

POLICY BRIEF

Recognizing Indigenous Interests: Labeling DSI with Provenance Metadata

Jane Anderson, Maui Hudson, Stephany
RunningHawk Johnson, KatieLee Riddle

Why Provenance Metadata Matter

Indigenous lands, resources and Traditional Knowledge (TK) are regularly the source or subject of genetic research, however, information about that source is often absent, obscured or missing from source metadata.^{1,2} This is due to research practices that have not historically aligned with Indigenous expectations about appropriate acknowledgement or attribution. The information systems that have been built to store and share genetic material also lack the critical capacity to reflect Indigenous provenance or cultural authority.³ This is a significant issue, as Digital Sequence Information (DSI) can now be generated with ease, and the general expectation is that it will be deposited in open data repositories.

As these historic information gaps and their ethical implications have become clearer, there is simultaneously widespread agreement across the research science community that metadata - or information about data including the how, where and who of data collection - is as important as the data itself.⁴ Metadata is what makes data 'findable and reusable', two key elements of the FAIR Principles.⁵ For example, geographical origin information helps contextualise genetic data, improves overall scientific utility and promotes responsible data sharing.⁶ Appropriate Indigenous metadata can also connect DSI with Indigenous Peoples and Local Communities (IPLC).⁷

Recognizing Indigenous **provenance** in metadata and establishing appropriate attribution protocols has emerged as key mechanisms for alignment with the FAIR & CARE Principles as well as protecting the rights and interests of IPLCs in relation to their genetic resources.⁸ Metadata will be around far longer than the systems and institutions that have generated it, and ensuring its interoperability

with existing standards is the only way to safeguard its survival and scientific utility into the future. Geographical origin data and Indigenous provenance data, will provide opportunities to recognize IPLC interests and support fair and equitable benefit sharing. Equitable benefit sharing is not only about the distribution of monetary funds, but the opportunity to participate in future research activities.

International Developments

CBD: COP15 Decision 15/9 established a multilateral mechanism to share benefits derived from DSI. While the decision notes that tracking and tracing of all DSI is not feasible, it does acknowledge the importance of IPLC considerations within DSI, and addresses the FAIR & CARE Principles in the sharing of DSI. COP16 will continue discussions about the nature of the multilateral mechanism, including the usefulness of geographical information and IPLC provenance information. Discussions within the CBD have identified one international mechanism - the Local Contexts TK and BC Labels and Notices - as a key tool for IPLC provenance metadata.

WIPO: In May 2024, a WIPO Diplomatic Conference ratified the new WIPO Treaty on Intellectual Property, Genetic Resources, and Associated Traditional Knowledge (aTK). This Treaty confirms the need for a disclosure requirement of country of origin for the use of GR, and where there is aTK, disclosure of the relevant IPLC who provided the TK is also necessary. While DSI has not been specifically mentioned in the WIPO Treaty due to the difference of opinion between the Parties, the need for greater transparency and disclosure of Indigenous interests in research and research outcomes is clearly a cross-forum issue.

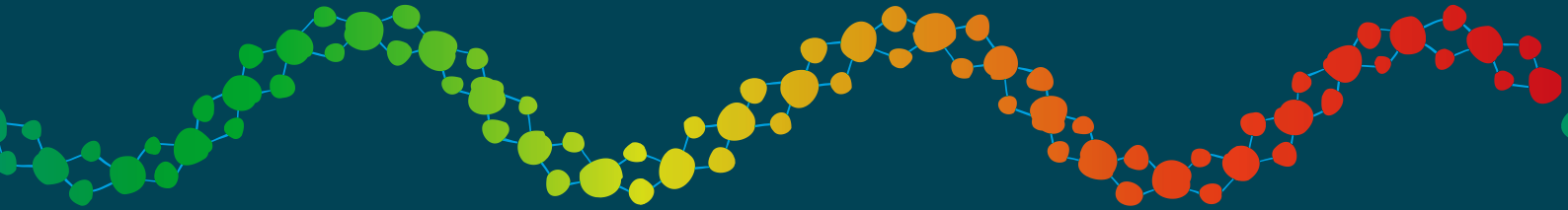


Figure 1: Local Contexts (www.LocalContexts.org)

