

Simon Easteal, Australian National University

National Centre for Indigenous Genomics



Bringing together the world's oldest living culture and its newest science.

John Curtin School of Medical Research



Data referencing in biomedicine started long long time ago

- 1979–82: Los Alamos National Laboratory created Genbank. Funding from NIH, NSF, DoE, DoD.
- 1982: DDBJ, EMBL and Genbank coordination.
- Mid 1980s: Moved to Stanford.
- Open access promoted through BIOSCI/ Bionet news groups (Stanford, Cambridge, ANU)
- 1992: Moved to National Centre for Biotechnology Information, National Library of Medicine.
- 1992: EBI established

Easteal S, Oakeshott JG. (1985). Estimating divergence times of *Drosophila* species from DNA sequence comparisons. *Molecular Biology and Evolution* 2 (2): 87–91.

D.simulans alcohol dehydrogenase gene, exons 1,2 and 3

GenBank: M19276.1

[FASTA](#) [Graphics](#)

Go to:

```
LOCUS      DROADHGAM                      829 bp    DNA        linear    INV 26-APR-1993
DEFINITION D.simulans alcohol dehydrogenase gene, exons 1,2 and 3.
ACCESSION  M19276
VERSION    M19276.1  GI:156897
KEYWORDS   alcohol dehydrogenase; dehydrogenase.
SOURCE     Drosophila simulans
           ORGANISM Drosophila simulans
           Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila; Sophophora.

REFERENCE  1 (sites)
           AUTHORS  Bodmer,M.W.
           JOURNAL  Thesis (1983) University of Cambridge
REFERENCE  2 (sites)
           AUTHORS  Bodmer,M. and Ashburner,M.
           TITLE    Conservation and change in the DNA sequences coding for alcohol
           dehydrogenase in sibling species of Drosophila
           JOURNAL  Nature 309 (5967), 425-430 (1984)
           PUBMED   6427630
REFERENCE  3 (sites)
           AUTHORS  Ashburner,M.
           JOURNAL  Unpublished
COMMENT    Original source text: D.simulans (library of M.Meselson) DNA.
FEATURES   Location/Qualifiers
           source          1..829
                       /organism="Drosophila simulans"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:7240"
           prim_transcript <1..>829
                       /note="adh mRNA and introns"
           CDS             join(127..225,293..697,767..>829)
                       /note="alcohol dehydrogenase"
                       /codon_start=1
                       /protein_id="AAA28360.1"
                       /db_xref="GI:156898"
                       /translation="MAFTLTNKNVIFVAGLGGIGLDTSKELLKRDNLKLVILDRIENP
                       AAIAELKVINPKVTVTFYPYDVTVPIAETTKLKTIFAKLKTVDVVLINGAGILDHDI
                       ERTIAVNYTGLVNTTALDFWDRKGGGGIICNIGSVTGFNAIYQVPVYSGNKAIV
                       VNFTSSLAKLAPITGVTAIXVNPVITRTT"
           exon            <127..225
                       /note="alcohol dehydrogenase"
                       /number=1
           intron          226..292
                       /note="adh intron A"
           exon            293..697
                       /number=2
           intron          698..766
                       /note="adh intron B"
           exon            767..>829
                       /note="alcohol dehydrogenase"
                       /number=3
ORIGIN     1 bp upstream of HindIII site.
           1 aagcttctgc gtacggatac ttctataaaa tacggggccg acacgaactg gaaaccaaca
           61 actaacggag cctctctccc attgaacag atcgaaagag cctgttaag caaaaaagaa
           121 gtcaccatgg cgtttacttt gaccaacaag aacgtgattt tcggtgccgg tctgggaggg
           181 atcggctcgg acaccagcaa ggagctgctc aagcgcgacc tgaaggtaac tatgcatgct
           241 ccacaggttc catggagggg gatggaggtt aatctcgtgt attcaatcct agaacctggt
           301 aatcctcgac cgcattgaga acccggctgc cattgccgag ctgaaggtaa tcaatccaaa
           361 ggtgaccgtc accttctacc cctatgatgt gaccgtgcc attgccgaga ccaccaagct
           421 gctgaagacc atcttcgcca agctgaagac cgtcgtatgc ctgatcaacg gagctggtat
           481 cctggacgat caccagatgc agcgcacat tgccgtcaac tacactggcc tggtaaacac
           541 cacgacggcc attctggact tctgggacaa gcgcaagggt ggtcccgggt gttatcatctg
           601 caacattgga tccgtcactg gtttcaatgc catctaccag gtgcccggtc actccggcaa
           661 caaggccgcn gtggtcaact tcaccagctc cctggcggta agttcatcga aggaaacgca
           721 aagttttcaa aaaataaaaa actaatctgt ttataaacac ctttagaacat tggcccccac
           781 taccggcgtg accgctnaca ccgtgaaccc cggcatacacc cgcaccacc
```

//

D.simulans alcohol dehydrogenase gene, exons 1,2 and 3

GenBank: M19276.1

[GenBank](#) [Graphics](#)

