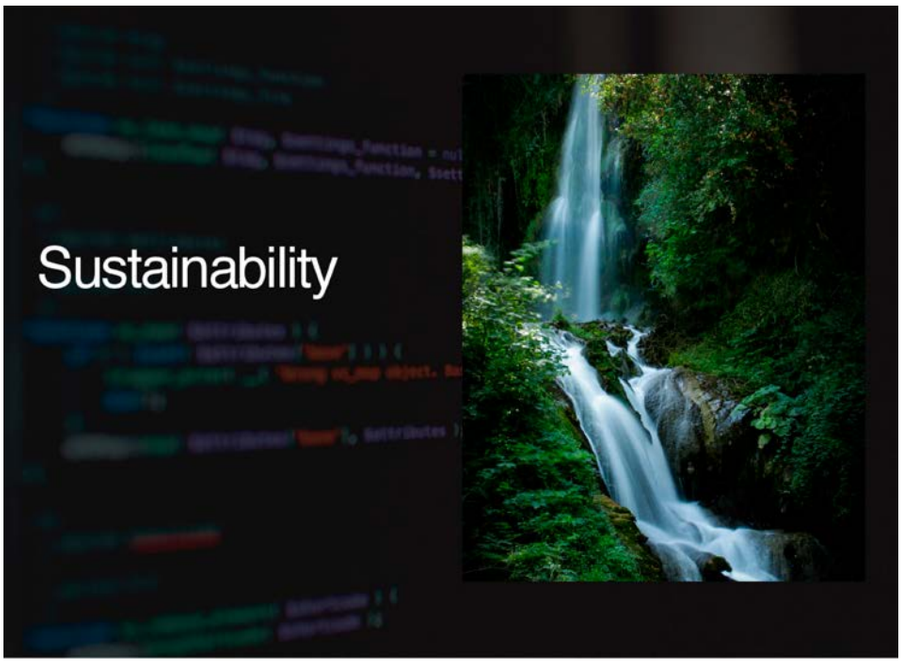
Chris Cormack



Reference:

Cormack, D., Kukutai, T., & Cormack, C. 2020. Not one byte more. In A. Chen (ed), *Shouting zeros and ones. digital technology, ethics and policy in New Zealand* (pp. 71-83). Bridget Williams Books.

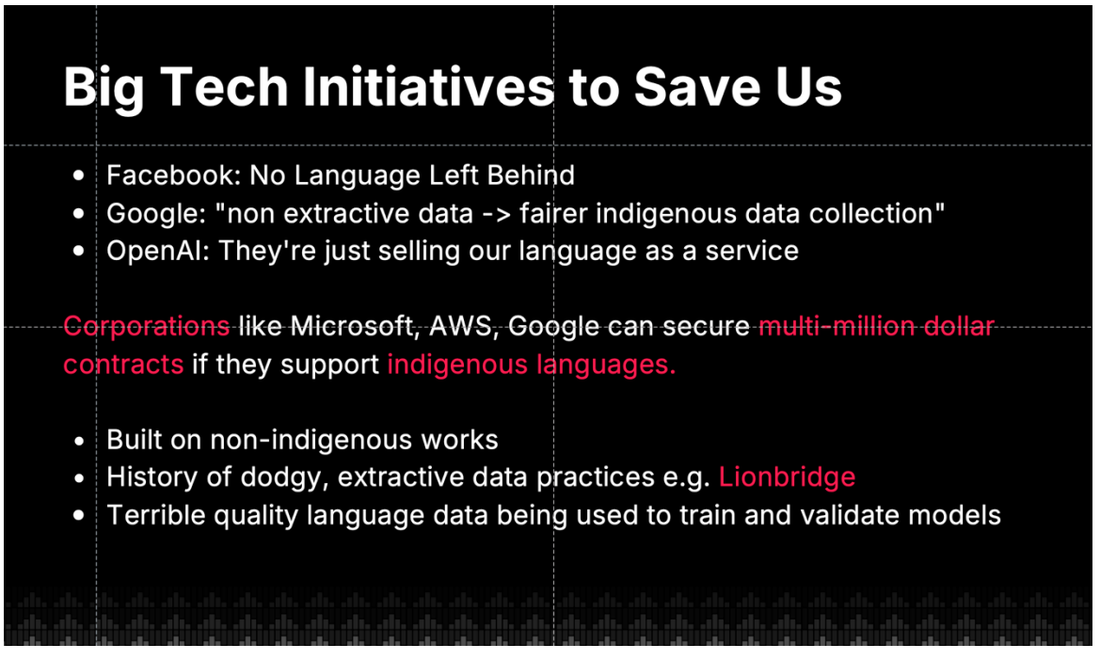




Section 4.2. Sovereign data infrastructure for Te Reo Māori

Peter-Lucas Jones

Te Hiku Media is a Māori charitable media organisation, set up by the tribes of the far north region more than 30 years ago, to revitalise and promote the Māori language. In recent years it has developed its own AI tools including a Natural Language Processing tool, a Māori language pronunciation app Rongo, a Kaitiakitanga license and a media platform app Whare Kōrero where third parties, mainly tribal radio stations, can upload their content. This presentation describes some of the initiatives that Te Hiku has been leading and its fight against Big Tech imperialism.



Supported languages

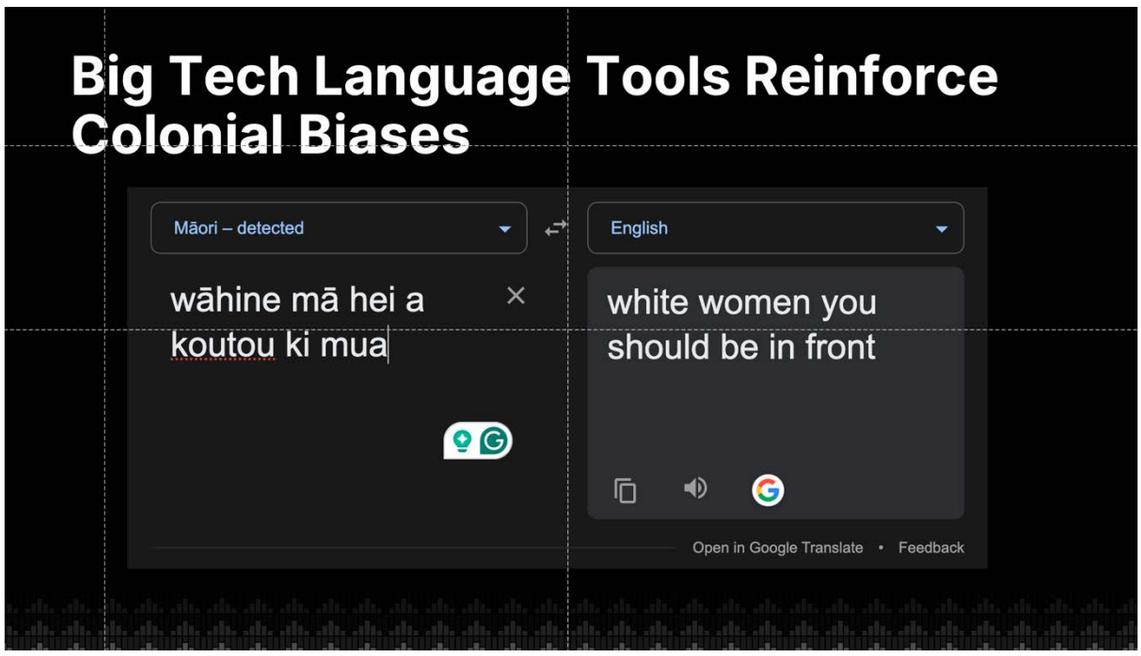
We currently support the following languages through both the `transcriptions` and `translations` endpoint:

Afrikaans, Arabic, Armenian, Azerbaijani, Belarusian, Bosnian, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, Galician, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Japanese, Kannada, Kazakh, Korean, Latvian, Lithuanian, Macedonian, Malay, Marathi, Maori, Nepali, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swahili, Swedish, Tagalog, Tamil, Thai, Turkish, Ukrainian, Urdu, Vietnamese, and Welsh.

While the underlying model was trained on 98 languages, we only list the languages that exceeded <50% word error rate (WER) which is an industry standard benchmark for speech to text model accuracy. The model will return results for languages not listed above but the quality will be low.

Result of Big Tech "Leading" for Us

- Half-baked tools that do damage to our languages
- Indigenous groups are not being funded because big tech have already done the work
- Corporations profit from our languages that were once outlawed and beaten out of our people
- Missed opportunities to grow capacity in indigenous communities
- Loss of economic opportunities



Kōrero Tuku Iho

- Observation
- Measurement
- Data
- Information (who, what, when, where)
- Knowledge (how)
- Wisdom (why)
- Implementation (outcome-focused)
- Data Pipeline
- Informing Community Led Data Science

The FLORES-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation

Naman Goyal[*], Cynthia Gao[*], Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato[†], Francisco Guzmán[†], Angela Fan[†‡]
Facebook AI Research, [‡]LORIA
flores@fb.com

Abstract

One of the biggest challenges hindering progress in low-resource and multilingual machine translation is the lack of good evaluation benchmarks. Current evaluation benchmarks either lack good coverage of low-resource languages, consider only restricted domains, or are low quality because they are constructed using semi-automatic procedures. In this work, we introduce the FLORES-101 evaluation benchmark, consisting of 3001 sentences extracted from English Wikipedia and covering a variety of different topics and domains. These sentences have been translated in 101 languages by professional translators through a carefully controlled process. from improvements in translation quality on their native languages. As a result, the field has been shifting focus towards low-resource languages.

Over the past decade, the research community has made a lot of recent progress on models for low-resource machine translation. Approaches like iterative backtranslation (Sennrich et al., 2015), multilingual machine translation (Johnson et al., 2016; Tang et al., 2020; Fan et al., 2020), and even unsupervised machine translation (Lample et al., 2018; Artetxe et al., 2018) have shown promising results. Beyond modeling, a major challenge for research in low-resource machine translation is evaluation. Low-resource evaluation is critical to the scientific progress of the field, because evaluation enables

"translated in 101 languages by professional translators through a carefully controlled process"

---------- Forwarded message ----------
From: **Lionbridge Recruitment** <recruitmentWM.tampere@lionbridge.com>
Date: Wed, Feb 28, 2018 at 1:52 AM
Subject: Lionbridge – Hawaiian Language Project
To: [redacted]

Dear [redacted]

My name i[redacted]and I am part of the Lionbridge Global Services for Machine Intelligence Recruitment Team in Tampere, Finland.

We are helping to develop modern speech recognition systems for computers and mobile devices. We believe those systems to be important for future development. Unfortunately, modern systems usually neglect to implement minority languages and focus mainly on commonly spoken languages like English, Spanish or Chinese.

We strongly believe in preserving a large diversity in languages and cultures and so we would kindly ask your support in keeping the Hawaiian language alive.

We are currently working on a language research project and its goal is the development of speech recognition for **Hawaiian** language. We would like to know if you would like to help us.

The tasks can be done from home using a mobile device (Android) or a computer with Chrome (no iPhone/iPad).

Workload: Approximately **1 hour** of engagement. The goal of the task is to record your own voice while reading randomized sentences in **Hawaiian** out loud. It is possible to work using a mobile device (Android) or a computer with Chrome (no iPhone/iPad).

Who can participate in this task: **Hawaiian** speakers of age 18+. Being native is not a key requirement, but participants must be able to speak and read sentences in Hawaiian fluently.

Estimated start date of this project: immediate involvement;

Compensation: **45 USD** for the entire task.

How to apply: to participate, please fill a short registration form at **http://your_voice_haw-US.register-lionbridge.com**

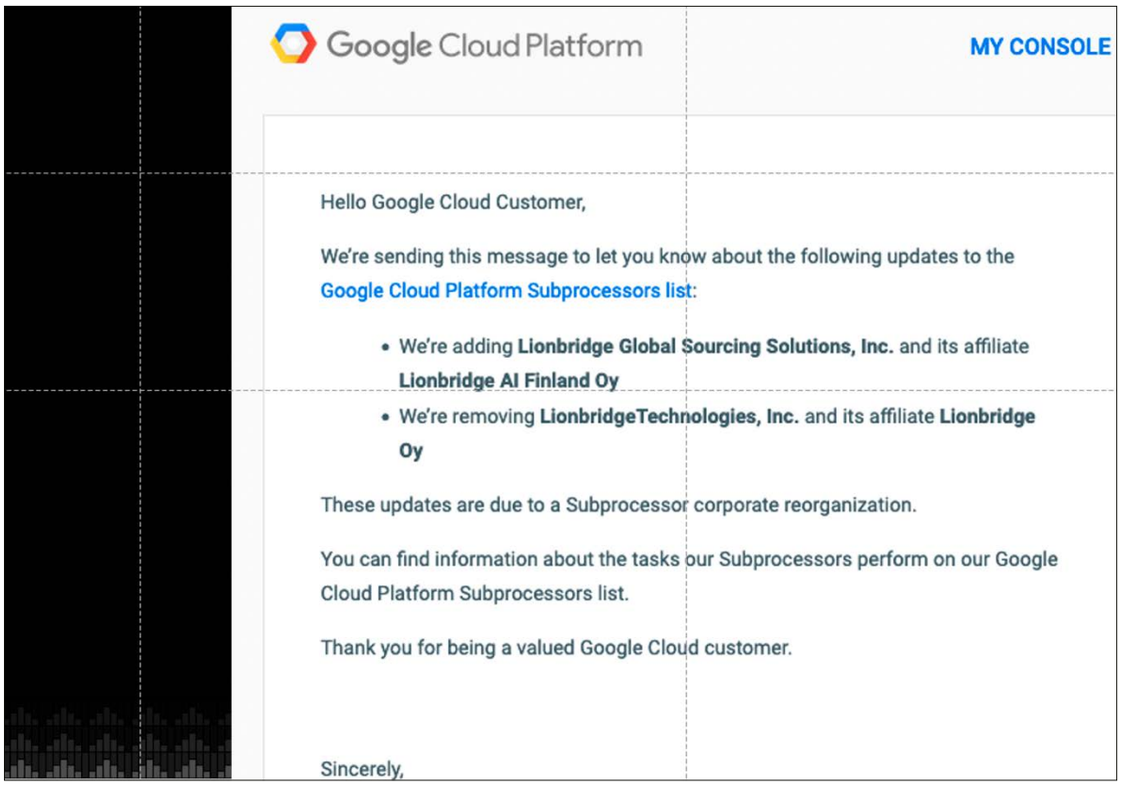
After you complete the registration, you will be contacted after a few days by the Production team.