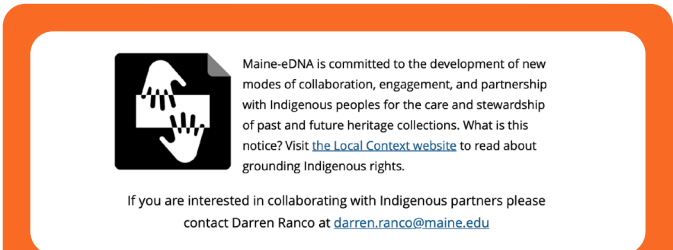


Figure 2: Maine-eDNA - Open to Collaborate Notice

The University of Maine's environmental DNA (eDNA) research programme uses the 'Open to Collaborate' Notice from Local Contexts. This Notice demonstrates the programme's commitment to new modes of collaboration, engagement and partnerships with Indigenous Peoples for the care and stewardship of past and future heritage collections.

Tribal Historic Preservation Officers (THPOs) in Maine receive notifications about eDNA samples collected from their tribal lands and can apply specific Labels reflecting contextual provenance, permissions and protocol information. This brings the THPOs into the management of the data and enables them to more effectively monitor research activities and future data applications. Recognizing data provenance ensures adherence to tribal protocols and permissions. (<https://umaine.edu/edna/open-to-collaborate/>)

Local Contexts

Local Contexts is a global non-profit that supports Indigenous communities with tools to reassert sovereignty and cultural authority in collections and data. It offers tools for institutions and researchers to disclose the existence of Indigenous rights and interests across a variety of research and non-research contexts.

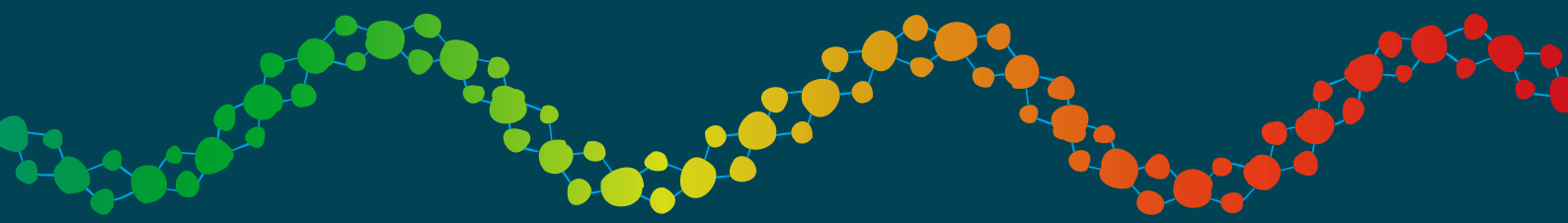
The Traditional Knowledge (TK) and Biocultural (BC) Labels and Notices work to enhance and legitimize locally based decision-making and Indigenous governance frameworks for determining access, and culturally appropriate conditions for sharing collections of cultural heritage and Indigenous data.

As extra-legal instruments, the Labels and Notices make it possible for researchers to disclose Indigenous interests, and for IPLC's to affirm the nature of their relationship to the data as well as protocols and permissions for re-use. Importantly, both Labels and Notices can be used in the context of genetic resources, DSI and associated traditional knowledge. Notices and Labels create pathways for partnership and collaboration with IPLCs.

Specifically, the Notices work to activate researcher and institutional responsibility to identify potential Indigenous rights and interests. The Labels establish Indigenous cultural authority and clarify Indigenous and local community rights, interests, and relationships to collections and/or data.

Traditional Knowledge and Biocultural Labels

Local Contexts Labels are becoming essential tools for recognizing Indigenous provenance, protocols and permissions in records for both natural ecosystems and cultural heritage.⁹ They make visible and emphasize the importance of IPLC knowledge systems in research and research outputs (c.f data, datasets and publications). The Traditional Knowledge (TK) and Biocultural (BC) Labels communicate user responsibilities to Indigenous data



including access and expectations for future circulation. Labels are designed to be customized by the community and reflect community authority.

When researchers are collaborating with communities and submitting sequence information to databases, they can properly attribute and connect provenance to their data using the Labels. They have a permanent identifier (PID) that can be linked through digital ecosystems including data aggregators.

The Labels and Notices are being integrated into digital research infrastructures including specific projects to develop interoperability with ORCID, DataCite, the international publishing community, Cyverse, Dataverse, GEOME and GBIF.



TK and BC Labels are for Indigenous communities

What do Labels do?

