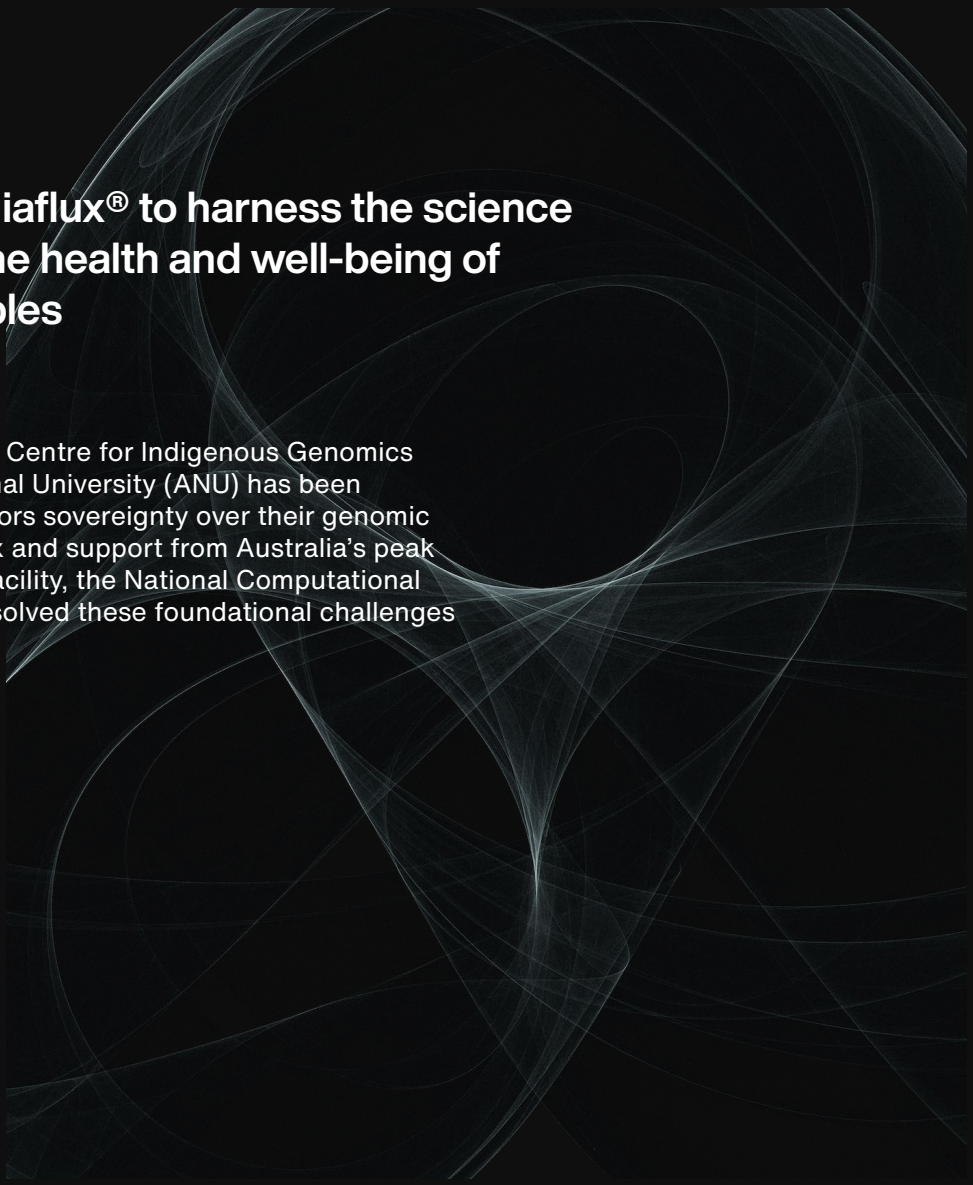
CASE STUDY

NCIG leverages Mediaflux® to harness the science of DNA to improve the health and well-being of Australia's First Peoples



For over 30 years, The National Centre for Indigenous Genomics (NCIG) at The Australian National University (ANU) has been working to give Indigenous donors sovereignty over their genomic data. With Arcitecta's Mediaflux and support from Australia's peak supercomputing and big data facility, the National Computational Infrastructure (NCI), NCIG has solved these foundational challenges of ethical genomic science.

www.arcitecta.com
talk@arcitecta.com

What do NCIG do?