We will also compensate you with **5 USD** per each person that you recommend and completes the task. If you are interested in this compensation for the recruitment, please contact us and we will send further instructions on how to proceed.

Thank you very much in advance, we are looking forward to working with you!

Best regards,

Ilaria



Google Cloud Platform MY CONSOLE

Hello Google Cloud Customer,

We're sending this message to let you know about the following updates to the Google Cloud Platform Subprocessors list:

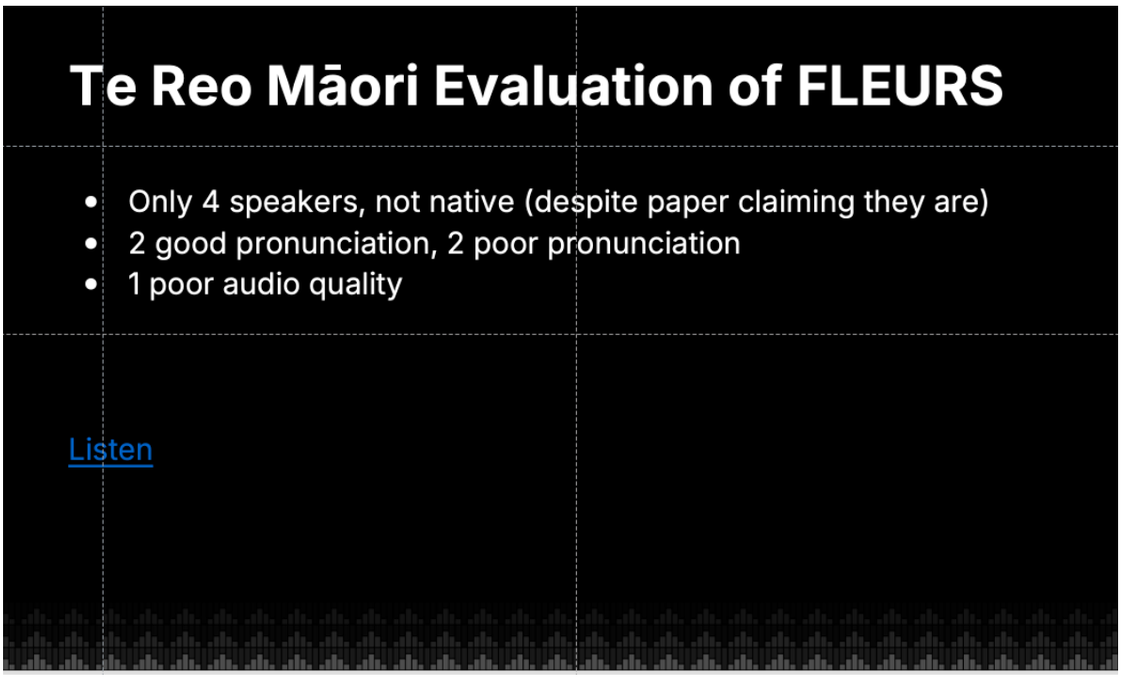
- We're adding **Lionbridge Global Sourcing Solutions, Inc.** and its affiliate **Lionbridge AI Finland Oy**
- We're removing **LionbridgeTechnologies, Inc.** and its affiliate **Lionbridge Oy**

These updates are due to a Subprocessor corporate reorganization.

You can find information about the tasks our Subprocessors perform on our Google Cloud Platform Subprocessors list.

Thank you for being a valued Google Cloud customer.

Sincerely,



Te Reo Māori Evaluation of FLEURS

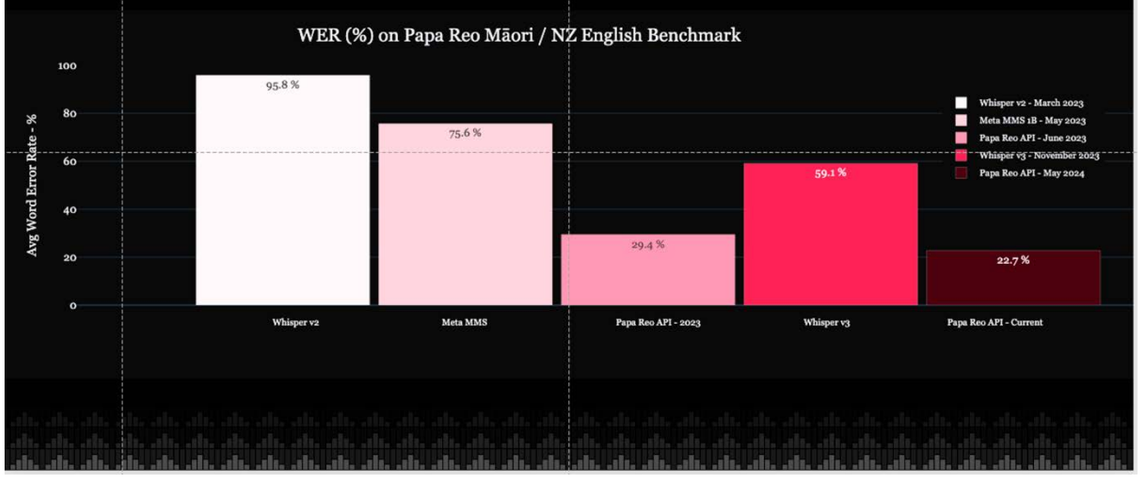
- Only 4 speakers, not native (despite paper claiming they are)
- 2 good pronunciation, 2 poor pronunciation
- 1 poor audio quality

