

Indigenous Sovereignty of Language Data - the Māori example



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Te Taka Keegan
Computer Science
University of Waikato



What is the fundamental issue?

- Science community
 - Open Science
 - Big Data
 - Open Data
- Indigenous community
 - Large collections of indigenous information
 - Lack of awareness, lack of ownership, lack of control
 - Issues are multiplying

WHAT IS **INDIGENOUS DATA**?

Data, information and knowledge, in any format, that impacts Indigenous Peoples, nations, and communities at the collective and individual levels:



Informed by British Columbia First Nations Data Governance Institute - BCFNDGI.COM

USINDIGENOUSDATA.ORG | @USIDSN

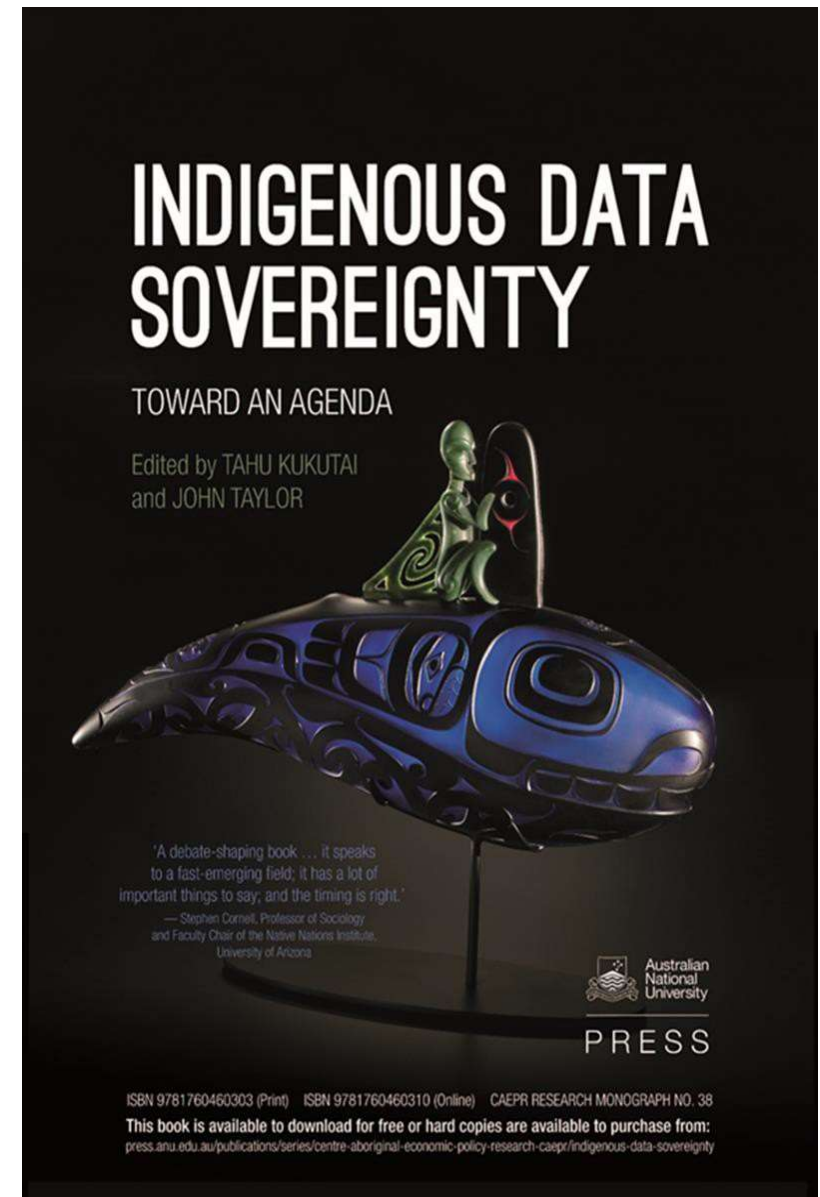


Strengthening Indigenous Governance

INDIGENOUS DATA SOVEREIGNTY

Indigenous Data Sovereignty is a discourse about rights and interests

1. **Data for Governance:** Gaining access to use data to transform the lives of our people.
2. **Governance of Data:** Managing access to data to ensure use is relevant and responsive.