NCIG is an Indigenous-led research organisation charged with managing over 7,000 human blood samples, their accompanying genomic data, and the associated archival documents from the 1960s to the 1990s when the samples were collected.

NCIG brings Indigenous decision-making into a high-performance biomedical research environment to make their genomic data resources available to researchers for a wide range of Indigenous-focused health and medical research, subject to appropriate access mechanisms. NCIG ensures that Australia's first people are included in the design and implementation and share in the benefits of this important frontier of medical research and discovery.

The Challenge

NCI's technical capacity and NCIG's robust Indigenous-led governance were needed to develop a platform within a unique cultural context that could solve one of the foundational challenges of ethical genomic science: donor sovereignty over the use of and access to their genomic data. Research into people's attitudes towards the use of genomic information shows that the majority want access to results from genome sequencing1, and there is a desire for this to be handled carefully. Yet the link between donors and the databases holding their genetic sequences is frequently lost. This disconnect occurs because legacy databases are researcher-centric rather than donor-centric, and it isn't easy to retrofit individual donor requirements into existing data management systems.

NCIG Deputy Director Azure Hermes says, "We as Indigenous people understand the importance of protecting our data. The sample we give to researchers is our story, our history, it's a link to our ancestors. But make no mistake, this is not just a problem for Indigenous people. It's an everybody problem. Every single person should ask the question, where is my sample, who is using it and for what purpose?"

Transparency and accountability were central challenges to solve with the data management platform to reflect the desires of donors and their communities. NCI Director Professor Sean Smith says, "First Nations people must be in control of their data. Real and meaningful consent from data owners is a requirement of any Indigenous genomics platform, and NCI is proud to provide solutions that meet the needs of Indigenous Australians and enable directly beneficial research collaborations."

NCIG is an Indigenous-led research organisation charged with managing over 7,000 human blood samples, their accompanying genomic data, and the associated archival documents from the 1960s to the 1990s when the samples were collected.

Why Mediaflux?

Australian National University (ANU), Australia's top university according to the 2021 QS World University Ranking, has been using Mediaflux for projects across the institution since it purchased an organisation-wide unlimited license for the RDSI project in 2014. The RDSI project at ANU used Mediaflux to implement a nationwide infrastructure to encourage access to and reuse data collections of national significance. Following its success, Andrew Howard, the Associate Director of Cloud Services at NCI, said: "When NCIG came to us to support a new donor-centric genomic data platform, it was an obvious choice to use Mediaflux. It certainly ticked all the boxes."

The Journey

The stewardship of the 7,000 Aboriginal blood samples extends back three decades. This rare and precious collection, holding in its DNA the life stories of thousands of indigenous Australians, had been preserved and protected by NCIG's founding Director, Professor Simon Easteal's, since the 1990s. It was Easteal's steadfast commitment to doing the right thing by those samples and Aboriginal people as a whole that led to the establishment of NCIG under his leadership in 2013.

Over the past 30 years, NCIG has conducted community-engaged research to develop practical guidelines and policies to assure Indigenous communities that their interests are protected, and to increase the participation of Indigenous leaders, communities, and individuals in the program.

In 2020, NCIG agreed with the participating stakeholders to a set of data permission that would offer donors and their communities a model of self-determination, which means the inherent right and capacity to ensure that data usage is consistent with their respective histories and values.

NCI and NCIG set out to develop a data management platform to support the principles of sovereignty, transparency and accountability before sequencing and migrating the data from consenting samples. Once sequencing was complete, NCIG repatriated the physical DNA samples back to Country to be reunited with the people who provided them.