>gi|156897|gb|M19276.1|DROADHGAM D.simulans alcohol dehydrogenase gene, exons 1,2 and 3

```
AAGCTTCTGCGTACGGATACTTCTATAAATACGGGGCCGACACGAACCTGGAAACCAACAACCTAACGGAG
CCCTCTTCCCATTGAAACAGATCGAAAGAGCCTGCTTAAGCAAAAAAGAAGTCACCATGGCGTTTACTTT
GACCAACAAGAACGTGATTTTCGTTGCCGGTCTGGGAGGCATCGGTCTGGACACCAGCAAGGAGCTGCTC
AAGCGCGACCTGAAGGTAACCTATGCGATGCCACAGGTTCCATGGAGGGGATGGAGGTTAATCTCGTGT
ATTCAATCCTAGAACCTGGTAATCCTCGACCGCATTGAGAACCCGGCTGCCATTGCCGAGCTGAAGGTAA
TCAATCCAAAGGTGACCGTCACCTTCTACCCCTATGATGTGACCGTGCCCATTGCCGAGACCACCAAGCT
GCTGAAGACCATCTTCGCCAAGCTGAAGACCGTCGATGTCCTGATCAACGGAGCTGGTATCCTGGACGAT
CACCAGATCGAGCGCACCATTGCCGTCAACTACACTGGCCTGGTCAACACCACGACGGCCATTCTGGACT
TCTGGGACAAGCGCAAGGGTGGTCCCGGTGATCATCTGCAACATTGGATCCGTCACTGGTTTCAATGC
CATCTACCAGGTGCCCGTCTACTCCGGCAACAAGGCCGNGTGGTCAACTTCACCAGCTCCCTGGCGGTA
AGTTCATCGAAGGAAACGCAAAGTTTTCAAAAAATAAAAACTAATTTGTTTTATAACACCTTTAGAAAC
TGGCCCCCATTACCGGCGTGACCGCTNACACCGTGAACCCCGGCATCACCCGCACCACC
```

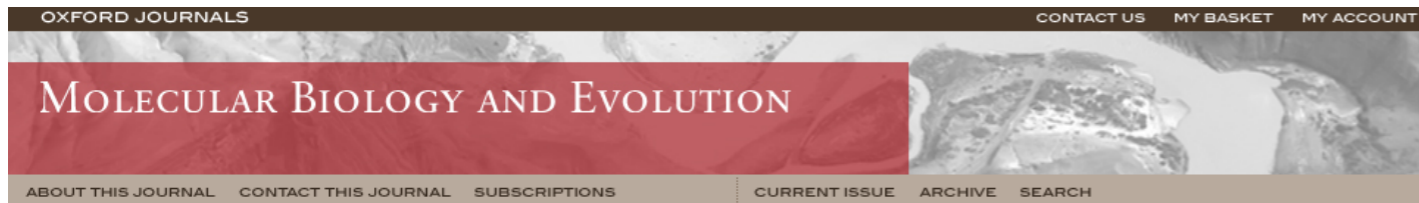
Andrews, TD, Jermiin LS, Easteal S. (1998). Accelerated evolution of the cytochrome b gene of simian primates: adaptive evolution in concert with other mitochondrial proteins? *Journal of Molecular Evolution* 47 (3): 249–257.

Methods

Other cytochrome *b* nucleotide sequences were obtained from the following published sources: *Homo sapiens* (human; J01415) (Anderson *et al.* 1981), *Colobus guereza* (black-and-white colobus monkey; U38264), *Saimiri sciureus* (squirrel monkey; U38273), *Lemur catta* (ring-tail lemur; U38271) (Collura and Stewart 1995), *Galago crassicaudatus* (thick-tailed bushbaby; U53579), *Nycticebus coucang* (slow loris; U53580) (Yoder *et al.* 1996), *Felis catus* (domestic cat; U20753) (Lopez *et al.* 1996),

Acknowledgement

The tarsier cytochrome *b* nucleotide sequence has been submitted to DDBJ/EMBL/GenBank and has the accession number AB011077.



Newly reported nucleic acid and amino acid sequences, microarray data, structural coordinates, and all other essential information must be submitted to appropriate public databases (e.g., GenBank; the EMBL Nucleotide Sequence Database; DNA Database of Japan; the Protein Data Bank; Swiss-Prot; GEO; and Array-Express). Methods must be described in sufficient detail to permit independent replication. Standard procedures can be referenced, provided that significant variations are adequately described. Evaluation of large-scale experiments such as transcript profiling using microarray- or deep sequencing-based approaches requires a complete and transparent description of each experiment, of the nature of the replication, and of the statistical analysis, including, for example, whether any multiple comparison correction was applied. Large-scale data sets necessary for peer review must be made available to reviewers at the time of submission.







Sequence alignments can also be deposited at EMBL-ALIGN.

Early open access based on bioinformatics has transformed life sciences. It is the foundation of modern biology.



The *EBI RDF Platform* aims to bring together the efforts of a number of EMBL-EBI resources that provide access to their data using [Semantic Web technologies](#). It provides a unified way to query across resources using the [W3C SPARQL](#) query language. We welcome **comments or questions** via our [feedback form](#).

Current RDF resources

Services	Quick links	Example query
 BioModels	<ul style="list-style-type: none"> ◦ Service description ◦ SPARQL endpoint ◦ Documentation ◦ RDF download 	All model elements with annotations to acetylcholine-gated channel complex (GO:0005892)
 BioSamples	<ul style="list-style-type: none"> ◦ Service description ◦ SPARQL endpoint ◦ Documentation ◦ RDF download 	Samples treated with alcohol
 ChEMBL	<ul style="list-style-type: none"> ◦ Service description ◦ SPARQL endpoint ◦ Documentation ◦ RDF download 	Find drug-like (but currently not approved) molecules which bind 7TM1 GPCRs with high affinity
 Expression Atlas	<ul style="list-style-type: none"> ◦ Service description ◦ SPARQL endpoint ◦ Documentation ◦ RDF download 	Under what experimental conditions is Ensembl gene ENSG00000129991 (TNNI3) expressed?
 Reactome	<ul style="list-style-type: none"> ◦ Service description ◦ SPARQL endpoint ◦ Documentation ◦ RDF download 	Pathways that references Insulin (P01308)
 UniProt	<ul style="list-style-type: none"> ◦ Service description ◦ SPARQL endpoint ◦ Documentation ◦ RDF download 	What are the preferred gene name and disease annotations of all human UniProt entries that are known to be involved in a disease?