- Puts data in context
- Makes visible provenance and the ethics of collection
- Foreground relationships that makes research possible
- Function as metadata that connects data to people, to environments and to relationships overtime



Notices are for researchers, institutions and data repositories

What do Notices do?

- Disclose Indigenous interests in collections
- Indicates openness to collaboration
- Demonstrates readiness for Indigenous metadata
- Supports transparency in research relationships

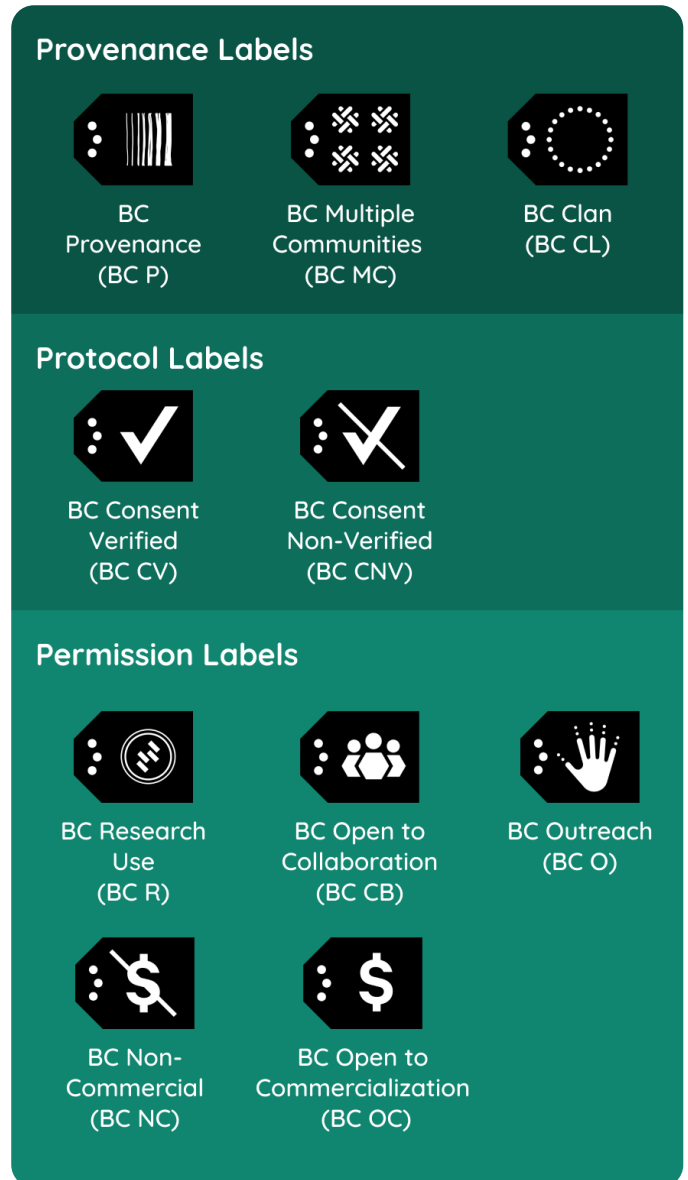
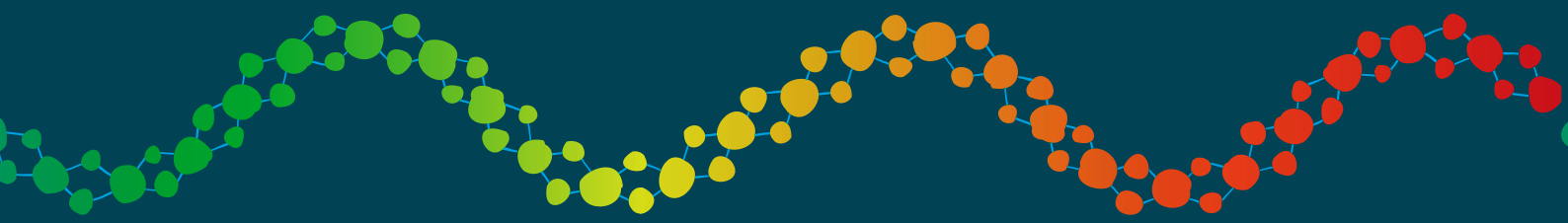


Figure 3: Biocultural Labels




Permissions


Project permits


Project title:
Biological specimens housed at Manaaki Whenua: Te rohe o Whakatohea