Listen

Issues with OpenAI's Whisper

- Unclear where the data comes from and what rights they had to use that data
- ~1300 hours of te reo Māori, ~400 hours of ʻŌlelo Hawaiʻi, many other indigenous languages
- Benchmarked on FLEURS

Māori Led & Developed ASR is Best

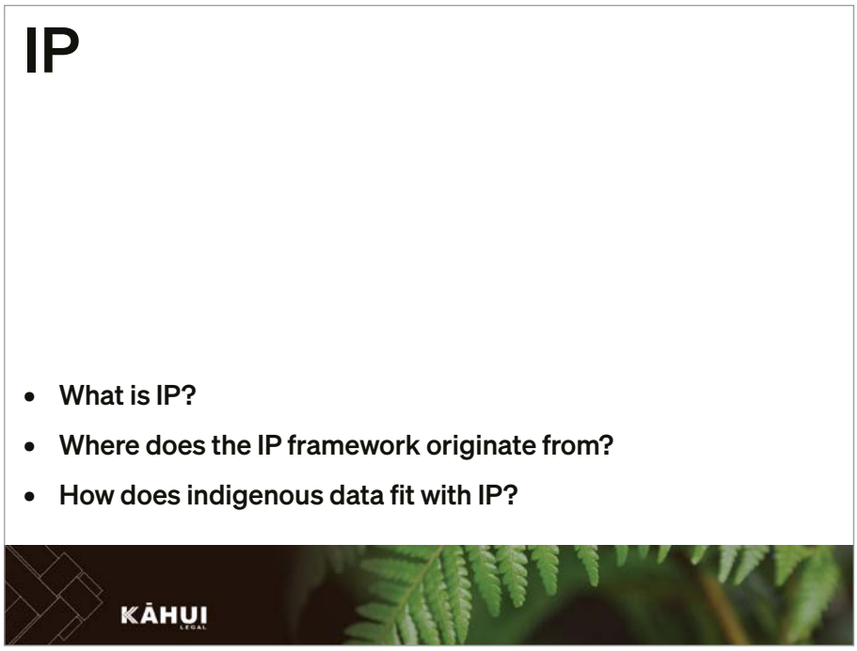


References:

Mahelona, K., Leoni, G., Duncan, S., & Thompson, M. (2023). Open AI's Whisper is another case study in colonisation. Available at: https://blog.papareo.nz/whisper-is-another-case-study-in-colonisation/

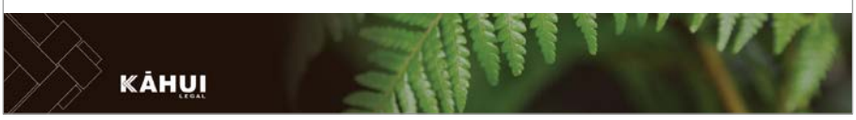
Section 4.3. IP, FTAs and keeping our data sovereign

Lynell Tuffery Huria



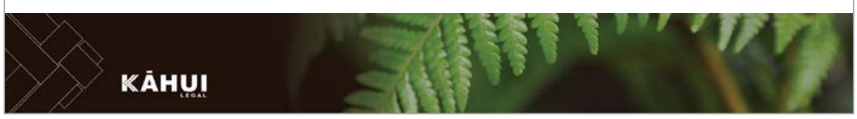
Wai 262

- agreed to various international agreements and obligations that affect indigenous flora and fauna and the intellectual property rights and rights to other taonga



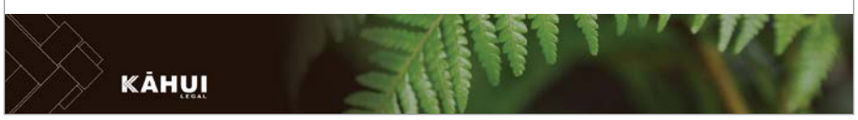
Wai 262

- Claim about recognising “Mātauranga Māori” (Māori knowledge)
- Attempts to reconcile Māori traditional knowledge system with European intellectual property system
- Creates a framework for recognising Māori authority where possible and at the same time balancing other interests



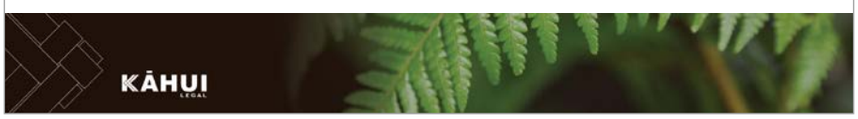
Wai 262

- **It is necessary to protect Māori culture and identity because that is how we protect New Zealand culture and identity**
- **New Zealand's law should make room for the relationships between kaitiaki and their taonga works, taonga species, and mātauranga Māori to flourish as a matter of national interest.**



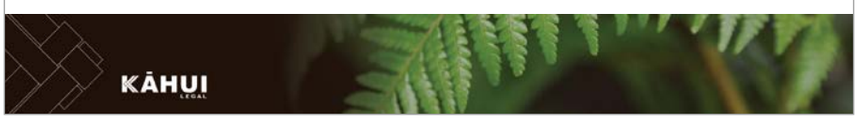
Wai 2522

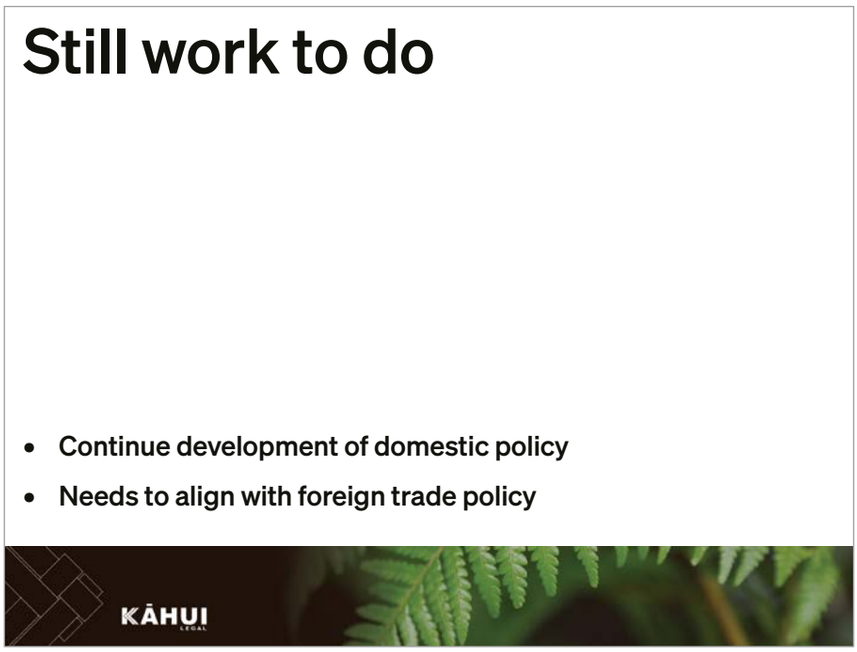
- Claim about the e-commerce chapter in TPPA
- E-commerce chapter introduces rules including:
  - Free and open internet
  - No digital customs duties
  - Unimpeded cross-border data flows
  - No localisation barriers
  - No forced technology transfers
  - Protection of source code