New challenges:

Scope

- Many kinds of data

Scale

- 'Genomical' amounts of data

Speed

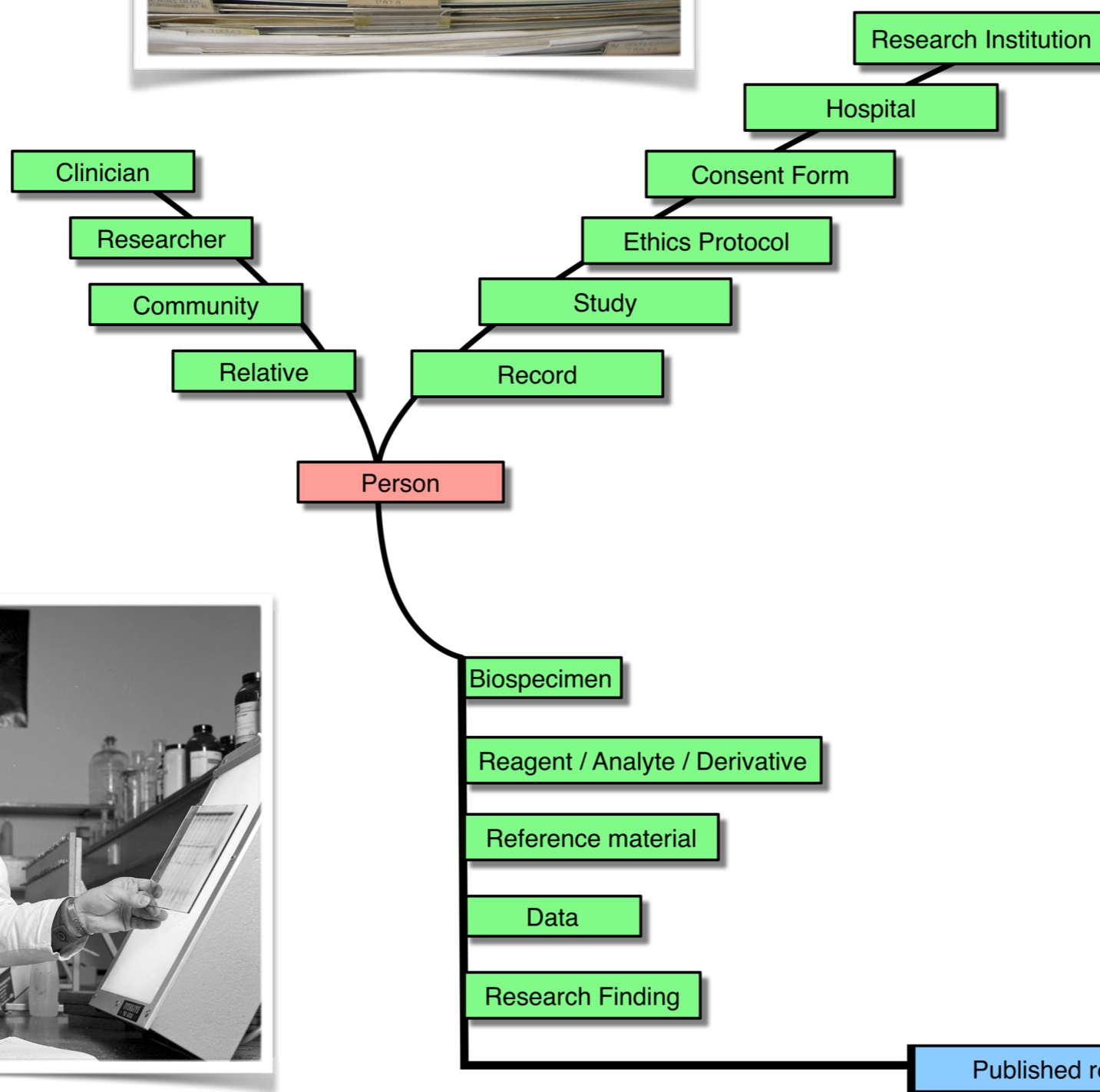
- Doubling time < 1 year

Sophistication

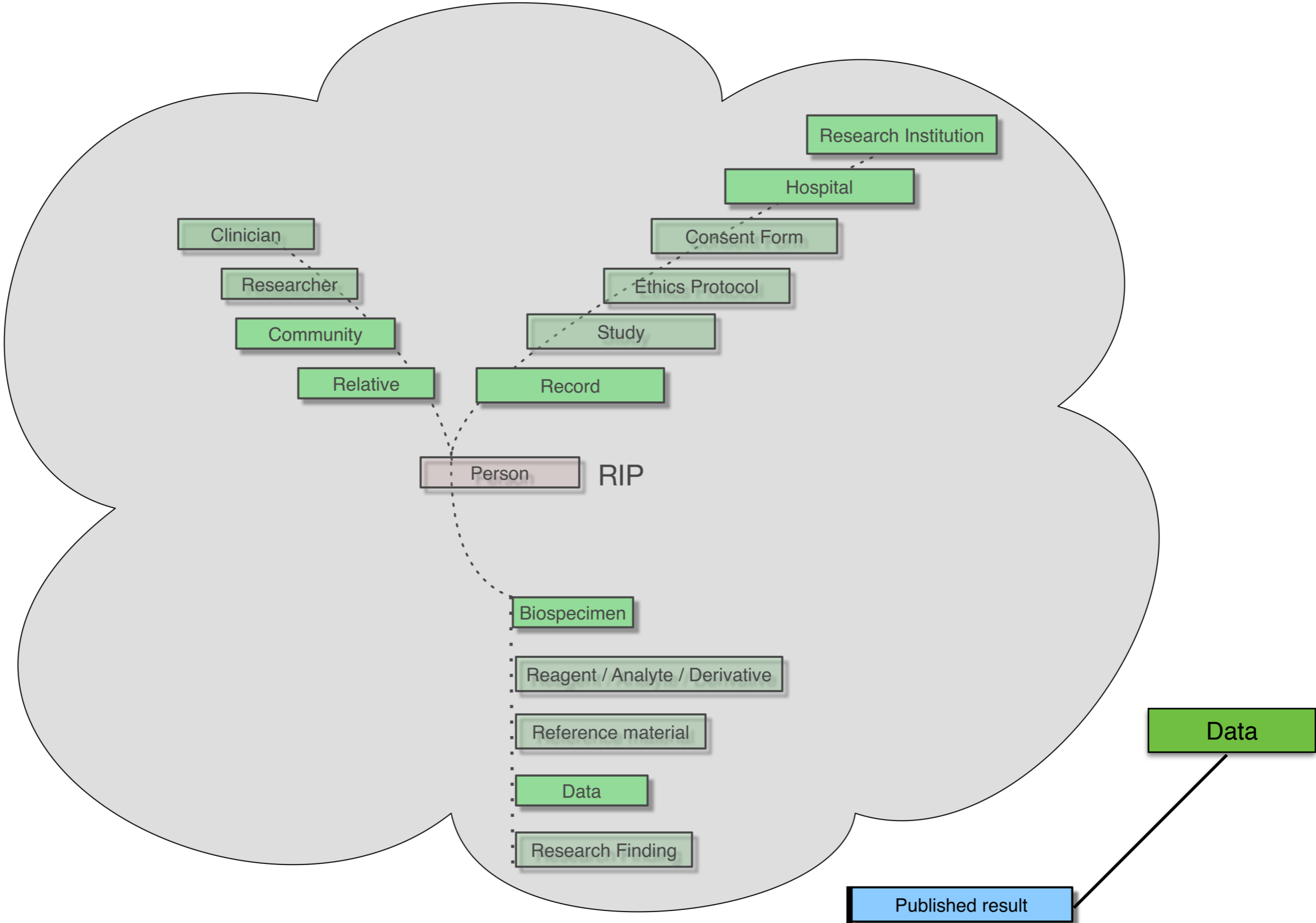
- Non-linearity, interactions, dependencies, error