Reference:
Local Contexts - Whakatohea

 BC Provenance (BC P) | Nā wai/ Nō hea

 BC Research Use (BC R) | Rangahau

 BC Open to Collaboration (BC CB) | Kotahitanga

 BC Open to Commercialization (BC OC) | Umanga

Project title:
Local Contexts - Allan Herbarium (CHR)

Reference:
CHR Collection - Local Contexts

 Biocultural (BC) Notice

Figure 4: Manaaki Whenua Landcare Research's Systematics Collection Database (SCD)

The Manaaki Whenua Landcare Research's Systematics Collection Database (SCD) has applied Biocultural Notices & Labels to almost 700,000 records in their system. Connecting through the Local Contexts Hub API allows them to update and add Labels to newly digitised records and is a good example of how small changes to digital infrastructures in research organizations can have major effects and improve provenance, transparency and the recognition of cultural authority. The database provides open access to specimen and culture data from a variety of national collections and the Notices disclose the potential for Indigenous interests in the specimens and data from biological samples. Collections can be filtered using Native Land Digital maps so Māori communities can easily find records associated with their territories. Records from the SCD are also shared with the Global Biodiversity Information Facility (GBIF). A work plan is being established to ensure Notices and Labels applied on local databases will also be visible on GBIF. <https://scd.landcareresearch.co.nz/>

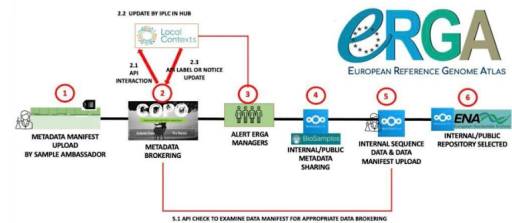


Figure 5: ERGA Label and Notice Implementation Workflow

Guidelines on the Implementation of the Traditional Knowledge and Biocultural Labels and Notices in the European Reference Genome Atlas for Biodiversity Researchers

The European Reference Genome Atlas (ERGA) is streamlining the collection and storage of ethical and legally compliant metadata for all genomic data across Europe. Recognizing that research portals are collections of repositories with different metadata brokers, ERGA has partnered with Local Contexts to implement the Traditional Knowledge and Biocultural Notices and Labels to translate across data ecosystems.

This has included creating requirements for disclosing Indigenous rights and interests in biodiversity data as users upload metadata. Additionally, ERGA manifest files (computing files that explain metadata versions, licences and constituents) include data fields with options for regulatory compliance, recognition of Indigenous rights, associated Traditional Knowledge contacts, and the identification of whether ethics, sampling and/or Nagoya permits are required and obtained.

Supporting Indigenous Data Governance

The Labels and Notices provide a practical application of the CARE Principles for Indigenous Data Governance.¹⁰ The CARE Principles¹¹ which are promoted internationally in policy documents including the [UNESCO Science Outlook](#), [World Data Systems Data Sharing Principles](#), & [Earth Biogenome Project Data Sharing and Best Management Practices](#).

Provenance Connection Control Governance	An expression of Indigenous Data Governance by explicitly recognizing the Indigenous rights and interests in Metadata <ul style="list-style-type: none"> Engages researchers by providing clear details about where data comes from and who to engage with for future use Engages institutions to establish appropriate relationships with communities over control and decision-making
Protocols Worldviews Education	An expression of Indigenous Data Governance by enabling Indigenous worldviews within digital infrastructures <ul style="list-style-type: none"> Educates user publics and limits inappropriate/derogatory use Engages institutions and researchers to apply proper care and Indigenous rules for knowledge sharing
Permissions Responsibility Transparency Integrity	An expression of Indigenous Data Governance by supporting ethical and equitable biodiversity research <ul style="list-style-type: none"> Make transparent Indigenous expectations and intentions of use and engagement Engages institutions and repositories to develop use policies and agreements

Table 1: Supporting Indigenous Data Governance

Shifts in metadata practices for Indigenous rights and interests recognize the contribution that TK brings to science broadly and genetic research specifically. Strategies that expand the metadata categories offer valuable solutions for non-Indigenous and Indigenous partners by adding more information to data.

The Labels and Notices also have a role in facilitating future repatriation/rematriation efforts of samples, aTK, and sequencing information in accordance with the CBD-adopted "[Rutzolijirisaxik Voluntary Guidelines](#)" on ethical repatriation/rematriation of IPLC samples.