Wai 2522

- Data can incorporate mātauranga
- Reaffirmed mātauranga is a taonga – should not be balanced against other trade objectives or interests of other citizens or sectors
- Lack of Māori involvement or consideration of Māori rights and interests
- Invokes obligations of active protection
- Exception clause not sufficient
- Need to develop a domestic policy first with Māori involved





References:

Kukutai, T., Cassim, S., Clark, V., Jones, N., Mika, J., Morar, R., Muru-Lanning, M., Pouwhare, R., Teague, V., Tuffery Huria, L., Watts, D. & Sterling, R. (2023). *Māori data sovereignty and privacy. Tikanga in Technology discussion paper.* Hamilton: Te Ngira Institute for Population Research. Available at: https://www.waikato.ac.nz/assets/Uploads/Research/Research-institutes-centres-and-groups/Institutes/Te-Ngira-Institute-for-Population-Research/MDSov-and-Privacy_20March2023_v2.pdf

Waitangi Tribunal (2021). *Report on the comprehensive and progressive Agreement for Trans-Pacific Partnership* (WAI 2522). Available at: https://forms.justice.govt.nz/search/Documents/WT/wt_DOC_178856069/CPTTP%20W.pdf

Waitangi Tribunal (2011). *Ko Aotearoa tēnei: A report into claims concerning New Zealand law and policy affecting Māori culture and identity* (WAI 262, Vol. 1). Available at: https://forms.justice.govt.nz/search/Documents/WT/wt_DOC_68356416/KoAotearoaTeneiTT2Vol1W.pdf

Activity 1: In small groups, think about what things we could do/build to achieve greater IDSov

Resources

Watch

Cormack, C. (2023). Libraries and Indigenous data sovereignty. PerlKohaCon. https://www.youtube.com/watch?v=7HHMlYXr6p4

Jones, P.L. (2023). Protecting our future – Indigenous data sovereignty. https://www.youtube.com/watch?v=y0muBYqlprE

Mahelona, K. (2023). AI and Big Tech imperialism. PAI Policy Forum – Lightning talk. https://www.youtube.com/watch?v=nNII6_l0s00

Tuffery Huria, L. Protecting taonga and mātauranga in contracts. https://www.facebook.com/reel/837188355220527

Read

Cormack, D., Kukutai, T., & Cormack, C. (2020). Not one byte more. In A. Chen (ed), *Shouting zeros and ones. digital technology, ethics and policy in New Zealand* (pp. 71-83). Bridget Williams Books.

Hao, K. (2022). A new vision of Artificial Intelligence for the people. *MIT Technology Review*. Available at: https://www.technologyreview.com/2022/04/22/1050394/artificial-intelligence-for-the-people/

Jones, P-L., Mahelona, K., Duncan, S., & Leoni, G. (2023). Kia tangata whenua: Artificial Intelligence that grows from the land and people. *Indigenous Communication Landscapes, 20*. https://doi.org/10.21428/0af3f4c0.9092b177

Mahelona, K., Leoni, G., Duncan, S., & Thompson, M. (2023). Open AI's Whisper is another case study in colonisation. Available at: https://blog.papareo.nz/whisper-is-another-case-study-in-colonisation/

Explore

Kaitiakitanga License. Available at: https://tehiku.nz/te-hiku-tech/te-hiku-dev-korero/25141/data-sovereignty-and-the-kaitiakitanga-license

Section 5. Growing Data 'Doers'

Kateriina Selwyn & Brittany Baltus (Te Mana Whakatipu, Te Kāhui Raraunga)

Learning objectives:

Objective 1	Whakamana te tirohanga Māori - Understand how to re-ground kaupapa in te Ao Māori
Objective 2	Whakamana ngā iwi Māori - Understand how to strengthen connections to iwi Māori
Objective 3	Whakamana ngā tāngata Māori - Understand how to grow 'data doers' / 'data disrupters'

Key Concepts and Definitions:

➢ Data lifecycle

- Refers to the processes or phases that data goes through – its "life".
  - "...the different stages of a unit of data undergoes from initial collection through to when it is no longer considered useful and deleted" (Sterling et al., 2023)
  - "The data lifecycle is the ongoing process of collecting, using, storing, transferring, and destroying or permanently archiving data. This process is repeated again and again with different data sets" (Indigenous Data Toolkit, n.d.)
  - "Data has a lifecycle: collect, describe, store short-term, analyse and check, use, save or destroy. You will often see 'plan' added to the beginning of this lifecycle, but we like to think that planning relates to all stages" (data.govt.nz, 2021).

➢ Coloniality of knowledge

- **Colonisation:** the word 'colonisation' locates it in the past.
  - "... as an "event" often talked about as historic or complete (D Cormack, personal communication, May 21, 2023)
- **Coloniality:** refers to the idea that 'colonisation' is an active, ongoing issue in the present day.
- **Knowledge & Power:** the idea that the eradication of knowledge systems was a deliberate move to disempower groups of people.
  - Challenges to knowledge systems were a fundamental part of colonisation and remain central to coloniality" (D. Cormack, personal communication, May 21, 2023)
  - ..."'Idea of 'epistemicide': the deliberate destruction of knowledge systems" (D. Cormack, personal communication, May 21, 2023)
- "What they mean by "literature review" is stuff written as literature by white people. Yet in our knowledge system, literature as written is only a recent innovation, but that does not mean that literature did not exist—it is there in the poetry, the songs, the history, the traditions of all Indigenous peoples. And yet to try to get a university to accept that way of knowing and seeing the world

constitutes a valid literature review, is one of those many issues that Indigenous peoples throughout the Western academy continue to wage" (Jackson 2016: 29-30)

- **Decolonising Knowledge:** reclaiming our own ways of knowing in the spaces we operate
  - "Opening the space for multiple knowledge systems, critiquing the idea of "universal" knowledge systems and pushing back against the idea of hierarchies of knowledge" (D. Cormack, personal communication, May 21, 2023).