Source

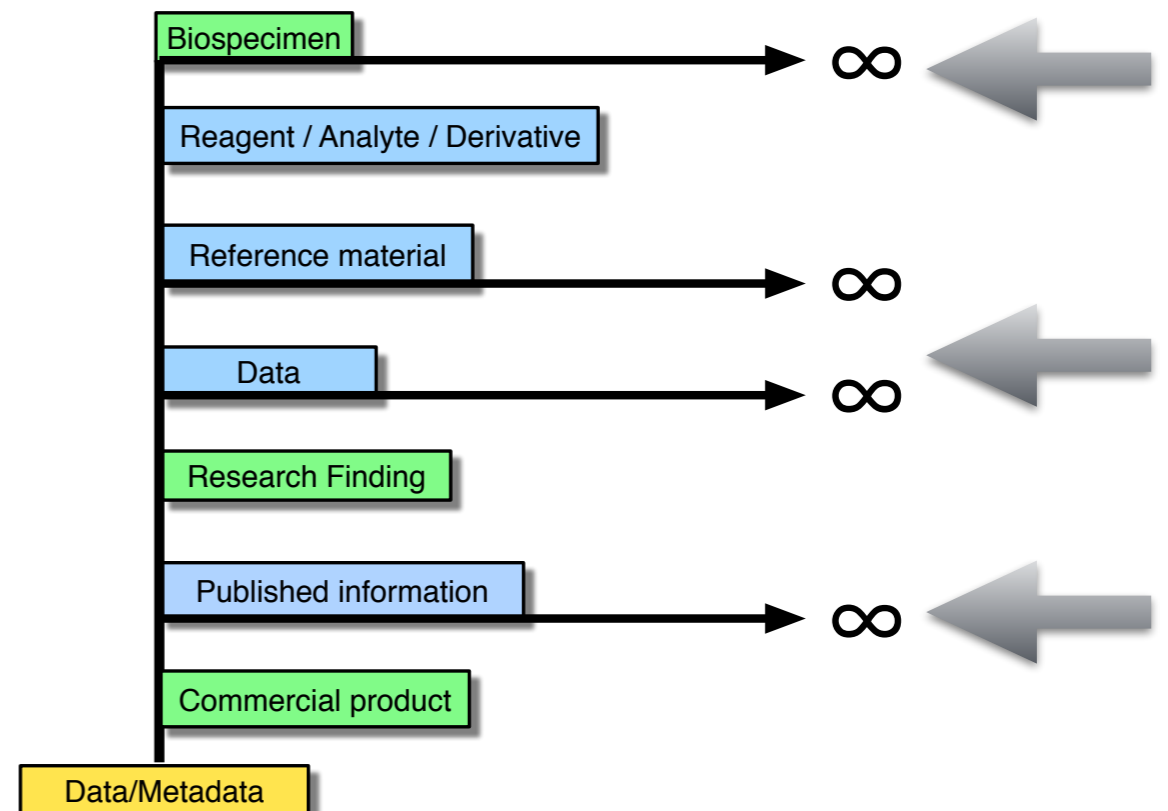
- [Link back to the specimen/person](#)
- Collection metadata
- Laboratory metadata
- Computational metadata