Principles of Māori Data Sovereignty

Tā Māra Raukunga | Brief #1 | October 2020



01 Rangatiratanga | Authority

1.1 Control. Māori have an inherent right to exercise control over Māori data and Māori data ecosystems. This right includes, but is not limited to, the creation, collection, access, analysis, interpretation, management, security, dissemination, use and reuse of Māori data.

1.2 Jurisdiction. Decisions about the physical and virtual storage of Māori data shall enhance control for current and future generations. Wherever possible, Māori data shall be stored in Aotearoa New Zealand.

1.3 Self-determination. Māori have the right to data that is relevant and empowers sustainable self-determination and effective self-governance.

02 Whakapapa | Relationships

2.1 Context. All data has a whakapapa (genealogy). Accurate metadata (story), at minimum, provide information about the provenance of the data, the purpose(s) for its collection, the context of its collection, and the parties involved.

2.2 Data disaggregation. The ability to disaggregate Māori data increases its relevance for Māori communities and lets Māori data shall be collected and coded using categories that provide Māori needs and expressions.

2.3 Future use. Current decision-making over data can have long-term consequences, good and bad, for future generations of Māori. A key goal of Māori data governance should be to protect against future harm.

03 Whanaungatanga | Obligations

3.1 Balancing rights. Individuals' rights (including privacy rights, risks and benefits) in relation to data need to be balanced with those of the groups of which they are a part. In some contexts, collective Māori rights will prevail over those of individuals.

3.2 Accountability. Individuals and organisations responsible for the creation, collection, analysis, management, access, security or dissemination of Māori data are accountable to the communities, groups and individuals from whom the data derive.

04 Kotahitanga | Collective benefit

4.1 Benefit. Data ecosystems shall be designed and function in ways that enable Māori to derive individual and collective benefit.

4.2 Build capacity. Māori Data Sovereignty requires the development of a Māori workforce to enable the creation, collection, management, security, governance and application of data.

4.3 Connect. Connections between Māori and other Indigenous peoples shall be supported to enable the sharing of strategies, resources and ideas in relation to data, and the attainment of common goals.

05 Manaakitanga | Reciprocity

5.1 Respect. The collection, use and interpretation of data shall uphold the dignity of Māori communities, groups and individuals. Data analysis that dignifies or blames Māori can result in collective and individual harm and should be actively avoided.

5.2 Consent. Free, prior and informed consent (FPIC) shall underpin the collection and use of all data from or about Māori. Less defined types of consent shall be balanced by stronger governance arrangements.

06 Kaitiakitanga | Guardianship

6.1 Guardianship. Māori data shall be stored and transferred in such a way that it enables and reinforces the capacity of Māori to exercise kaitiakitanga over Māori data.

6.2 Ethics. Tikanga, haka, protocols and mātauranga knowledge shall underpin the protection, access and use of Māori data.

6.3 Restrictions. Māori shall decide when Māori data shall be controlled, rapid or open (real) access.

¹ <https://www.wgtn.govt.nz/development/maori-indigenous-people/indigenous-people-to-free-prior-and-informed-consent-as-most-genuine-possible-rights-and-a-guard-against-the-real-consent-barrier/>

Principles of Māori Data Sovereignty

01 Rangatiratanga | Authority

02 Whakapapa | Relationships

03 Whanaungatanga | Obligations

04 Kotahitanga | Collective benefit

05 Manaakitanga | Reciprocity

06 Kaitiakitanga | Guardianship

Kaitiakitanga — Guardianship

'Kia u ki te whakapono, kia aroha tetahi ki tetahi.'

Te Hiku Media have developed a Kaitiakitanga licence, which states that data is not owned but as cared for under the principle of kaitiakitanga and any benefit derived from data flows to the source of the data. Kaitiakitanga is a principle that expresses guardianship rather than ownership of data. Te Hiku Media are merely caretakers of the data and seek to ensure that all decisions made about the use of that data respect it's mana and that of the people from whom it descends.