References:

Data.govt.nz. (2021, April 23). *Introduction to data management*. https://www.data.govt.nz/toolkit/data-management/introduction-to-data-management/

Indigenous Data Toolkit. (n.d.). *Indigenous data sovereignty and governance*. https://indigenousdatatoolkit.ca/

Grosfoguel R. (2002). Colonial difference, geopolitics of knowledge, and global continuity in the modern/colonial capitalist world-system. *Review, 25*(3), 203-224.

Jackson, M. (2016, November 8–10). Keynote address [Conference presentation]. Lowitja Institute International Indigenous Health and Wellbeing Conference, Melbourne, Australia. https://www. youtube.com/watch?v=oWjewpB6UX8

Smith L. T. (2012). *Decolonizing methodologies*, 2nd edition. Otago University Press.

Sterling, R., Blake, M., Jones, N., Hartshorn, R., & Kukutai, T. 2023. *The Research Data Landscape in Aotearoa New Zealand: A report undertaken in partnership with the Aotearoa New Zealand Committee on Data in Research (CoDiR).* 10.15663/UoW.RDLA.DEC2023

CASE STUDY 1: Iwi-led Census Collections 2023

Background

During the 2018 Census roll out, there were a number of issues and changes which impacted Māori responses to Census. These included:

- A lack of emphasis on field collection, and instead an emphasis on online collection.
- A lack of employment of sufficient staff to undertake field collection to support the full enumeration of low response areas & populations.
  - Most of those low response areas had high populations of Māori.
- A lack of printed English & Te Reo Māori forms to enable field staff to deliver forms prior to Census Day.

As a result of the above factors, 2018 Census failed to deliver high-quality, Māori data.

Te Mana Whakatipu

"Whakatipuria tō mana i runga i tō mana whakaheke"
Harness the potential and gifts instilled in you from your tūpuna, in order to help you realise your own potential

In response to poor response rates in 2018, Te Kāhui Raraunga developed a programme to implement Iwi-Led Census Collections for the 2023 Census period. 'Te Mana Whakatipu' is an iwi-led data collection and data analytics capability and capacity development programme. The programme has two key components:

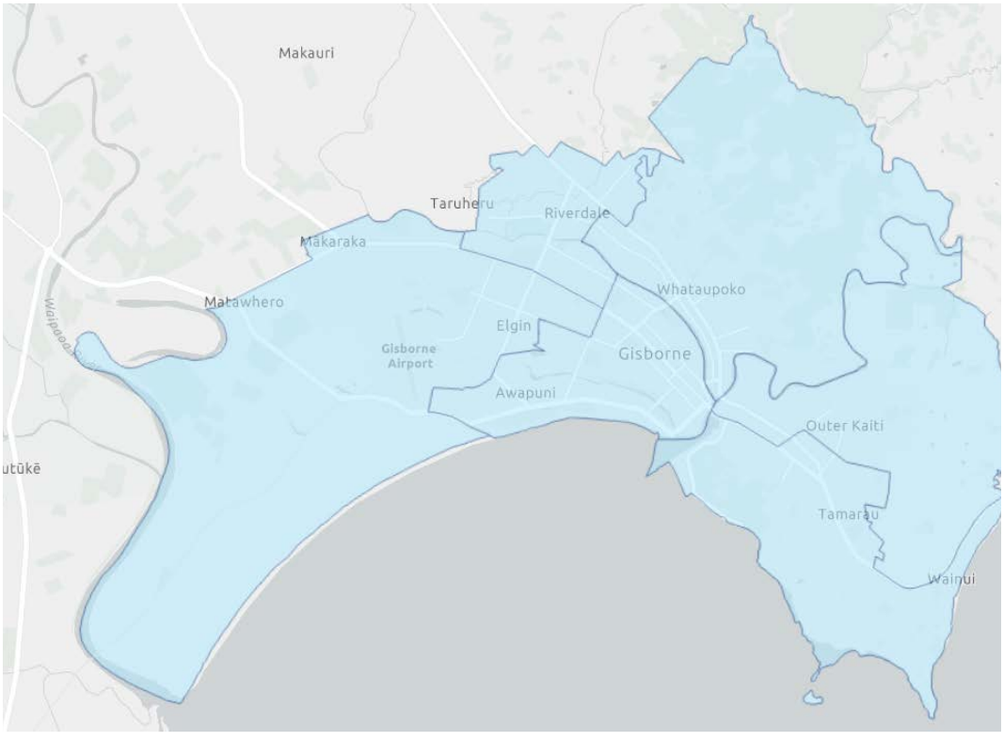
1. Iwi leading a significant field data collection operation – Census 2023.
2. Build iwi Māori capacity and capability to undertake data analytics.

2023 Iwi-Led Collections

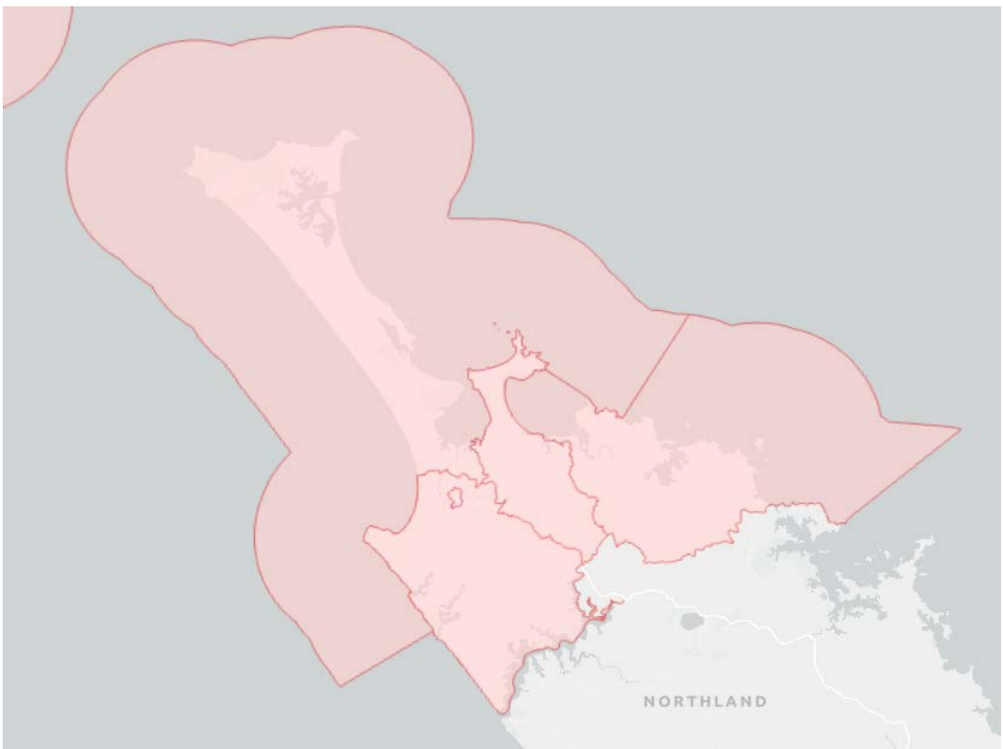
As part of Te Mana Whakatipu, three iwi collectives participated in the Iwi-Led Census Collections for the 2023 Census. All three iwi-led areas were a few of the lowest responding areas for 2018 Census.

