

# Establishing the means for providing suitable representation of indigenous knowledge within The Dataverse Project

Steve McEachern, Australian Data Archive

Sonia Barbosa, The Dataverse Project, IQSS, Harvard University

Jane Anderson, Engelberg Center for Innovation Law and Policy in the  
Law School at New York University

# Acknowledgement of Country

I acknowledge the Traditional Owners of the land on which this talk is taking place, the land of the Ngunnawal and Ngambri, and pay my respect to their Elders past, present and emerging. I also acknowledge any Aboriginal people joining us today in person and online.

# Generalist repositories and indigenous knowledge

- Many archives have large holdings of content about (and from) Indigenous communities and people
- Increasing recognition of the need to understand, include and represent indigenous forms of knowledge and rights in repositories and archives
- This presentation is an introduction to work currently occurring in the Dataverse community to support this
- Key aim is to recognise and enable indigenous data sovereignty: **“The right of Indigenous Peoples and nations to govern the collection, ownership, and application of their own data.”**

# WHAT ARE **INDIGENOUS DATA**?

Data, information and knowledges, in any format, that impacts Indigenous Peoples, nations, and communities at the collective and individual levels:

## DATA ABOUT OUR NON-HUMAN RELATIONS

Land, water,  
geology, titles, air,  
soil, sacred  
ecosystems,  
territories, plants,  
animals, etc.

## DATA ABOUT US AS INDIVIDUALS

Administrative, legal,  
health, social,  
commercial,  
corporate, services,  
etc.

## DATA ABOUT US AS COLLECTIVES

Traditional and  
cultural information,  
languages  
knowledge systems,  
ancestral and clan  
knowledges, etc.

**USINDIGENOUSDATA.ORG**  
**@USIDSN**

Informed by British Columbia First Nations Data Governance Institute - BCFNDGI.COM

**GIDA-GLOBAL.ORG**  
**@GidaGlobal**

# INDIGENOUS DATA SOVEREIGNTY

The *right* of Indigenous Peoples and nations to govern the collection, ownership, and application of their own data.

1

Derives from inherent rights to govern peoples, lands, and resources.

2

Genesis in traditions, roles, and responsibilities for the use of community held information.

3

Positioned within a human rights framework and court cases, treaties, and/or recognition.

4

Knowledge belongs to the collective and is fundamental to who we are as peoples.

See Kukutai T & Taylor J. (Eds). (2016). Indigenous Data Sovereignty. Canberra: Australian National University Press.

[USINDIGENOUSDATA.ORG](http://USINDIGENOUSDATA.ORG) | [@USIDSN](https://twitter.com/USIDSN)

[GIDA-Global.ORG](http://GIDA-Global.ORG) | [@GIDAGlobal](https://twitter.com/GIDAGlobal)

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



**Collective Benefit** Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.

**Authority to Control** Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data.

**Responsibility** Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.

**Ethics** Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

Research Data Alliance International Indigenous Data Sovereignty Interest Group. (September 2019). "CARE Principles for Indigenous Data Governance."  
The Global Indigenous Data Alliance. GIDA-global.org

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<https://www.gida-global.org/care>

# About Local Contexts

- Local Contexts is a global initiative that supports Indigenous communities with tools that attribute cultural authority of heritage and data. By focusing on Indigenous Cultural and Intellectual Property and Indigenous Data Sovereignty, Local Contexts helps Indigenous communities repatriate knowledge and gain control over how their data is collected, managed, displayed, accessed, and used in the future.

# TK/BC Labels

## Indigenous communities

- reinforce rights by applying TK and BC Labels
- Traditional Knowledge (TK) and Biocultural (BC) Labels establish Indigenous cultural authority and governance over Indigenous data and collections by adding provenance information and contextual metadata (including community names), protocols, and permissions for access, use, and circulation.
- The Label descriptions can be customized by Indigenous communities to ensure that their local and traditional protocols are embedded in those Labels.
- When Indigenous communities apply those Labels to relevant Projects, the Labels will be used as an identifier to define how their cultural material is accessed and used

## Institutions and researchers

- disclose interests by placing Notices
- Notices are tools for institutions, repositories, and researchers to support ethical use and reinforce relationships with Indigenous communities. They also work to educate the public around Indigenous rights and interests in historical and future collections and data.
- Institutions and researchers can create Projects for cultural materials they protect and apply Notices to them.
- By applying Notices it will notify Indigenous communities of Indigenous rights and interests in these collections and data.



# Presenting and representing indigenous knowledge

## Systematics Collection

CHR 559987 A – *Gentianella lutea*



Data provider:	Allan Herbarium
Barcode:	CHR 559987 A
Type status:	Holotype
Specimen type:	Sheet
Database record added:	13 June 2002
Database record updated:	24 February 2023

## Components

### Primary component

Active identification

<b>Habitat:</b>	Blowout in granite sand on ridge, with <i>Chionochebe pulvinaris</i> <i>Dracophyllum pronum</i> , <i>Poa colensoi</i> , <i>Luzula pumila</i> .
<b>Notes:</b>	Monocarpic, perhaps biennial or triennial but some plants flowering each year. Pollinator seen: bumble bee.  Common from this point north along the range, abundant wherever there is bare granite sand. <i>Gentianella</i> "Lookout". Specimen over 2 sheets.