The Smithsonian Institution holds biological materials collected during an expedition to Moorea, Tahiti in 2009. Over 50,000 samples were taken during this expedition and the sample metadata is also maintained on GEOME - a metadata database which has enabled the use of Local Contexts Notices and Labels. The project team recently applied BC Notices to all 50,000 samples collected during that expedition. Using the Community Notification function in the Local Contexts Hub, the Attita Center was notified of the existence of the samples and can add their community Labels giving clarification about consent and future use. Interoperability is an important factor and the Labels are also visible on iSamples and can be harvested into GBIF as soon as GBIF has that capability.

Links

Geome

- <https://geome-db.org/workbench/project-overview?projectId=75>

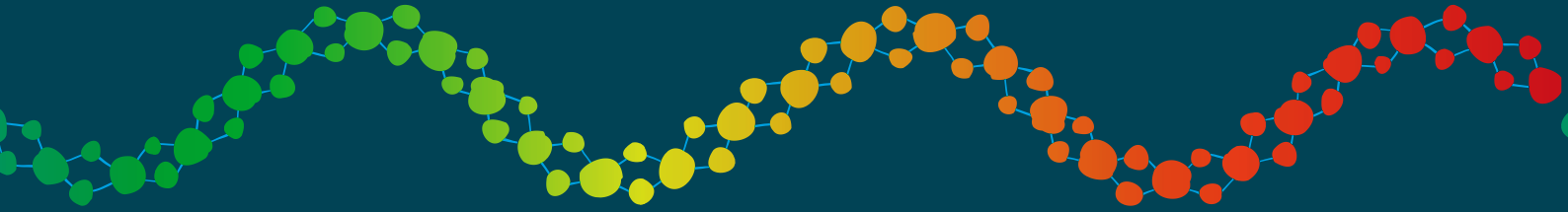
- https://geome-db.org/record/ark:/21547/CVJ2BMOO_00010

iSamples

- https://central.isample.xyz/isamples_central/thingpage/ark:/21547/CVJ2BMOO_00010

Local Contexts ID

- <https://localcontextshub.org/projects/71b32571-0176-4627-8e01-4d78818432a7/>



Supporting Indigenous Communities

Indigenous communities utilize the Local Contexts TK and BC Labels to define attribution, access, and use for their intellectual and cultural property, cultural heritage, environmental data, and genetic resources. To find out more about Indigenous community use of TK & BC Labels see Local Contexts Films. <https://localcontexts.org/films/>



Figure 6: E kore au e ngaro - The Connection Remains

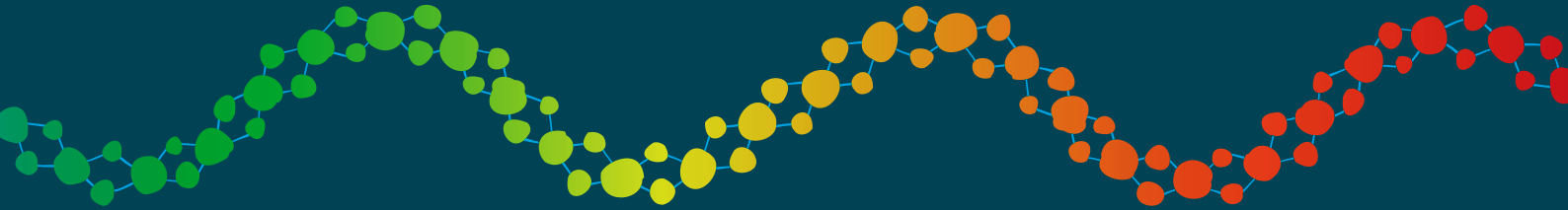


Figure 7: Awasəwehlawəłətina wikəwamok – They Returned Home

Sarayaku Community and the Living Forest - In 2018, the Sarayaku community in the Ecuadorian Amazon released the Kawsak Sacha (Living Forest) Declaration. The community is now collaborating with the NYU MOTH Project, the Fungi Foundation, and the Society for the Protection of Underground Networks (SPUN) to identify unique underground more-than-human beings living in their territories, as part of the implementation of the Kawsak Sacha initiative. The Sarayaku, along with the Living Forest, steward these underground networks of life. The Fungi Foundation's Ethnomycology Ethical Guidelines provide a framework for decision-making and conduct for ethnomycological research including projects with Indigenous Peoples, traditional societies, and local communities. The Sarayaku, MOTH, Fungi Foundation, and SPUN now have a new research collaboration. All parties have decided to utilise Local Contexts Labels to ensure appropriate acknowledgement and attribution to both the community and the Living Forest as jointly holding the cultural authority over any data, films, and sounds generated by the project.