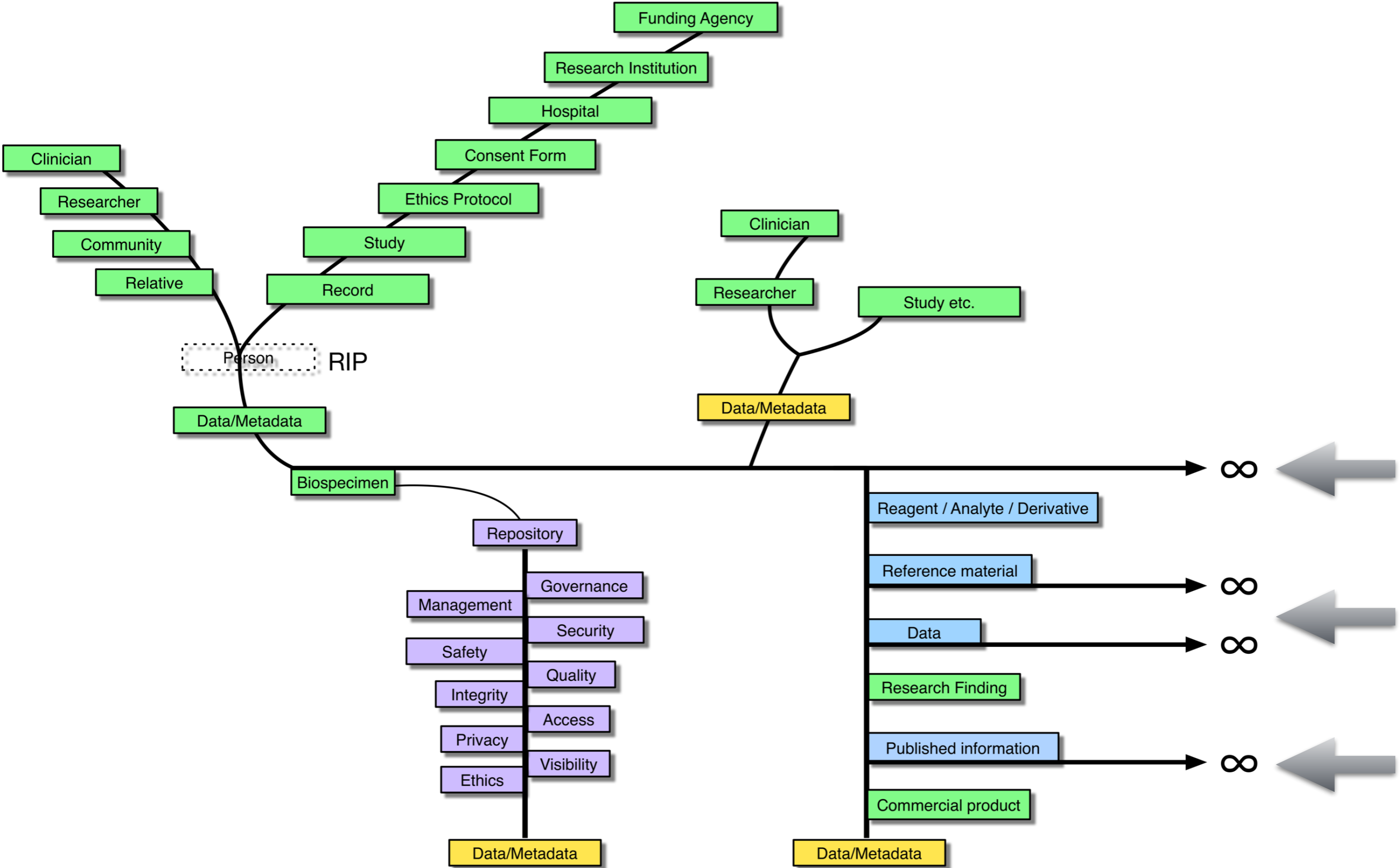
Great value from open access



Really huge value in preserving provenance data or biospecimen



Huge value from retaining the connection to source



National Centre for Indigenous Genomics



Bringing together the world's oldest living culture and its newest science.

Mission:

“To manage a repository of Indigenous biospecimens, genomic data, and documents, under Indigenous governance, for research and other uses that benefits the Indigenous donors, their communities and descendants, the broader Indigenous community, and the general Australian community.”



Bringing together the world's oldest living culture and its newest science.

National Centre for Indigenous Genomics: Rationale

We know almost nothing about the genomes of Indigenous Australians.

Without this knowledge, Aboriginal and Torres Strait Islander peoples will be excluded from the benefits that flow from human genomic research.

Rather than helping to close the health gap, these developments in medical science may actually cause it to open up even further.

... but there's a problem

For the past 15 years, it has been more difficult to engage Indigenous Australians in genetics and genomics than in any other area of health and medical research.

... for which we have a solution

A national framework under Indigenous governance that respects and protects Indigenous cultures is needed to allow Indigenous people to participate in and access the benefits of genomic research without risking social or cultural harm.

Right Way Research

Indigenous people have also long argued that health and medical research that is not conducted in line with Indigenous cultural values and in partnership with Indigenous people is unlikely to lead to health benefits and may even cause harm.

Indigenous people will only receive the benefits of genome science if research is carried out through effective engagement with Indigenous communities.

The NCIG Collection



Biological samples collected from ~7,000 Indigenous Australians.