The iwi collectives who took part in Iwi-Led Collections were:

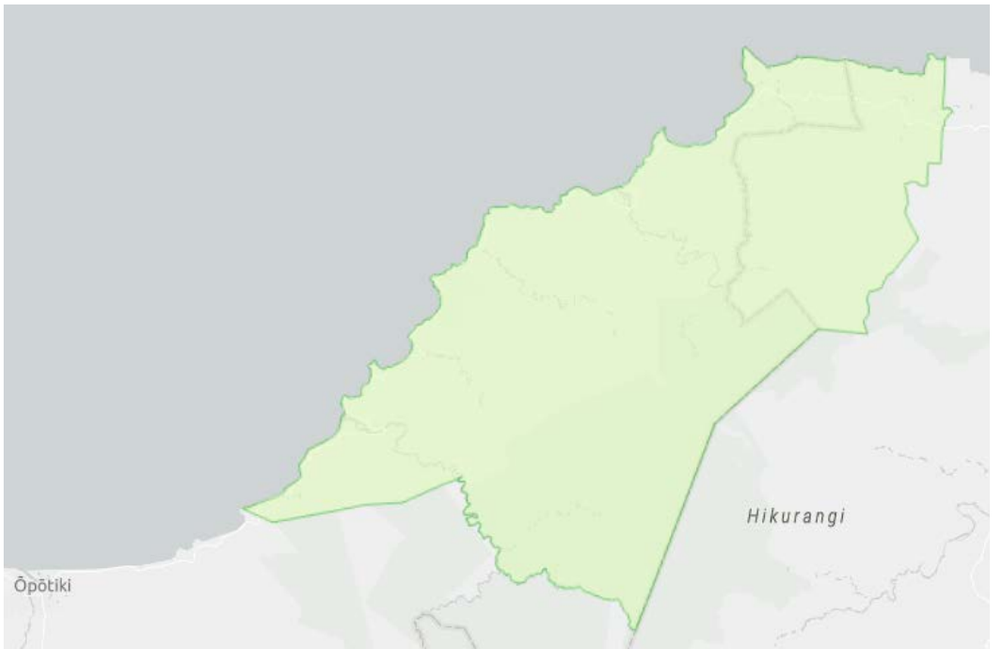
Toitū Tairāwhiti: *Ngāti Porou | Ngāi Tāmanuhiri | Te Aitanga a Māhaki*



Ōhua: *Te Aupōuri | Ngāti Kahu | Te Rarawa | Ngāti Kuri | Ngāi Takoto | Ngāti Kahu ki Whaingaroa* – Ngāpuhi ki *Whaingaroa*



Te Whānau a Apanui: *Hawai | Kahurautao | Kaiaio | Kauaetangohia | Maraenui | Maruhaeremuri | Omaio | Pararaki | Potaka | Rutaia | Te Ehutu | Waiorore | Whitinga*



Below outlines the geographical areas and approximate number of dwellings that the iwi collectives collected Census from:

	Toitū Tairāwhiti	Ōhua	Te Whānau a Apanui
Geographical Area	Central Gisborne	Far North	Tribal boundary of Te Whānau a Apanui
No. of dwellings collected (approx.)	13,000	11,000	1,100

Re-indigenising Census 2023

To re-indigenise the approach to iwi-led collections, the iwi developed their own model that encompassed strategies that would work for them, in their areas, for their people.

"You can't fit Rangi's foot in Cinderella's slipper" (Rikirangi Gage, 2023)

Organisational Structure

The iwi collectives set up a clear organisational structure that linked back into Stats NZ, and then into the Te Mana Whakatipu kaupapa itself. This ensured that there were clear lines of escalation and relationships between the iwi collectives, Te Kāhui Raraunga (as stewards of the kaupapa) and Stats NZ.

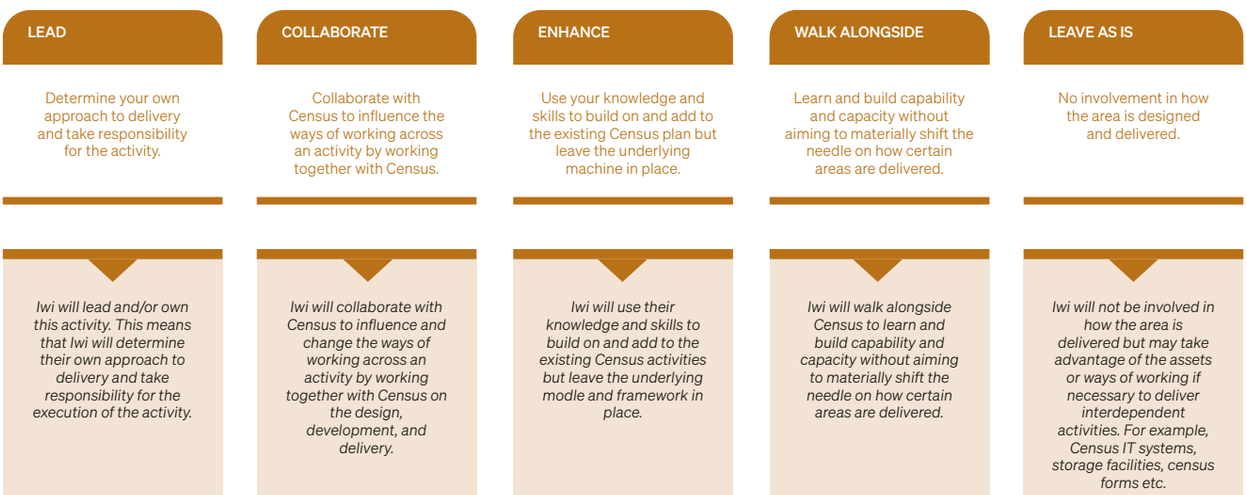
Local Leadership Group (LLG)

The LLG served as a centralised decision-making body for the iwi collectives. These groups ensured that senior leadership input and oversight were applied across all operational aspects of the programme. In turn, ensured that the direction of the project and any key decisions, were made with the values and interests of the local iwi firmly front and centre.

