

## CARE-ing for Indigenous nonhuman genomic data — rethinking our approach

[TAHU KUKUTAI](#) AND [AMANDA BLACK](#)

SCIENCE

1 Aug 2024

Vol 385, Issue 6708

DOI: [10.1126/science.adr2493](https://doi.org/10.1126/science.adr2493)

Earlier this year, an international group of scientists published a paper in *ScienceAdvances* on the [draft genome](#) of the little bush moa (*Anomalopteryx didiformis*), one of about nine species of extinct flightless birds endemic to Aotearoa (New Zealand). The genome was sequenced from the ancient DNA of a “[poorly provenanced](#)” fossil bone acquired by the Royal Ontario Museum. It held important clues about the moa’s evolutionary history and aspects of its biology.

Recognizing that Indigenous Māori are the kaitiaki, or guardians, of these ancestral treasures, the authors deposited the sequence data to the [Aotearoa Genomic Data Repository](#) (AGDR), where access is controlled to uphold Indigenous data governance. However, conforming with “Global North” norms, the sequence data had also been deposited in the open access [GenBank®](#), a public database of nucleotide sequences run by the National Centre for Biotechnology Information. The moa example is illustrative of a wider challenge that Indigenous communities face in trying to exert sovereignty over their genomic data: that control is often illusory, even when scientists are acting with good intent.

Indigenous rights to and interests in nonhuman genomic data are recognized by the Convention on Biological Diversity (CBD) and its [Nagoya Protocol](#). Indigenous peoples are [often described](#) as communities, peoples, and nations that have a historical continuity with precolonial societies that developed on their territories and consider themselves distinct from the dominant societies in which they live. The CBD is the international legal instrument for the conservation of biological diversity (genetic resources from plants, nonhuman animals, and microorganisms), whereas the protocol is a global agreement dedicated to realizing its access and benefit-sharing obligations. Efforts are also underway to establish a [separate, multilateral benefit-sharing](#) mechanism for the use of digital sequence information on genetic resources. To date, the promise of the protocol for Indigenous peoples has been hampered by a [range of challenges](#), including high transaction costs, low levels of legal certainty, a narrowly defined benefit-sharing base, and no guarantee of benefits. In addition, the United States is the only United Nations member state that has not yet ratified the CBD.

Unsurprisingly, the sharing of nonhuman genomic data remains problematic for Indigenous peoples around the world as well as for peoples in the Global South. With continued use of public data repositories, research biases are exacerbated and historical inequities perpetuated, pushing Indigenous researchers and those working with Indigenous communities to the [periphery](#). In recent years, genomic studies involving ancient nonhuman organisms have proliferated, made possible by advances in high-throughput sequencing and a ready availability of specimens. A good many of these specimens hold cultural importance for Indigenous peoples or have been sourced from Indigenous lands. This is not confined to relics of the past. The growing popularity of environmental DNA in biodiversity protection, for example, [has increased the frequency and scale of sampling and data generation from Indigenous lands](#). The data generated from such activities can range from important diagnostics for managing unwanted organisms to a bioprospecting goldmine for new biological tools.

Within the international scientific community, data produced from nonhuman genomic studies are routinely shared to large-scale data repositories for others to use for future research. Data sharing is often a mandated requirement from funders, as well as [journals](#). The rationale for this is [well traversed](#).

In the past, such requirements were rarely questioned. But the terrain is shifting amidst a growing recognition that the benefits of open science are unequally distributed and, in some instances, can erode inclusion or cause [harm](#). Indigenous scientists and those engaged with [Indigenous data sovereignty](#) are leading the charge for a more balanced approach to data sharing, including the implementation of Indigenous governance.

