

Bringing Data to Life: Co-Designing a Language Data Commons

A Summary Report

Dr Harriet Sheppard 0000-0002-8525-5699,
Dr Simon Musgrave 0000-0003-3237-9943
December 2022

CONTENTS

| | |
|---|-----------|
| CONTENTS | 1 |
| ACKNOWLEDGMENTS | 2 |
| PREAMBLE | 3 |
| PANEL 1 - CO-DESIGN | 4 |
| Panellists: | 4 |
| Summary: | 4 |
| PANEL 2 - INDIGENOUS LANGUAGES | 6 |
| Panellists: | 6 |
| Summary: | 7 |
| PANEL 3 - ORAL HISTORY | 8 |
| Panellists: | 8 |
| Summary: | 8 |
| ROUNDTABLE | 10 |
| Governance | 10 |
| Communities | 10 |
| Capacity Building | 10 |
| Enriching data - 'Singing data to life' | 11 |
| NEXT STEPS | 12 |
| SUPPORTING MATERIALS | 13 |

DOI: [10.5281/zenodo.7425924](https://doi.org/10.5281/zenodo.7425924)

Suggested citation:

Sheppard, H, Musgrave, S. Bringing Data to Life: Co-Designing a Language Data Commons. 2022.

Viewed online at:

<http://doi.org/10.5281/zenodo.7425924>

ACKNOWLEDGMENTS

The report authors acknowledge and pay respect to the Elders and Traditional Owners of the lands on which we live and work.

This report draws on notes by Dr Leah Gustafson 0000-0003-2760-1044 and Jo Savill 0000-0001-6933-1025.

The Language Data Commons of Australia (LDaCA) project received investment ([1](#) and [2](#)) from the Australian Research Data Commons (ARDC). The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS).

PREAMBLE

Bringing Data to Life: Co-Designing a Language Data Commons, held on Yuggera and Turrbal land at the Brisbane Powerhouse, brought together Aboriginal and Torres Strait Islander and non-Indigenous leaders from academia, galleries, libraries, archives, and museums (GLAM) institutions to discuss co-designing a [Language Data Commons for Australia](#).

The event was moderated by Gooreng Gooreng man Grant Sarra, and attended by over 70 people from academia and the GLAM sector both in person and online on 4 August 2022.

The day began with 3 panel discussions in the morning and was completed by a roundtable discussion in the afternoon. These sections are summarised here.



The audience at the Bringing Data to Life event. Image: ARDC/Marc Grimwade

PANEL 1 - CO-DESIGN



Panel 1: Rose Barrowcliffe, Sandra Phillips, Louise Denoon, Michael Haugh. On the left is Grant Sarra, the event facilitator. Image: ARDC/Marc Grimwade

Panellists:

Sandra Phillips, Wakka Wakka and Gooreng Gooreng woman, (University of Queensland)

Rose Barrowcliffe, Butchulla woman, (Queensland State Archives)

Louise Denoon (State Library of Queensland)

Michael Haugh (University of Queensland)

Summary:

Co-design cannot be a piece of jargon behind which we hide: it has to have specific meaning and it should be based on articulated principles (such as CARE and FAIR) and their operationalisation. The co-design process has to begin from the acknowledgement that we are on stolen land. Understood in this way, co-design must crucially involve communities, giving them equal power in the process, and

seeking benefits for them. This is a slow process which depends on institutions listening rather than assuming, adopting trauma informed practice, and learning to acknowledge what is uncomfortable.

Language data is diverse and complex, covers a range of types of material in many languages, and originates with people whose rights must be respected. Institutions are already rethinking legal and moral rights that they apply to such data. In a language data commons, data is not the common property and, therefore, is not, in its entirety, automatically accessible to anyone. Rather, the structures and processes which handle data are the commons. Co-design therefore can apply at two levels: in deciding how particular data should be handled and in developing the common elements. Although the way forward may be difficult, doing nothing is not an option.

PANEL 2 - INDIGENOUS LANGUAGES



Panel 2: Beau Williams, Des Crump, Clint Bracknell and Nick Thieberger. Kirsten Thorpe and Lauren Booker participated virtually. Image: ARDC/Marc Grimwade

Panellists:

- Clint Bracknell, Noongar man (University of Queensland)
- Des Crump, Gamilaroi man (University of Queensland)
- Beau Williams, Murrawarri man (First Languages Australia)
- Nick Thieberger (University of Melbourne)
- Kirsten Thorpe, Worimi woman (UTS Jumbanna)
- Lauren Booker, Garigal woman (UTS Jumbanna)

Summary:

There are still barriers in accessing Indigenous language data in Australia. In some cases, the difficulties are practical and can be (and are being) addressed. For example, variation in naming and spelling can be barriers to discovering materials but the standard language codes (AustLang) developed by AIATSIS can help (where they are used). In other cases, policies and practices inhibit access and institutions may need to take a more proactive role. For example, since language data can be buried in jargon-ridden academic publications, researchers could do much better in making the results of their work accessible.