“Having a stable, mature product that has been well tested in the industry and can deal with petabytes of data natively is a big asset.”

Andrew Howard,
Associate Director of Cloud Services at NCI

The Solution

A critical step in creating a donor-centric genomic data platform was to embed donor preferences about the use of their data for research at the repository design level. NCI and NCIG implemented this dynamic consent using existing tools within the Mediaflux platform. Importantly, this work aligned with the metadata standards of the European Genome-phenome Archive (EGA) and the Global Alliance for Genomics and Health (GA4GH) for responsible genomic data sharing within a human rights framework.

“Mediaflux meets these standards right out of the box, so it was one of the first capabilities we turned on to make sure that even if there was a data leak or someone got access to a backup copy, it would be completely encrypted,” says Howard.

As a mature and rich software platform, Mediaflux is used by NCI to curate, manage, protect, and disseminate large volumes of structured and unstructured data across its many service offerings. The Mediaflux platform allows for metadata-driven granular control for data access, tracks the entire data transaction history with detailed audit trails, implements data encryptions and automated workflows through APIs. Mediaflux also manages the underlying storage to tier assets based on policies to ensure that storage is optimised automatically behind the scenes. All this is built petabyte-scale data in mind.

“These capabilities provide the necessary foundation for implementing NCIG’s use case using industry-standard security and leverage a range of NCI storage and compute services for end-to-end data workflows”, says Howard.

And Howard further adds “Having a stable, mature product that has been well tested in the industry and can deal with petabytes of data natively is a big asset. So too is having a real product when you want to pick up the phone to the support team or request new functionality for edge cases which has been invaluable for saving time and development resources.”

The Implementation

An immense amount of work was completed by NCIG over the past 30 years to design a system to support research partnerships that are collectively respectful, ethical, and culturally safe. The technical implementation is a vault in which this precious tribal data and intellectual property is stored, with Mediaflux providing an interface for researchers to access the data where appropriate.

Take, for example, a researcher is conducting a study into a new treatment for diabetes. The individual donor or their community could benefit from consenting to include in the study a genetic sample they have the rights to. Diabetes in their community may be a big problem, so the researcher could then ask, “would you like to participate or not?”. The samples’ owners or families are contacted, and only once they have given permission, their data is visible to the researcher doing the study. Consent then comes down to what kind of benefits researchers can provide to the community by authorising their genomics data in a particular study.

Further to requiring consent, is a secondary mechanism to inhibit the violation of ethics agreements, such as the wrongful selling of the data to an insurance company. Rather than providing researchers access to a data source that is then copied into their work area, this donor-centric model requires researchers to bring their study to NCIG, where the study can be vetted and executed on behalf of the researcher. This directing acts as an intermediary both ways, ensuring that the community’s interests are absolutely maintained and that researchers are only allowed to study what was consented to in their ethics agreement for these data sets. Of course, this extra layer of the workflow may not be necessary with all genomics data, say of plants and animals, but it is critical in the human data space.

The Result

This case study is an example of how an interdisciplinary team led by Indigenous people can foster trust and mutual respect, allowing them to act as cultural navigators of the change that many Indigenous communities and organizations continue to strive for today.

Together, NCIG and NCI have joined to be innovative in this area of national and international significance. Research bodies around the world are working on the challenge of controlled and appropriate access to genomic data collections, and this partnership has provided a key step forward. Future collaborations, including with NCRIS-funded infrastructure platform, the Australian BioCommons, and internationally with the Broad Institute Data Science Platform, will ensure FAIR (Findable, Accessible, Interoperable and Reusable) and CARE (Collective benefit, Authority to control, Responsibility and Ethics) implementations for genomics data.

What is next for NCIG?

“Now that we have completed the pilot with the most complex data, the 7,000 human samples, we’re looking at all the non-human species data as a natural fit for this. That includes around 20,000 plant and animal specimens, about 13Pb of data, that Mediaflux will ingest in the next couple of weeks”, says Howard.