Indigenous data sovereignty—A new take on an old theme

TAHU KUKUTAI 0000-0001-5080-2296
SCIENCE
1 Dec 2023
Vol 382, Issue 6674
DOI: 10.1126/science.adl4664

A new kind of data revolution is unfolding around the world, one that is unlikely to be on the radar of tech giants and the power brokers of Silicon Valley. Indigenous Data Sovereignty (IDSov) is a rallying cry for Indigenous communities seeking to regain control over their information while pushing back against data colonialism and its myriad harms. Led by Indigenous academics, innovators, and knowledge-holders, IDSov networks now exist in the United States, Canada, Aotearoa (New Zealand), Australia, the Pacific, and Scandinavia, along with an international umbrella group, the Global Indigenous Data Alliance (GIDA). Together, these networks advocate for the rights of Indigenous Peoples over data that derive from them and that pertain to Nation membership, knowledge systems, customs, or territories. This lens on data sovereignty not only exceeds narrow notions of sovereignty as data localization and jurisdictional rights but also upends the assumption that the nation state is the legitimate locus of power. IDSov has thus become an important catalyst for broader conversations about what Indigenous sovereignty means in a digital world and how some measure of self-determination can be achieved under the weight of Big Tech dominance.

Indigenous Peoples are, of course, no strangers to struggles for sovereignty. There are an estimated 476 million Indigenous Peoples worldwide; the actual number is unknown because many governments do not separately identify Indigenous Peoples in their national data collections such as the population census. Colonial legacies of racism; land dispossession; and the suppression of Indigenous cultures, languages, and knowledges have had profound impacts. For example, although Indigenous Peoples make up just 6% of the global population, they account for about 20% of the world's extreme poor. Despite this, Indigenous Peoples continue to assert their sovereignty and to uphold their responsibilities as protectors and stewards of their lands, waters, and knowledges.

With the increasing datafication of all aspects of human (and increasingly nonhuman) life, Indigenous Peoples are also having to defend their information from new forms of exploitation. This terrain is not entirely unfamiliar. There are numerous cases of unethical research practice involving Indigenous Peoples and their data, with a prominent case involving the misuse of DNA samples from the Havasupai tribe. However, the explosion of big data, open data, and data technologies has greatly amplified the risks of Indigenous data misappropriation and misuse.

IDSov confronts these practices head on. Most IDSov networks have developed their own data sovereignty principles, guidelines, and tools. The First Nations Information Governance Centre in Canada were the first to set the wheels of IDSov in motion with the OCAP principles, which assert First Nations ownership, control, access, and possession of their data. The GIDA CARE Principles for Indigenous Data Governance front-foot the inherent tensions with open data, noting that "[e]xisting principles within the open data movement … primarily focus on characteristics of data that will facilitate increased data sharing among entities while ignoring power differentials and historical contexts." CARE has been affirmed or adopted by the Research Data Alliance, United Nations Educational, Scientific, and Cultural Organization (UNESCO) Recommendation on Open Science, and Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) Code of Ethics for Aboriginal and Torres Strait Islander Research, among others. Whereas data justice movements tend to focus on the protection of

individual data rights, IDSov is more focused on protecting and upholding collective rights. All of the CARE principles (collective benefit, authority to control, responsibility, and ethics) speak directly to collective rights and responsibilities. The Māori data sovereignty principles go one step further, stating that in some data contexts, "collective Māori rights will prevail over those of individuals." Data that have a clear collective dimension include DNA and genomic data, tribal nation data, and traditional knowledge.

The ethos of collective, intergenerational stewardship is another key feature of IDSov. As Northern Cheyenne IDSov scholar Desi Rodriguez-Lonebear writes, "our peoples have always been data gatherers." From the winter counts of Native American northern plains tribes, to Pacific voyaging and astronomical traditions, the collection and transmission of information and knowledge were crucial for the survival of Indigenous Peoples and their lifeways. Cultural protocols safeguarded how information was stored, shared, and used. Many of these protocols and values continue to guide IDSov to shape data practices that are respectful and aligned with the worldviews and priorities of the communities to whom the data relate. Collective consent is a core concern, reflecting the importance of free, prior, and informed consent (FPIC) in wider Indigenous rights discourse. FPIC is crucial for ensuring that Indigenous communities are fully informed about the purpose of data collection and how the data will be disclosed and used, including secondary uses by humans and machines.
In asserting their rights, IDSov networks leverage the distinctive rights of Indigenous Peoples enshrined in domestic treaties and recognized in global human rights instruments such as the UN Declaration on the Rights of Indigenous Peoples (UNDRIP). IDSov is seen not only as an extension of more substantive forms of Indigenous sovereignty but also as a critical enabler of Indigenous nation-building, wellbeing, and development. The UN Special Rapporteur on the right to privacy has called data a "cultural, strategic, and economic resource" for Indigenous Peoples and called on governments and corporations to "recognize the inherent sovereignty of indigenous peoples with respect to data about them or collected from them."

As yet, there are few tangible examples of regulatory mechanisms that enable Indigenous Peoples to directly control and benefit from their information. In Aotearoa, the Māori Data Governance Model was designed by Māori data experts for use by public service agencies as part of a treaty-based relationship agreement between the national tribal forum and the National Statistics Office, Stats NZ. Underpinned by the vision of "data for self-determination," the model sets clear expectations for the system-wide governance of Māori data and provides direction on the actions needed to meet them. For example, in contrast to the collection and production of Māori "deficit" data for government policy agendas, the model stipulates that the collection of Māori data should prioritize Māori and tribal information needs, uphold communal values of respect and dignity, and embed FPIC. However, the model has yet to be fully implemented, and in the absence of enabling IDSov legislation, there are concerns that agencies will do little to change poor data practices. In Canada, the federal government has invested considerably in the First Nations Data Governance Strategy that provides for a national network of regional First Nations information governance centres. Unlike Aotearoa, First Nations have the UNDRIP Act and its associated UNDRIP Act Action Plan to lean on as a mechanism for holding the government to account on IDSov matters.

Ultimately, the promise of IDSov may be realized more through bottom-up collective action and innovation than top-down regulation. And there is no shortage of examples. In Aotearoa, for example, a small nonprofit Māori radio station Te Hiku Media has built its own speech-to-text engine to transcribe hundreds of hours of Māori language from elders and other community members as part of its homegrown Māori language revitalization efforts. At the same time, it has taken aim at Open AI and Lionbridge for digital misappropriation of Māori

language data—which it describes as the last frontier of colonization. Scientific American has featured a number of Indigenous initiatives developing data storage technology to push back against privacy-invasive apps and give their users and communities greater privacy and control.

Although there is much at stake for Indigenous communities, many of the aspirations and values that drive IDSov have wider relevance for ethical and equitable data practices worldwide.