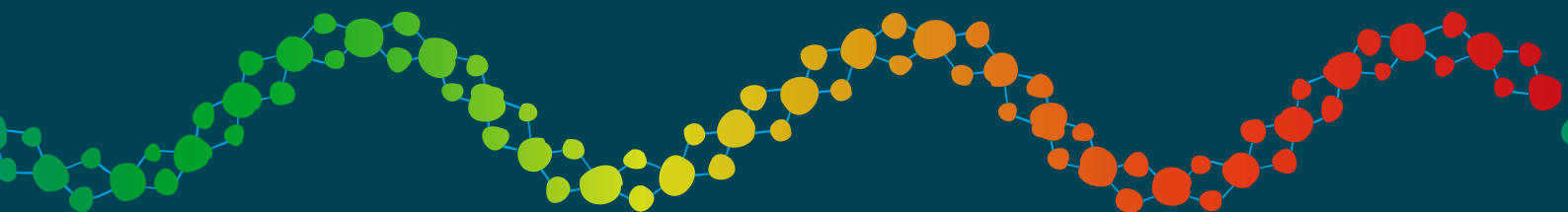
References:

- <https://amazonwatch.org/assets/files/2016-kawsak-sacha-proposal-english.pdf>
- <https://sumauma.com/en/jose-gualinga-montalvo-a-floresta-e-um-ser-vivo-inteligente-e-consciente/>
- <https://www.ffungi.org/campaign/ethnomycology-ethical-guidelines>

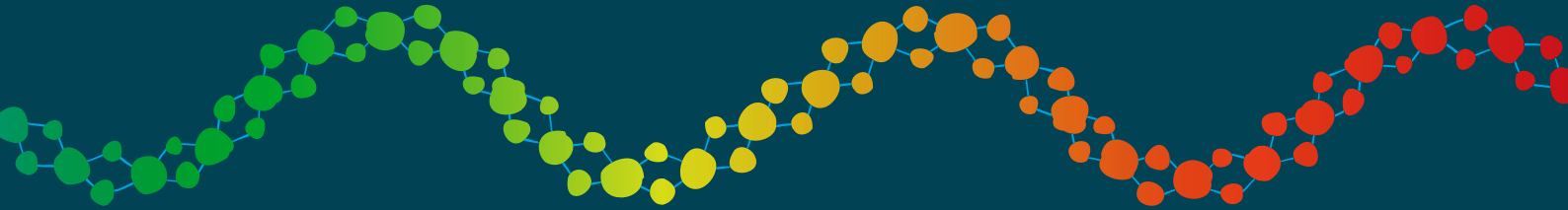


Implications for CBD Negotiations on Digital Sequence Information

Questions	Answers
Is this a track and trace system?	No, the Labels are designed to enhance cultural metadata associated with samples and DSI related datasets derived from Indigenous Peoples and Local Communities lands and territories.
Do Labels create legal obligations for researchers and/or companies?	No, the Labels are an extra-legal mechanism. Legal obligations are created by the CBD and subsequent national legislation.
Do Labels create ethical obligations for researchers and/or companies?	Yes, the Labels are expressing the ethical expectations of the relevant Indigenous Peoples and Local Communities for future use of genetic resources and DSI. Ethical data practices are pushing towards better provenance information and attribution of Indigenous and Traditional communities.
Why are Labels being connected to repositories and publishing processes?	Indigenous communities have expressed a need to understand where genetic resources are and how they are being used. Research generates datasets and produces research papers. The Labels can help communities identify what research has been done and who is utilizing genetic resources and DSI. It enables communities to have their name associated with those data resources and publications.
How does it support data governance?	You can't govern data if you don't know where it is. The CARE Principles are being adopted alongside the FAIR Principles and the Labels provide a practical and standardized mechanism for recognizing provenance, and sharing protocols and permissions.
How might the Labels support researchers?	Labels provide researchers with additional information which can better inform their research and/or commercialisation activities. It provides a pathway for gaining FPIC if that is required for the planned activities.
Do I have to engage with Indigenous communities when using DSI on open databases?	No, if Labels are present on a dataset it is up to the researcher/company to decide if they need to engage. The information can support the process and may indicate that the community is open to those types of discussions. Engagement with communities may be beneficial for accessing certain types of research funds and/or engaging in certain commercial processes.
Why are Indigenous communities utilising BC Labels?	The BC Labels allow Indigenous Peoples and Local Communities to be included on records about genetic resources and associated DSI. This creates opportunities for future collaborations and partnerships.