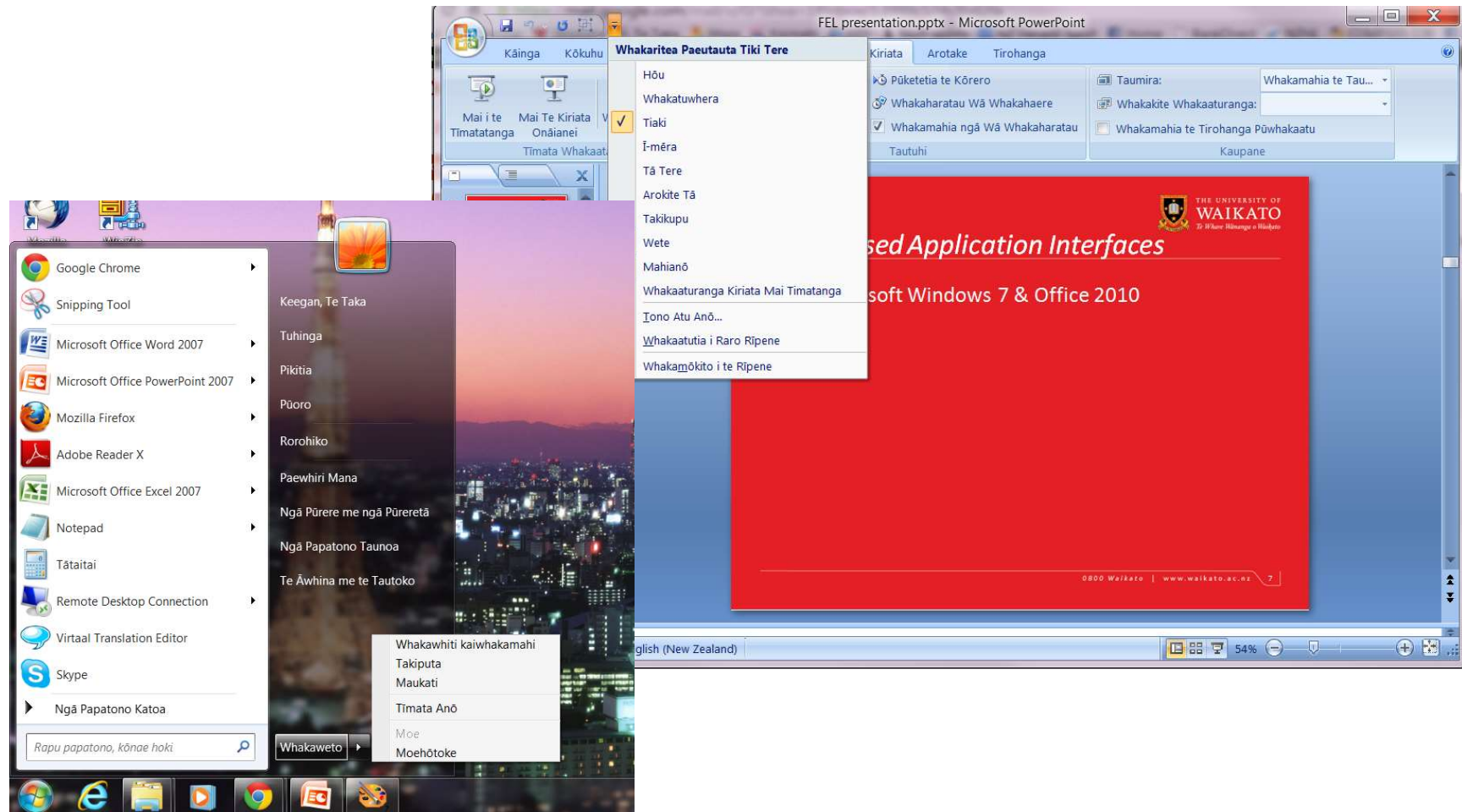
Māori data will not be openly released, but requests for access to the data, or for the use of the tools developed under the platform, will be managed using tikanga Māori. Te Hiku Media have been invited to speak on their kaitiakitanga licence and it has been adopted by a government department and a social enterprise.