Model of Iwi Level of Involvement in Census 2023

The iwi collectives then developed a method to help determine their level of involvement at each stage of the design and delivery of the Census operation.

The model below shows the method used to help Iwi think about, and determine, their preferred level of involvement in the design and delivery approach of various Census activities. This model was suggested to stimulate dialogue and engagement with the various activities and enable participants to articulate their preferred level of involvement better. This model was not intended to be prescriptive or restrictive. It was acknowledged that all parties may needed to be flexible in their approach to involvement.



Iwi-Led Communications

One area where the iwi collectives played a crucial part, alongside the field collection itself, was communications and engagement. This area was where iwi could make a huge difference in communicating out and showcasing how Census data could help iwi Māori. Alongside other Stats NZ activities, the iwi collectives designed other kaupapa like, whānau days, and pushing communications through iwi-specific channels, ensuring the messages were clear and relevant to the people inside those iwi-led areas.

CALL TO ACTION TOUR

Iwi will work collaboratively with the 2023 Census team to run Call to Action Tours in their rohe.

WHĀNAU DAYS

Iwi may choose to organise and run whānau days as Assisted Completion Events in their rohe.

Iwi will utilise their knowledge of the people and places in their rohe, as well as their COVID 19 vaccination rollout experience to create and execute these events.

STATS NZ ENGAGEMENT

Iwi may choose to work collaboratively with the 2023 Census teams to enhance their existing/previous engagement in their respective rohe.

IWI SPECIFIC CHANNELS

Iwi may choose to work with organisations they have relationships with, who have links to hard-to-reach communities.

This may look like running events together, supporting initiatives (boots on the ground or financial), or other engagement that is thought to be relevant, to build trust with these populations/audiences and to generate awareness of Census and its importance.

Outcome of Iwi-Led Efforts

As a result of the efforts of the iwi collectives in 2023 Census, there was a significant increase in response rates in all iwi-led areas – including both Māori response rates, and overall response rates in those areas.

CASE STUDY 2: He Ara Pūkeko

Te Tirohanga Whānui | Background

After the efforts of the 2023 iwi-led Census collections, Te Mana Whakatipu programme has shifted into strengthening and increasing the overall data analytics capability and capacity for iwi Māori. This evolution has led to the creation and delivery of six unique, data analytics focused initiatives for ngā iwi Māori o te motu. One of these initiatives is, 'He Ara Pūkeko'.

Te Mana Whakatipu Data Analytics Initiatives.

He Ara Pūkeko | Iwi Data Apprenticeship

He Ara Pūkeko provided an opportunity for 8 participants to be part of an apprenticeship programme over 9 months. This kaupapa is aimed at developing Māori apprentices' data analysis skills, their application of mātauranga Māori, and their ability to exercise their tirohanga Māori in data. It combines theoretical learning with on-the-job practical experience to pathway apprentices onto careers in data analysis.

Te Whakatakotoranga | Layout

TAHA MĀORI

Tauira will spend three months with an Iwi Māori organisation where they will gain exposure to the data lifecycle from an Iwi Māori perspective including the expression of te Ao Māori concepts in data.

TAHA KĀWANATANGA

Tauira will dedicate three months to working with a government organisation, acquiring exposure to the data lifecycle from a governmental perspective.

RANGAHAU

Tauira will dedicate three months to a research project, synthesizing the tools and knowledge acquired in the initial six months to create an independent research piece.

Te Hoahoatanga | The Design

"Don't just be a Māori doing things, do things in a Māori way"
(Te Mauri Kingi, all the time)

If we are to radically change the spaces that we operate in, then we need to radically change the ways we develop our people.

By embracing our own ways of knowing, understanding and being, in addition to Western ways of knowing and understanding, we grow capability within our people that enables them to thrive and disrupt.

Ngā Herenga | Connections

TE KĀHUI RARAUNGA

Te Kāhui Raraunga has one of the largest critical masses of Māori data analysts and scientists. We were able to wānanga with our mātanga to understand their perspectives on the unique attributes of Māori data analysts. This fed into the development of our curriculum whose delivery has been supported by those mātanga as well as others.

Key learnings:

- **Me Māori te tūāpapa** | Māori foundations
- **Me ū ki tō ake raina** | playing to your strengths
- **He mea whakahirahira te wānanga** | the importance of wānanga
- **He pae whakaaro** | a continuum of thought

IWI

When thinking about building iwi capacity, consideration needs to be given to what is put in place to continue to strengthen relationships between apprentices and iwi so they both benefit.

Key learnings:

- **Tautoko** | endorsement
- **Mana whakairo hinengaro** | intellectual property
- **Me whai take** | finding purpose and relevance
- **Whakawhanaungatanga** | sustaining ongoing relationships

KĀWANATANGA

Collaboration with the government is crucial for ensuring that iwi data needs are met (Te Kāhui Raraunga, 2021). Working with our government partners to co-design and deliver 'Taha Kāwanatanga' provided the opportunity to execute a partnership based approach to this programme and, importantly, to develop apprentices' data analysis skills.

Key learnings:

- **Te painga o te mahi ngātahi** | the value of working together
- **Me ū ki tō ake raina** | playing to your strengths
- **Me hāngai pū** | ensuring alignment between partners
- **Mana** | dynamics of power

Activity 1: If you were to run a data capability development initiative in your community, how would you re-ground it in your own cultural context?

Resources

Watch 

Papa, R. (2023). Te Mana Whakatipu. *Waatea News*. https://waateanews.com/2023/10/27/rahui-papa-chair-te-mana-whakatipu/

Read 

Te Mana Whakatipu sees Māori census participation rates up to 92 per cent in pilot rohe. *Radio New Zealand*, 1 November 2023. Available at: https://www.rnz.co.nz/news/te-manu-korihi/501473/te-mana-whakatipu-sees-maori-census-participation-rates-up-to-92-percent-in-pilot-rohe

2023 Iwi led census – a resounding success. *Scoop*. Available at: https://www.scoop.co.nz/stories/PO2310/S00172/2023-iwi-led-census-a-resounding-success.htm

Iwi bring the numbers to Census. *Waatea News*. Available at: https://waateanews.com/2023/10/30/iwi-bring-the-numbers-to-census

What the census tells us about the Māori population. *Radio New Zealand*, 29 May, 2024. Available at: https://www.rnz.co.nz/news/te-manu-korihi/518182/what-the-census-tells-us-about-the-maori-population

Explore 

Te whata iwi data platform. https://tewhata.io/

iirc.ac.nz
maramatanga.ac.nz