## Specimen notes

<b>Type:</b>	HOLOTYPE of <i>Gentianella luteoalba</i> Glenny. <i>New Zealand Journal of Botany</i> 42(3): 361-530 (2004).
--------------	--

## Permissions

### Project permits

**Project title:**  
[Local Contexts - Allan Herbarium \(CHR\)](#) 

**Reference:**  
CHR Collection - Local Contexts



Biocultural (BC) Notice

[Terms of use](#) | [Feedback & enquiries](#)

© 2013 - 2023 Version 2.3.0.1323

# About the Dataverse Project

- The Dataverse Project is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, journals, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility.
- A Dataverse repository is the software installation, which then hosts multiple virtual archives called Dataverse collections. Each Dataverse collection contains datasets, and each dataset contains descriptive metadata and data files (including documentation and code that accompany the data).
- The central insight behind the Dataverse Project is to automate much of the job of the professional archivist, and to provide services for and to distribute credit to the data creator.
- <https://dataverse.org/about>

# 99 Installations



# How do we represent indigenous knowledge and rights in the Dataverse platform?

## Local Contexts and TC/BC Labels and Notices: Displaying Notices

- The text, icons, and titles for the Notices cannot be changed.
- The only changes that can be made to the descriptive text is replacing the placeholder “our institution” with the name of the institution displaying the Institution Notices.
- **The Notices are intended to be displayed prominently** on public-facing institutional websites, on digital collections pages, and or in finding aids.
- **When displaying the Notices, the title, icon, and text must all be displayed.**
- The Notices should be applied at a metadata field level.

## Dataverse:

- DDI based
- Capable of presenting broad range of metadata
- Includes licensing, discovery metadata, access controls, ...
- BUT:
- Designed for representation of metadata consistent with traditional notions of rights, knowledge and access
- How do we enable and appropriately reflect indigenous knowledge and rights within a “traditional” repository framework

# What would we need to incorporate?

- Representation of knowledge (e.g. TK/BC Labels)
- Recognition of indigenous ownership rights (data access)
- Curation of knowledge
- Presentation of this information in the Dataverse portal

# The approach

- Established discussion between Dataverse, Local Contexts and active indigenous and non-indigenous service providers, researchers and developers, following the 2022 Dataverse Community Meeting
- Series of meetings from September 2022 to present, looking at scoping of possible project requirements and use cases
  - Included case from the ADA associated with the Indigenous Data Network project in Australia
- Opportunity to pilot some integration as part of the NIH GREI program, of which Dataverse is a part
- Proposal currently in development with NIH through Harvard as Dataverse project leads

# Pilot tool (pre-project)

- RIDAGOP toolkit: <https://toolkit.ridagop.net>
- Developed by Ramin Soleymani-Fard and the team at the Social-Ecological Systems Research Group, Laboratory for the Analysis of Social-Ecological Systems in a Globalised world (LASEG) is a research group at the Institute of Environmental Science and Technology, Universitat Autònoma de Barcelona (ICTA-UAB)
- Collaboration with Local Contexts
- Demonstration of toolkit: <https://dataverse.org/dataversetv>, April 2023

## Assign Local Contexts Labels to dataverse datasets



Local  
Contexts

The  
**Dataverse**  
Project

### Local Contexts Project ID

Copy the Project ID of your Local Contexts Project, from which you want to use the labels. The ID can be taken from the Identifiers section of the project description or the project address

Fetch Labels

▼ See the project details here, once you fetched them

### Dataverse details

Please specify the dataverse instance (e.g. <https://data.harvard.edu/dataverse>), the repository DOI (which you find in the metadata) and the API key

Dataverse instance:

DOI:

API Key:

Read dataset



# Integrated with Dataverse CV support

Dataverse depos

One metadata field linked to many ontologies

A screenshot of the Dataverse deposit form. It shows several sections: Geographic Coverage (Country/region: Amsterdam, City:), Geographic Unit, Geographic Bounding Box (West Longitude, North Latitude), Unit of Analysis (Vocabulary: un|, thesaurus, grid, agrovoc), Universe, and Time Method (Vocabulary: unesco, Vocabulary URL).

Every field could be linked to

A screenshot of the Dataverse Subject field configuration interface. It shows a 'Subject' dropdown menu and a 'Keyword' section. Below these are several rows of configuration for different vocabularies. Each row includes a 'Vocabulary' field, a 'Vocabulary URL' field, and a 'Term' field. The 'Term' field is highlighted in blue, and a list of suggested terms is shown below it.

Vocabulary	Vocabulary URL	Term
unesco	http://skos.um.es/unescothes/C01489	Family
thesaurus	http://vocabularyes.irstea.fr/thesaurus/T-	family labour
agrovoc	http://aims.fao.org/aos/agrovoc/c_2785	families
iptc		fam

- family
- family planning
- famine

Language switch in Dataverse will change the language of suggested terms!

# Next steps

- Ongoing engagement with NIH on funding for pilot project
- Scoping of requirements
  - Integration methods for connecting Dataverse to Local Contexts (via LC API services and
  - Suitable user interface for presentation
  - Curation workflows for including labels and notices
    - That align with community and repository practices and capabilities
  - How do align to notices from specific communities
- Proposal to NIH expected in second half of 2023
  - More on this from Sonia