Research on other indigenous languages that is carried out under this platform will be for the primary benefit of those peoples. The language platform will also collect data on NZ English and, where appropriate, this will be released under an open-access licence. Machine learning software, that is independent of the language communities, will be made openly accessible where appropriate.

Some Key Data Sovereignty Questions

1. Who's data is it?
2. Who receives benefit?
3. Who will have control over it?
4. Who receives financial gain?

2005 - Microsoft Office and Windows



2009 - Google



Rerenga e Iwa modified Feb 9 by tetaka1

File Edit View Help

Original text: You can allow users to individually select their timezone, or force a timezone for everyone. To use this functionality you need to update your tables with all the tracking information for existing posts. A link from an already existing set of links that are also used by other questions in this category. Wednesday, 20 May, 2009 Wisdom is better than strength, nevertheless the poor man's wisdom is despised and his words are not heard. All go unto one place; all are of the dust, and all turn to dust again. Provide neither gold, nor silver, nor brass in your purses. 20:05:09 How are you? She was concerned about the quality of his work Translation Memories can provide good assistance to Māori

Translation: English (Australia) » Māori 27% complete, 125 words

You can allow users to individually select their timezone, or force a timezone for everyone. To use this functionality you need to update your tables with all the tracking information for existing posts. tētahi hononga mai i tētahi huinga hononga tiariari kē, e whakamahia ana hoki e tētahi atu pātari i tēnei kāwai Wednesday, 20 May, 2009 Wisdom is better than strength, nevertheless the poor man's wisdom is despised and his words are not heard. E haere ana te katoa ki te wāhi kotahi; nō te puehu nei te katoa, ka hoki anō te katoa ki te puehu. kua he parahi ki roto ki ō koutou whāki; work assistance to Māori language translators

E pāwhea ana koe?

Characters: 17

Automatic Translation Search Custom Translation Search

Translation Search Results

How are you? (Exact match) Source: Global, shared TM

E pāwhea ana koe? ☆☆☆☆ No rating

Use translation

How are you? (Exact match) Source: Global, shared TM

Kei te pāwhea koe? ☆☆☆☆ No rating

Use translation

how are you? Source: Global, shared TM

E pāwhea ana koe ☆☆☆☆ No rating

Use translation

Whakamāori Google

Secure | <https://translate.google.com/#auto/mi/What%20is%20the%20time%3F>

Apps 25 Calendar HRRR contacts DataSheet iWaikato Moodle Kaimahi Stuff NZH Other bookmarks

Google

Whakamāori Whakawetohia te whakamāori aunoa

Ingarihi Pāniora Wiwi Ingarihi - kua kitea

Ingarihi Māori Wiwi Whakamāori

What is the time? 17/5000

He aha te wa?

Tiro hoki: the, time, is

Google Whakamāori mō te Pakihi Kete Utauta Kaiwhakamāori Whakamāori Tukutuku

A Māori Language Corpus

- Sourced from my own collections, language friends, public sources, Taura Whiri, Muka, Mātāpuna sources, NZDB, MBC, dictionaries, paipera tapu, etc.
- 3,750,709 words of raw text
- 1,835,001 words of translations
- 679,977 words in dictionaries