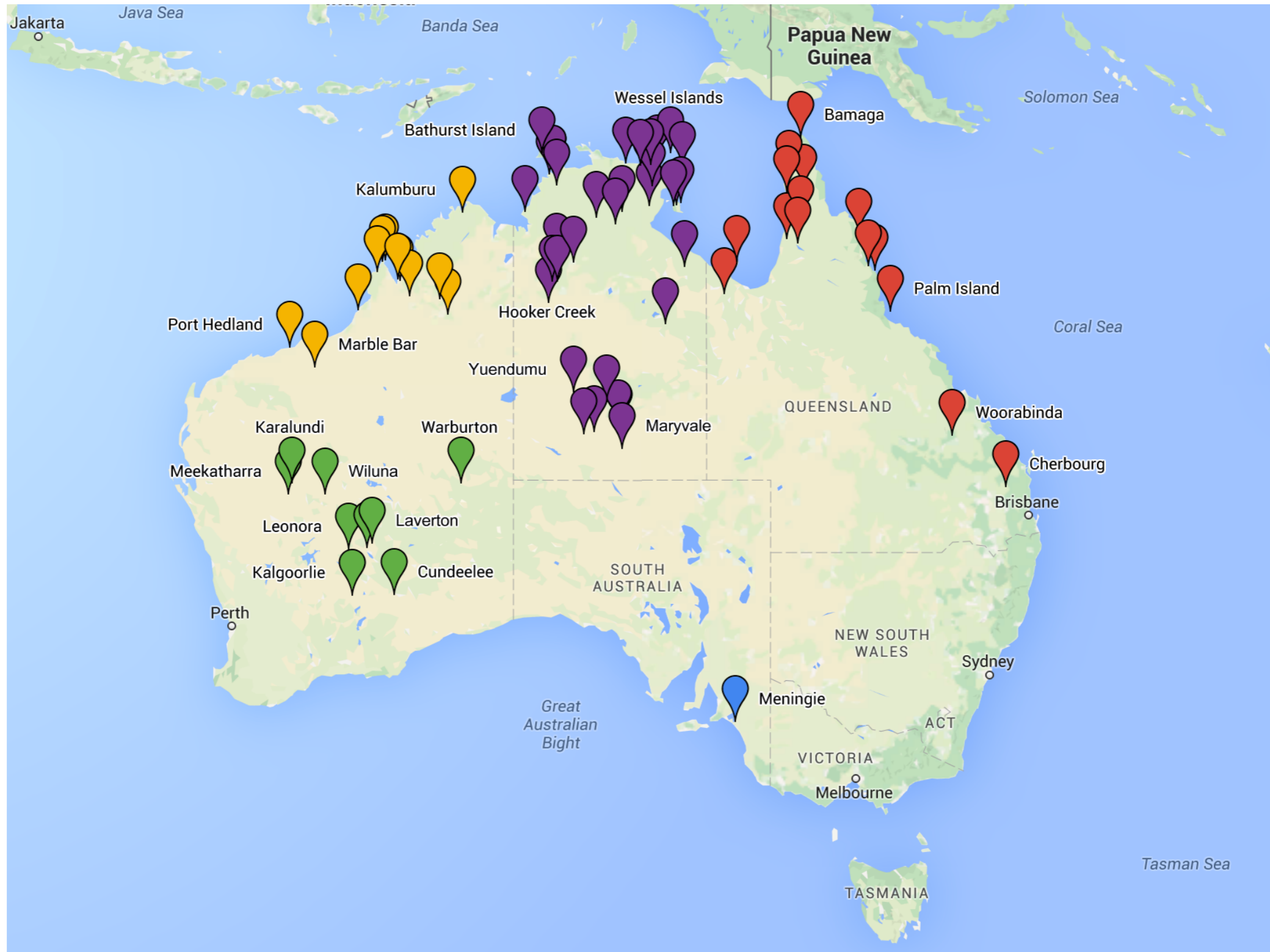
Field notes, correspondence and information about people, families and communities



NCIG was born out of a need to manage this collection.

It now also provides a more general framework, under Indigenous governance, that will enable Indigenous Australians to benefit from advances in genome science.

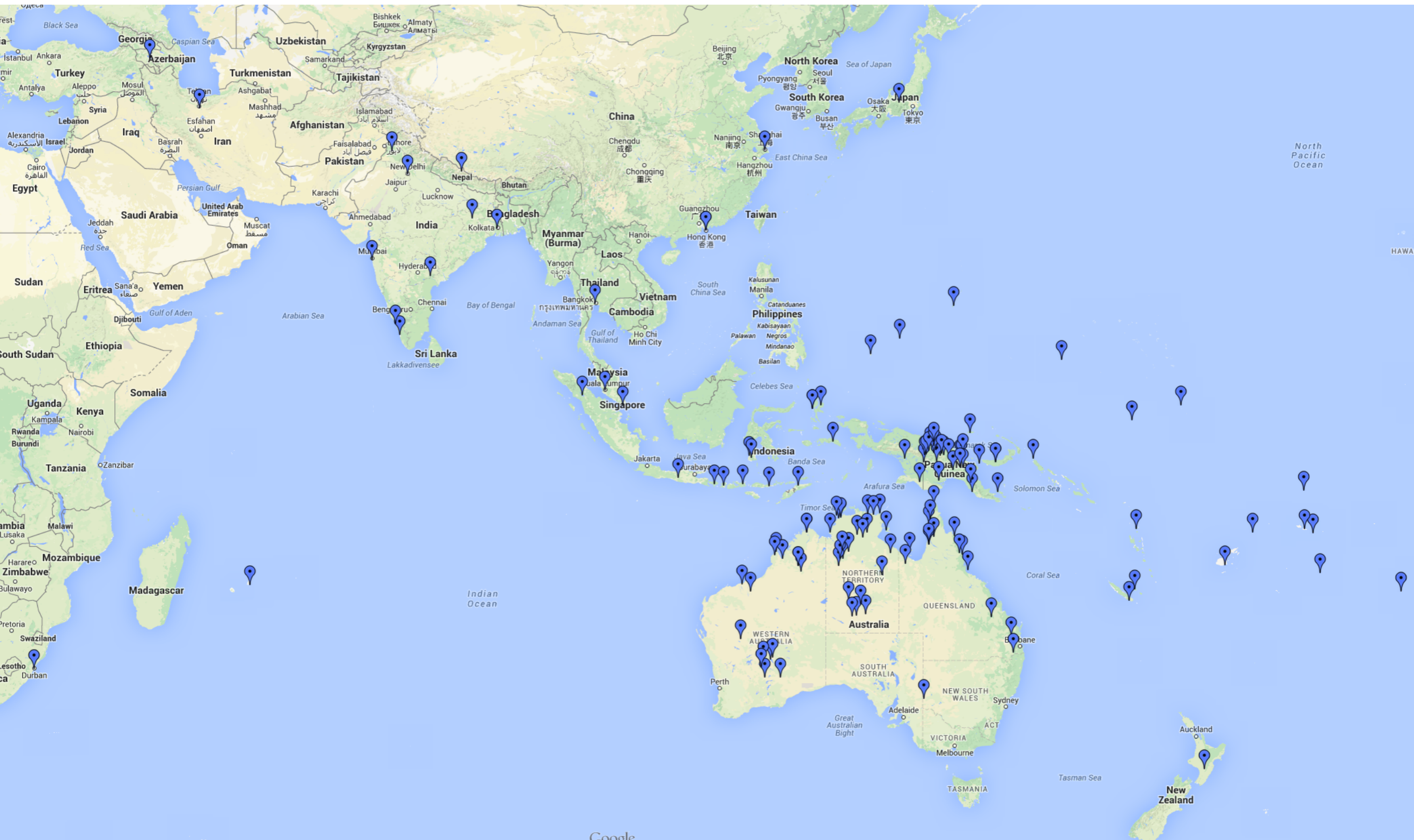
Samples collected from ~7,000 people from > 50 distinct Aboriginal communities