Questions	Answers
How could it be utilised within INSDC?	BC Notices and Labels have already been integrated into genomic data repositories (e.g. GDV), metadata databases that connect to the INSDC (e.g. GEOME), National Collections that connect to GBIF (e.g. Manaaki Whenua Landcare in New Zealand, and Smithsonian Institution in US). INSDC just need to agree to include a new field in their metadata schema for Indigenous provenance. It would only be used if provenance information was known and available.
How could you deal with Labels being applied on parts of an aggregated DSI dataset	Many aggregated datasets could include Labels from a single or multiple communities. It will be up to the researcher to decide how they deal with this situation. It may be appropriate to recognise the Labels in subsequent research outputs but we expect the research community to develop standardised practices for appropriate attribution for these cases (e.g. Datacite, ORCID, Dataverse, Crossref etc). The MLM provides an appropriate pathway for companies utilising aggregated data from multiple sources.
Are Labels necessary in the context of a Multi-Lateral Mechanism?	Yes, the Labels provide useful provenance information. This might be a balance factor in the distribution of funds from the MLM. However, the primary use of the Labels is to support opportunities for future bilateral research and/or commercial collaborations with Indigenous Peoples and Local Communities.
What would happen if the use of Labels wasn't supported for DSI?	The MLM aims to increase the funding available for benefit sharing to IPLC's from commercialisation activities involving DSI (and potentially genetic resources). Compared to the existing Bilateral mechanism in the Nagoya Protocol, the MLM creates a beneficiary relationship with Indigenous Peoples and Local Communities. Labels create pathways towards more direct relationships with Indigenous Peoples and Local Communities.
How do Labels relate to the new WIPO Treaty on Patent Disclosures?	The new WIPO Treaty requires disclosures when traditional knowledge and genetic resources from IPLC's are utilised, if known, when patent applications are being made. Notices and Labels may provide a critical signal and information source for Patent Offices checking the veracity of claims in patent applications.



Biocultural Labels: Selected examples - Why use this Label?



BC Consent Verified (BC CV)

Why use this Label?

Indigenous peoples have the right to permission the use of information, biological collections, data, and digital sequence information (DSI) that derives from associated lands, waters and territories. This Label verifies that there are consent conditions in place for uses of information, collections, data, and digital sequence information.



BC Outreach (BC O)

Why use this Label?

This Label should be used when you would only like your biocultural materials and/or data used for educational outreach activities. Outreach activities means to share works outside the community in order to increase and raise awareness and education about your family, clan and/or community. Sites for outreach activities can include schools, universities, libraries, archives, museums, online forums, and small learning groups.



BC Multiple Communities (BC MC)

Why use this Label?

This Label should be used to indicate that multiple communities have responsibility, custodianship and/or ownership over the geographic regions where this species or biological entity originates/is found. This Label recognizes that whilst one community might exert specific authority, other communities also have rights and responsibilities for use and care.



BC Provenance (BC P)

Why use this Label?

Indigenous peoples have the right to make decisions about the future use of information, biological collections, data, and digital sequence information (DSI) that derives from associated lands, waters and territories. This Label reflects a significant relationship and responsibility to the species or biological entity and associated scientific collections and data, and supports the practice of proper and appropriate acknowledgement into the future.



BC Open to Commercialization (BC OC)

Why use this Label?

Indigenous peoples have the right to benefit from information, biological collections, data, and digital sequence information (DSI) that derives from traditional lands, waters and territories. This Label is being used to indicate an express interest in being a primary party to any future negotiations if future commercialization opportunities arise.



BC Research Use (BC R)

Why use this Label?

This Label should be used to permission use of information, collections, data, and digital sequence information (DSI) for unspecified research. This Label does not give permission for commercialization activities.

