How Indigenous communities in New Zealand are protecting their data

TAHU KUKUTAI 0000-0001-5080-2296

Concerns about the ethical use of data, privacy, and data harms are front of mind in many jurisdictions as regulators move to impose tighter controls on data privacy and protection, and the use of artificial intelligence (AI). Although efforts to hold corporations to account for their deployment of data and data-driven technologies have been largely welcomed by academics and civil society, there is a growing recognition of the limits to individual data rights, given the capacity of tech giants to link, surveil, target, and make inferences about groups. Questions about whether collective data rights exist, and how they can be recognized and protected, have provided fertile ground for researchers but have yet to penetrate the broader discourse on data rights and regulation.

The notion that groups have rights and responsibilities is nothing new to Indigenous peoples—the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) is fundamentally about upholding collective rights. It is thus unsurprising that Indigenous data sovereignty is largely concerned with the rights of collectives—rather than individuals—in relation to their data.

In recent years, the small island nation of Aotearoa (New Zealand) has become a hub for efforts to assert collective Indigenous data rights. There are several reasons for this. One has to do with the Indigenous landscape. The Māori people are a substantial demographic, comprising around 17% of the population. While Māori (like Indigenous peoples in other settler-colonial states) are overrepresented in all measures of social and economic disadvantage, Māori economic development has become a key part of the national economy, and Māori "authorities" (kinship-based entities) and small and medium-sized enterprises have higher rates of innovation and research and development than Aotearoa firms generally

In the absence of a formal constitution, the 1840 Te Tiriti o Waitangi (Treaty of Waitangi) is considered the country's founding document. Among other things, the Treaty upholds Māori authority over their "taonga"—treasures of tangible and intangible value. Increasingly, data are seen as a modern taonga, a view shared by the Māori data sovereignty network and the national iwi (tribal) leaders forum. Tribes are a considerable political and economic force, partly owing to contemporary Treaty-based settlements, and the vast majority of Māori record at least one tribal affiliation in the census.

Aotearoa is also one of the world's most digitally advanced countries, and the government has invested heavily in data integration. This proved useful when operational failures with the 2018 Census meant that 16% of the total population, and nearly one-third of Māori, were missed. The national statistics office, Stats NZ, was able to deploy its data linkage capabilities to retrospectively add people and households to the census dataset from other government administrative data. Data sharing between agencies is helped by the country's unicameral political system and low level of perceived corruption.

Together, these structural features make for a context that is conducive to the assertion of Indigenous data rights and Indigenous-led innovation. One such innovator is Te Hiku Media, a charitable media organization set up by the tribes of the far north region more than 30 years ago, to revitalize and promote the Māori language. Suppressed by the government for many decades, the Māori language has been the focus of revitalization efforts since the 1970s and

is [legally recognized as a taonga](#) that belongs to the Māori people. To aid its language efforts, Te Hiku has developed AI tools including a natural language processing tool, a Māori language pronunciation app "Rongo," and a media platform app "Whare Kōrero" to which tribal radio stations can upload their content. It has also designed [Kaitiakitanga Licenses](#) (stewardship licenses) for both of its apps as well as the "Papa Reo" application programming interface (API) that Rongo uses to process user data to enable features such as speech detection. The licenses draw on tribal tikanga (protocols) to specify conditions of data access, sharing, and use. The API license prohibits its use for surveillance, tracking, building Māori corpora, mining Māori data, and anything that is inconsistent with [Māori data sovereignty principles](#). To gain access to the API, all interested parties must sign a legally binding agreement in which they agree that they won't use the tools for any of the prohibited uses.