To date, the governance of Indigenous genomic data has been mainly concerned with protecting Indigenous human genomic information from extractive practices and further misuse. Indigenous-controlled [biorepositories](#) have been established to ensure that advances in genetics and health research benefit the Indigenous communities from which the data derive. One such repository is the [Native BioData Consortium](#) (NBDC). Located in the sovereign territory of the Cheyenne River Sioux tribe, the NBDC is the first nonprofit research institute and biorepository in the United States led by Indigenous scientists and tribal members. Key features include keeping samples and data local, building tribal capacities in STEM (science, technology, engineering, and mathematics), and doing research that benefits tribes. We need similar efforts to empower Indigenous governance of nonhuman genomic data.

Research data repositories still largely fail to recognize the cultural importance of nonhuman genomic data and do not have standards and processes that respect those relationships. In Indigenous world views, human relations with rivers, mountains, flora, and [fauna](#) are just as important as interpersonal connections owing to the intrinsic interconnectedness between all living and nonliving entities. Ecological relationships are naturally extended from the interconnectedness between people and are key to cultural identity, [placing individuals and groups within the wider ecosystem](#). In the same way that the [buffalo is seen as kin to Lakota](#), and [arthropods to Kānaka Ōiwi](#) (Native Hawaiians), so too are [moa to Māori](#).

One way forward is to diversify the options for Indigenous governance. In Aotearoa, the AGDR provides a secure within-nation option for the storage, management, and sharing of nonhuman genomic data generated from biological and environmental samples originating in Aotearoa. The database supports the [CARE](#) principles (collective benefit, authority to control, responsibility, and ethics) for Indigenous data governance alongside the more familiar FAIR principles (findable, accessible, interoperable, and reusable). Māori have oversight of data from culturally important species, and decisions about who can access those datasets are decided entirely by kaitiaki. For example, access to genome sequences for samples from the flowering plant mānuka (*Leptospermum scoparium*) collected in the rohe (tribal lands) of Rangitāne o Manawatū must be authorized by the tribe. Access to the genome sequences of the endangered giant parrot known as kākāpō (*Strigops habroptilus*) is managed by the conservation arm of the government in partnership with the Ngāi Tahu tribe. Although increasing numbers of researchers in Aotearoa are using AGDR for storing and sharing genomic data, this needs to expand to international researchers generating data from samples originating in Aotearoa.

As well as infrastructure that enables different modes of data sharing, the genomics community needs [policies](#) that recognize the inherent sovereignty of Indigenous peoples and provide meaningful mechanisms for them to be involved in decision making over their genetic resources. International journals also have a role to play by providing clear guidelines and alternatives to the mantra of open data. The implementation of the CARE principles should be embedded alongside FAIR data [to ensure that the use of Indigenous genomic data upholds Indigenous rights](#). [Biocultural \(BC\) Labels](#) can also be used to define community expectations about

appropriate use of biocultural collections and data, including accurate provenance and representation of Indigenous rightsholders in the metadata.

In reimagining the status quo, there must also be a place for the repatriation of data; this goes to the heart of Indigenous data sovereignty. Lessons can be learned from the reckoning of the cultural heritage sector with its history of extraction and imperialism and the [return](#) of thousands of human remains back to lineal descendants or descendant communities. In the United States, [new federal law](#) mandates that museums obtain free, prior, and informed consent from lineal descendants, American Indian Tribes, and Native Hawaiian organizations before they exhibit human remains or cultural items. A [recent commentary](#) on the legacy of palaeontology expeditions in Native American lands argued that fossils of the extinct ungulates (*Brontotheriidae*) taken from the lands of the White River Badlands should be returned to that community. The key difference between repatriating data versus physical artifacts is that data remaining on public repositories can be duplicated, whereas an artifact cannot. Data repatriation will also be technically challenging owing to the additional infrastructure, expertise, and data security needs.

Scientists working with nonhuman species that connect to or come from Indigenous communities ought not be absolved from ethical standards and responsibilities to which their colleagues who work with artifacts and human remains must attend. Given the long and sorry history of exploitation and extraction from Indigenous lands and bodies, there are exciting opportunities to think differently and do better with Indigenous data.