References

- 1 Anderson, J. & Christen, K. (2019). Decolonizing Attribution: Traditions of Exclusion. *Journal of Radical Librarianship*. Vol 5.
- 2 Golan J, Riddle K, Hudson M, Anderson J, Kusabs N and Coltman T (2022) Benefit sharing: Why inclusive provenance metadata matter *Front. Genet.* 13:1014044. doi: 10.3389/fgene.2022.1014044
- 3 Ambler, J., Diallo, A. A., Dearden, P. K., Wilcox, P., Hudson, M., and Tiffin, N. (2021). Including digital sequence data in the Nagoya protocol can promote data sharing. *Trends Biotechnol.* 39 (2), 116–125. doi:10.1016/j.tibtech.2020.06.009
- 4 Scholz, A. H., Freitag, J., Lyal, C. H. C., ... Overmann, J. (2022). Multilateral benefit-sharing from digital sequence information will support both science and biodiversity conservation. *Nature Communications*, 13(1). <https://doi.org/10.1038/s41467-022-28594-0>
- 5 Leigh, D.M., Vandergast, A.G., Hunter, M.E. et al. Best practices for genetic and genomic data archiving. *Nat Ecol Evol* (2024). <https://doi.org/10.1038/s41559-024-02423-7>
- 6 Raposo, D.S., Cepeda, M.L., Ebert, B., & Scholz, A.H. (2023). Challenges and opportunities of geographical origin information in DSI benefit-sharing: a global analysis from the academic sector. https://www.dsiscientificnetwork.org/wp-content/uploads/2023/10/PolicyBrief_Geographical_Origin.pdf
- 7 Liggins, L., Hudson, M., and Anderson, J. (2021). Creating space for Indigenous perspectives on access and benefit-sharing: encouraging researcher use of the local contexts notices. *Mol. Ecol.* 30 (11), 2477–2482. doi:10.1111/mec.15918
- 8 Mc Cartney, A.M., Head, M.A., Tsosie, K.S. et al. Indigenous peoples and local communities as partners in the sequencing of global eukaryotic biodiversity. *npj biodiversity* 2, 8 (2023). <https://doi.org/10.1038/s44185-023-00013-7>
- 9 Anderson, J., and Hudson, M. (2020). The Biocultural Labels Initiative: Supporting Indigenous rights in data derived from genetic resources. *Biodivers. Inf. Sci. Stand.* 4, e59230. doi:10.3897/biss.4.59230
- 10 Jennings L, Anderson T, Martinez A, Sterling R, Chavez DD, Garba I, Hudson M, Garrison NA, Carroll SR. (2023) Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. *Nat Ecol Evol.* 2023 Oct;7(10):1547–1551. doi: 10.1038/s41559-023-02161-2. PMID: 37558804
- 11 <https://www.gida-global.org/care>

Biographies

Jane Anderson

Jane Anderson is a Professor at New York University in Lenapehoking (NYC). They are the Co-Founder of Local Contexts and Co-Director of ENRICH, Equity for Indigenous Research and Innovation Coordinating Hub.

Maui Hudson (Whakatōhea)

Maui is an Associate Professor and Director of Te Kotahi Research Institute at the University of Waikato. He is a Co-Director of ENRICH, Equity for Indigenous Research and Innovation Coordinating Hub, and a Council member and Strategic Advisor for Local Contexts.

Dr. Stephany RunningHawk Johnson

Dr. Stephany is a Citizen of the Oglala Lakota Nation and is the Founding Executive Director of Local Contexts. She was formerly at Washington State University, where she held a position as Assistant Professor in Cultural Studies and Social Thought in Education.

KatieLee Riddle (Rongowhakaata)

Katie is an ENRICH Scholar, Ngā Pae o te Māramatanga and SING Alumni, and enrolled Solicitor and Barrister of the High Court of New Zealand, specialises in Māori and Indigenous Intellectual Property, Digital Sequence Information (DSI), Genetic Resources, and Indigenous Data Sovereignty. She advises on these issues for the United Nations Convention on Biological Diversity and is a representative for New Zealand and the Pacific at the International Indigenous Forum on Biodiversity. She completed her Bachelor of Laws with Honours at the University of Waikato in 2020 and is a founding member of the Interdisciplinary Researchers working in DSI (iDSI) group.

DOI: 10.5281/zenodo.13154842

Preferred Citation:

Anderson J., Hudson, M., RunningHawk, S., & Riddle., K.L. Recognizing Indigenous Interests: Labeling DSI with Provenance Metadata. *Policy Brief. ENRICH* 2024.

Designed by

Indigenous Design Innovation Aotearoa (IDIA)