Camp at Pineapple Bore, Hall's Creek, 1961

These are part of a much larger collection of samples (~400,000) mainly from the Asia-Pacific region



Indigenous Consultative Committee



Prof Ian Anderson (Chair). Assistant Vice Chancellor of Indigenous Higher Education Policy and Director of the Murrup Barak Institute of Indigenous Development, University of Melbourne.

Prof Kerry Arabena, Director of Research, School for Indigenous Health, Monash University.

Mr Mick Gooda, Aboriginal and Torres Strait Island Social Justice Commissioner, Human Rights Commission.

Dr Misty Jenkins, Research Fellow, Peter MacCallum Cancer Centre.

Prof Marcia Langton, Chair of Australian Indigenous Studies, University of Melbourne.

Mr Glenn Pearson, Manager of Aboriginal Research, Telethon Institute for Child Health Research, University of Western Australia.

Assoc Prof Mark Wenitong, School of Tropical Public Health, James Cook University and Board Member, Australian Institute of Aboriginal and Torres Strait Islander Studies.

Consultative Committee on Indigenous Collection: Recommendations

- **The ANU Collection has immense cultural, historical and scientific importance**
- The ANU Collection should be preserved and developed as a ‘managed collection’ for appropriate research purposes according to proper, respectful, ethical and legal requirements
- **This will mark a ‘watershed’ moment in the history of Indigenous research and bioethics in Australia**
- The outcome of this process must aim for high-quality research with substantial beneficial outcomes for the Indigenous people of Australia
- **The outcome of this process should provide a model for the conduct of genetic research with Indigenous populations both in Australia and elsewhere in the world**
- The ANU establish a Governance Board for the collection with Indigenous leadership and Indigenous-majority membership

Consultative Committee on Indigenous Collection: Recommendations

“In a mainstream context, the fact that collection practices were appropriate for the time would usually mean that samples could be used with a waiver of consent. However, given the sensitivity of research in Indigenous Australian contexts, it is more appropriate to consult with donors or, where donors have passed away, their descendants and communities.”

This simple requirement turned an interesting, but difficult project into an opportunity to re-write the book, guided by Indigenous participants

NCIG Governance Framework 2.1.4.d

“NCIG has particular respect for the Indigenous understanding that the present and future are inseparable from the past, and for the extended kinship bonds on which the integrity of Indigenous Communities is based.”

Rewriting the book

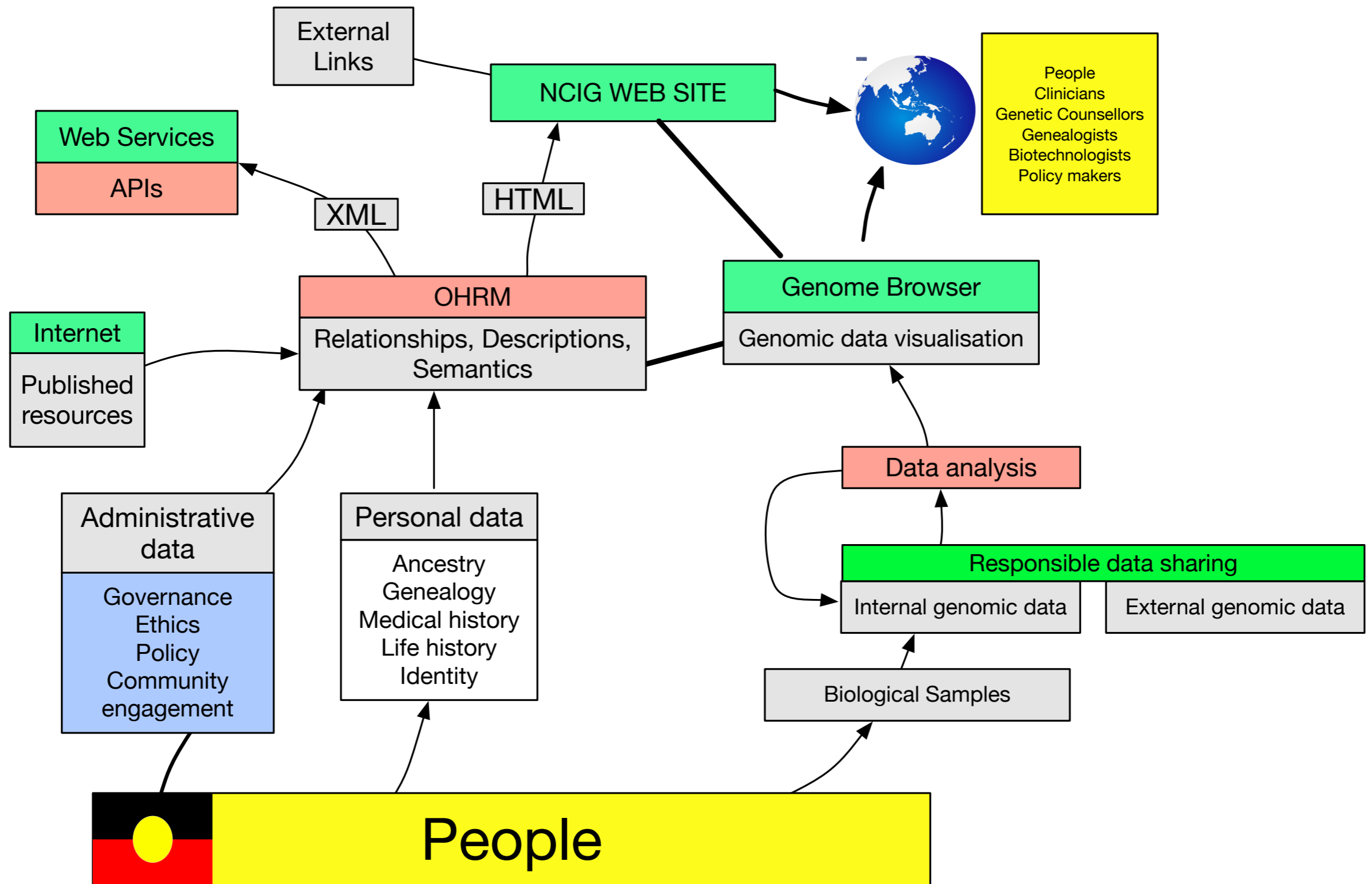
- Community engagement
- Free, prior and informed consent
- Research governance
- Data management and sharing