Whakapapa (genealogical) data are also highly valued by Māori and are widely considered to be tapu (sacred). [Whakapapa](#) includes ancestry and [DNA](#) data along with Traditional Knowledge (TK) connecting humans to their wider environments, including ancestral mountains and [rivers](#), as well as flora and fauna. Māori tech company [Āhau](#) has developed a decentralized data platform that helps whānau (families) and tribal communities capture, preserve, and share whakapapa and other forms of collective information in secure, whānau-managed databases. Using a free, open-source app, whānau are able to record their own ancestry data and share cultural knowledge to preserve its intergenerational transmission. All of the information is encrypted on the users' devices or to their own kinship group "pātaka" (storehouse), a computer connected to the internet that serves as a communication link between devices. The company itself does not have servers or databases that keep a record of any information shared by whānau—an approach that departs sharply from the profit-driven model of genealogical websites and genetic genealogy services.

Data are a key focus of the national tribal leaders forum and a charitable trust, [Te Kāhui Raraunga](#) (TKR), was established in 2019 to progress data initiatives on its behalf. Data from Māori are collected in much the same way as from other New Zealanders, through participation in Stats NZ surveys (required by law) and through contact with government agencies (e.g., health and education services). Tribal affiliation information is also captured in the census and in some other government datasets. However, until fairly recently, tribes and Māori communities struggled to gain access to the data needed to support their own goals and decision-making.
One TKR initiative designed to meet tribal governance needs is a data platform called "Te Whata" (a whata being a noncarved Māori storehouse). Most of the data accessed through Te Whata are provided by government agencies, including Stats NZ. Te Whata also has closed-access functionality that enables tribal information managers to add their own information and analysis, and to integrate the data into their reporting. During the COVID-19 pandemic response, tribes were able to access aggregated data on Māori vaccination rates through Te Whata so that they could target their vaccination outreach efforts at the neighborhood block level. A closed-access vaccine data hub was recently added to provide tribes with data insights into the uptake of key childhood immunizations. Te Whata also provides tribes with access to interactive maps and toggles to explore their ancestral lands and sites of cultural significance.

The lack of enforceable intellectual property (IP) rights can be a formidable barrier for Indigenous peoples wanting to exercise sovereignty over their collectively owned IP, including cultural knowledge and heritage, as affirmed by the UNDRIP. In Aotearoa, this remains an [ongoing concern.](#) Originating in the US, the TK labels were designed as an extralegal tool

for communities to add provenance information and contextual metadata to their cultural heritage records, and to assert their rights in relation to them. Stewarded by the Local Contexts Hub, the labels are digital tags that can be attached to cultural material held in public repositories. The labels provide community-specific conditions regarding access and use, in ways that uphold their own protocols and values. One notable example is the TK labels applied to the 1890 wax cylinder recordings of the Passamaquoddy tribe of northeastern North America in the Library of Congress. Although the TK labels are not legally enforceable, they enable communities to assert their collective rights in relation to their cultural heritage and provide a ready resource for institutions and researchers wanting to do the right thing.

Building on the TK labels, the Biocultural (BC) labels recognize Indigenous rights in research data, including digital sequence information, on genetic resources. Codesigned by Māori and US-based researchers, the 10 labels contain community information relating to provenance, protocols, and permissions and make explicit their expectations and consents for appropriate data use. For example, the BC Consent verified label indicates that a community has consent conditions in place for the use of the information, while the Open to collaboration and Open to commercialization labels indicate a community's willingness to engage and collaborate. In instances where several communities have rights in relation to the information, the Multiple communities label can be used. In Aotearoa, three tribes have applied BC labels to a nationally important biodiversity database that includes plants, fungi, and microbial specimens. Another tribe has applied the BC Provenance label to eDNA samples (DNA shed by organisms as they move through the environment) collected by a commercial DNA testing laboratory as part of a national environmental monitoring project. Beyond ensuring proper attribution and protocols, the labels can be a tool for more equitable benefit sharing with Indigenous communities, particularly in the context of commercialization.

Although some of the examples given here may be considered to be at the "soft" end of data rights, rather than formal "hard" law, they nevertheless provide a foundation for future collective action that is more ambitious in scope. Lacking a Māori data sovereignty law, or the recognition of collective privacy rights in Aotearoa's Privacy Act, the best line of defense against data misuse and exploitation may well be Māori communities themselves.