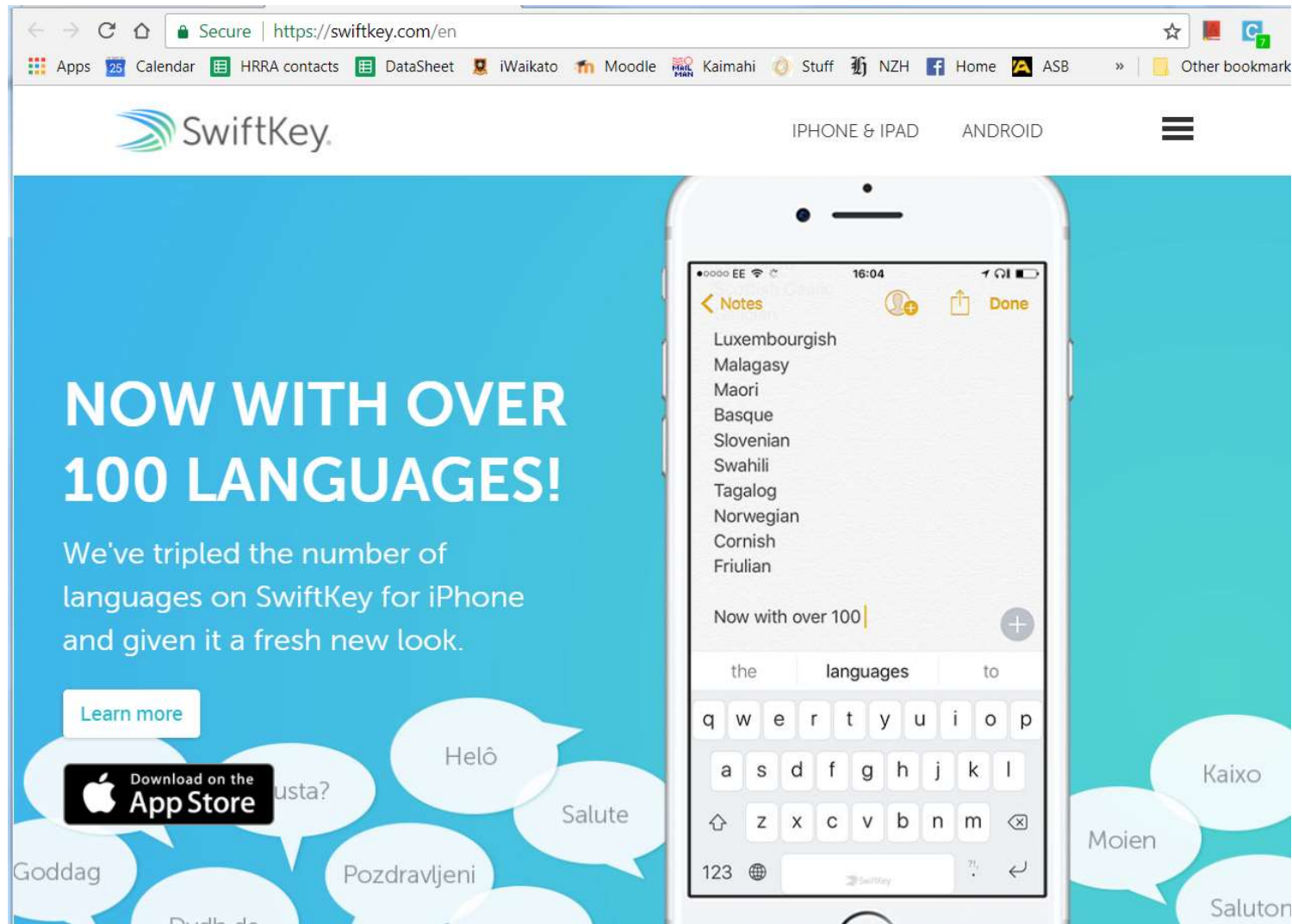
2012 - Macroniser

The screenshot shows a web browser window with the address bar displaying `community.nzdl.org/macron-restoration/jsp/servlet/DirectInput`. The browser's toolbar includes various icons and a search bar. Below the browser window, the web application header features the logo of the Māori Language Commission (Ngā Pae o te Māramatanga) and the title "The Māori Macron Restoration Service" with the subtitle "Automatically add macrons to Māori documents".

The main content area has two tabs: "Add Macrons by Direct Input" (selected) and "Add Macrons by File Upload". Under the "Add Macrons by Direct Input" tab, there is a text input field labeled "Enter the Text" containing the text "ka noho nga wahine ki te rongo i nga korero". To the right of the input field, the text "ka noho ngā wāhine ki te rongo i ngā kōrero" is displayed, showing the result of the macronisation process.

At the bottom of the form, there is a section for "Advanced Options" and a red button labeled "Add Macrons".

2017 - SwiftKey



Some Key Data Sovereignty Questions

1. Who's data is it?
2. Who receives benefit?
3. Who will have control over it?
4. Who receives financial gain?