Indigenous governance, decision-making and control

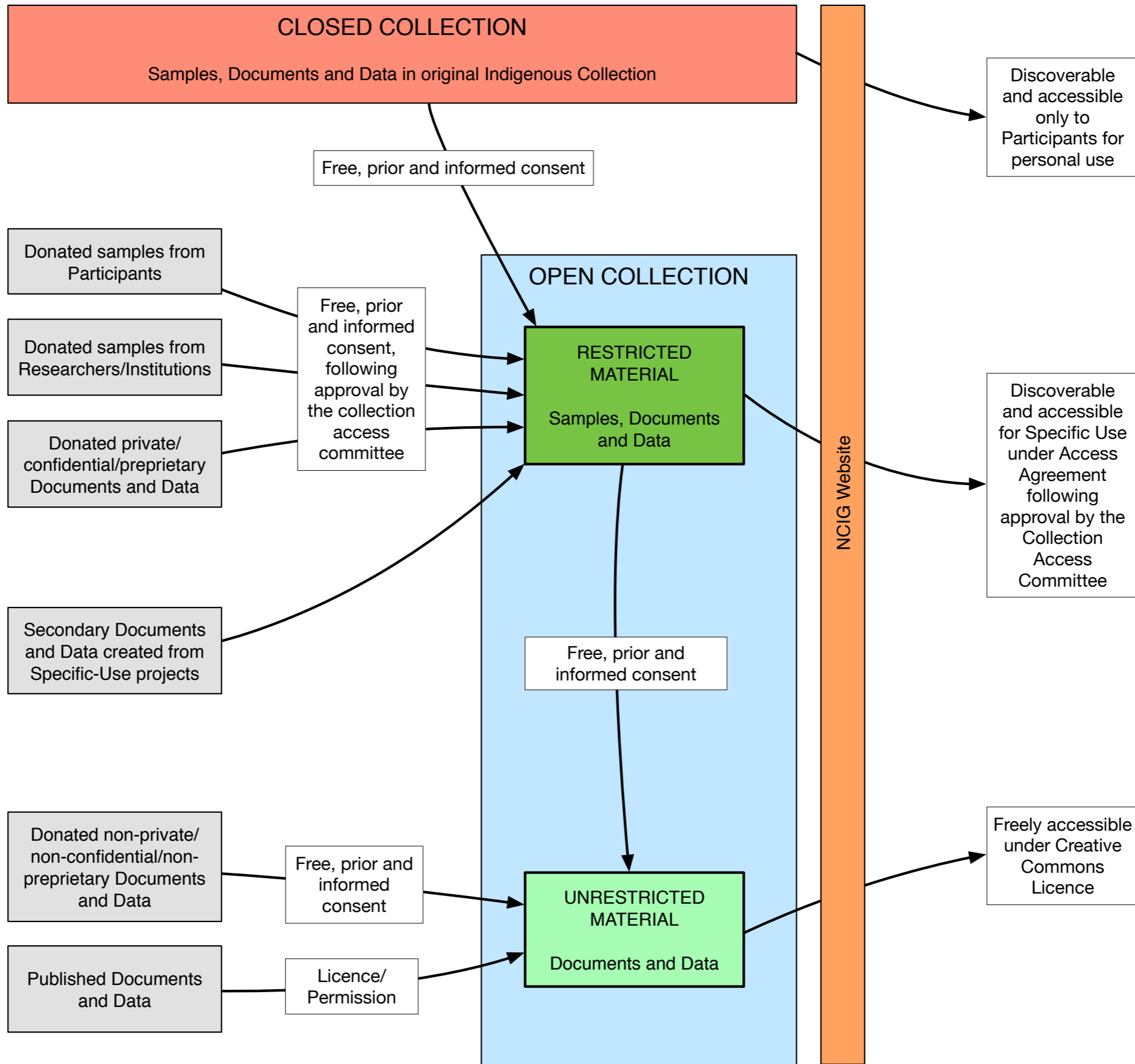


- Indigenous-led Governance Board (Chair: Mr Mick Gooda; Aboriginal and Torres Strait Islander Social Justice Commissioner).
- An independent Indigenous Access Committee controls access to the NCIG Collection.
- Research Advisory Committee (RAC) has Indigenous members.
- Governance Framework covering all aspects of NCIG's activities.
- Ethics Protocol approved by ANU HREC and Aboriginal and Torres Strait Health Ethics Committees in specific jurisdictions.
- Indigenous community representatives notified of NCIG Operations affecting their communities.

NCIG data management



Closed Collections and Restricted and Unrestricted Material in Open Collection



Reciprocal transformation through engagement

- “The promise of genomics is that it will transform how we manage our health and treat our diseases.”
- Our aim is to ensure that this promise is extended to Indigenous Australians.
- We and our research practices are already being transformed in the process
- We hope others might learn something from this engagement. Maybe Indigenous communities can show how ‘right way research’ is a good thing for everyone.



Bringing together the world's oldest living culture and its newest science.