

Te Tiriti-guided national DNA reference library wānanga series

Wānanga 5: Cross-project fertilisation of best practice

The following questions were submitted during our wānanga but due to time restrictions were only addressed in the chat. Some answers have been modified in the interest of clarity.

To Tania Laity/Cam Slatyer: Thanks for sharing the Restricted Access Species framework. For the data service and the secure analysis environment, could you talk to the outcomes and benefit sharing of the federated data / consent system to the specific organisations and communities involved?

It is very early days for the data service. We are still in the process of loading organisational metadata (and will be for months). We've been targeting government and conservation organisations, a) because these are the largest datasets people in Australia wish to access, and b), because we are allowing the First Nations consultation to run. The secure data environment (developed by Griffith) was really developed without much consultation but also has a lot of flexibility as to how security works which I anticipate (once consultation is finished) will be important for communities. For example, the data custodians, not the system administrators could control access.

To Nick Jones/Claire Rye: Is it possible to implement a secure research environment on NeSI like what Cam (Slatyer) described? I.e., where multiple users can access/read remote files but not download or make local copies?

The short answer is yes, but not straight out of the box. We have done this with the Rakeiora project so do have experience building a secure research environment.

Additional comments contributed from participants in relation to the above question

I imagine this could be an important part of making an Ao/NZ reference library work (where a major use-case is identifying unknown DNA sequences that could potentially match to data owned by many different hapū/iwi, but with no way to know who a priori). Perhaps a high-level governance committee could approve applications to query against the library in a secure environment, but then - based on the results - approval could be sought from specific hapū/iwi who are the owners of individual sequences to download, publish, and/or do further work.

I feel this is the model that the Aotearoa Genomic Data Repository (AGDR) is settling on: an AGDR advisory/governance group that might do a first assessment of data requests that come in, and then also nominated kaitiaki, and the optional use of Biocultural Notices and Labels, for specific datasets.

To Sujeevan Ratnasingham: I'm also keen to hear about how this might be similar to implementation within the BOLD Species ID workbench - some data is protected and cannot be transmitted but can be analysed against?

BOLD employs a similar framework to support secure storage, collaboration, and analysis. It's good to see convergence on this strategy.

To Libby Liggins: The point you raised around governance of different data is a really important one. What are your thoughts on how governance over decisions for different datasets (e.g. barcodes vs genomes) may be different or similar?

I cannot speak to what is appropriate, but I can point to the examples out there. For instance, my experience with Genomics Aotearoa would indicate that by default, whole genomes are access-restricted. The European Reference Genome Atlas and Earth BioGenome Project are open access by default. So, there may also be a data type x Indigenous provenance nuance to consider. "Barcodes" by default (or assumption by researchers), I would say, are made open access. I am interested to know what popular opinion is regarding "megabarcodes", genome-skimmed data, reduced representation of genomes (i.e. SNPs)..