Examples exist both of researchers taking the initiative and making material more accessible, and of complex co-design processes having excellent outcomes. Actions which lead to the reduction or elimination of such barriers to access should not be driven by the goals of researchers but should be seen as benefits to communities. Data collection and the access to the ensuing data can bring with it the possibility of trauma. Our discovery procedures should deal with such a possibility. Giving communities access to as much data as possible can lead to a range of positive outcomes including better opportunity for language education and for sustaining (and even extending) culture.

PANEL 3 - ORAL HISTORY



Panel 3: Alistair Thomson, Jacqui Uhlmann, Vanessa Russ, Sophia Sambono. Grant Sarra is on the far left. Image: ARDC/Marc Grimwade

Panellists:

Alistair Thomson, (Monash University)

Jacqui Uhlmann, (National Film and Sound Archive (NFSAC))

Vanessa Russ, Ngarinyin and Gija woman, Indigenous Data Network (University of Melbourne)

Sophia Sambono, Jingili woman, (Queensland Art Gallery & Gallery of Modern Art (QAGOMA))

Summary:

Comparing oral history data with Indigenous language data reveals a number of common problems. What can be considered oral history is not always clear, in the same way that Indigenous language data may be part of materials which an institution classifies under some other label. Both types of data are stored in numerous locations, leading to problems of discovery and access, with some collections in need of urgent preservation work. The two types of data also have similarities in the relationship between the data and the community from which the data originates.

This suggests that strategies for enabling communities to enrich data descriptions, such as advising on the use of Traditional Knowledge labels, can be applied to both types of data. Additionally, in both cases, the data can be the basis for further cultural activity, which emphasises the idea that cultural activity needs people rather than physical infrastructure.

However, there are important differences between the two types of data. One obvious difference is that there is an assumption that the participants in oral history research wanted their stories to be told and heard and that, therefore, data governance issues are less complex.

ROUNDTABLE

While the roundtable discussion covered a wide range of topics, some themes emerged.

Governance

Data governance was an important theme in the discussion, centring on the closely linked topics of access and risk management. Participants spoke of the eagerness of people to access data and the barriers which often prevent this, while acknowledging that it was necessary to control access in some cases and that finding those with the authority to make such decisions was not straightforward. GLAM institutions have to find a balance between safety and accessibility and an endeavour such as LDaCA also faces the same problem. Risk aversion leads to overly restrictive access controls but the wide range of people participating in a discussion of such issues was welcomed as the start of positive moves. Legal rights and moral rights affect access in different ways and institutions increasingly make the distinction between these rights in their practice. However, it was suggested that there are legal issues which needed to be addressed and that the roundtable group and the LDaCA project provide an opportunity to do this.

Communities

Community was a central theme of discussion throughout the afternoon. Although the discussion focussed on Aboriginal and Torres Strait Islander communities, many of the points raised could be of relevance for other communities, particularly those that use non-indigenous minoritised languages. Participants emphasised that meaningful community involvement and relationship building need to be central for work with language data but pointed out that such processes do not typically align with traditional institutional funding timelines and, therefore, time and resources for these activities should be explicitly built into timelines and funding applications. Participants spoke of the importance of tapping into Indigenous networks to connect with communities but to also be aware of potential gatekeeping and be conscious of the fact that establishing who has authority in a community can be complex. Difficulties surrounding community members accessing data were also raised, particularly when such records are often held by institutions and not on Country. Navigating institutions can be difficult and participants recognised the importance of strengthening and leveraging networks between institutions housing or collecting data in order to help communities locate and access data.

Capacity Building

Capacity building was another theme that emerged. It was noted that an important difference in our discussion and those that were possible several decades ago is the presence of a very capable cohort of Indigenous researchers today. This, however, was taken as a starting point on which to build further with people valued as the most important infrastructure we have. Sharing of knowledge and skills is part of the co-design process, but a concern was expressed that in too many cases this kind of work falls on a few community members who risk being overburdened. One way to mitigate this is by involving community members as employees and at governance level in projects and organisations.

Enriching data - ‘Singing data to life’

The need to enrich existing data was discussed at length. The provenance of objects, particularly of older ones, is often lacking, missing information about where an object originated, names of people associated with the object, language name(s), and language codes (such as AustLang and ISO 639 codes). Adding missing metadata to objects and collections aids future discoverability, making it easier for community members and researchers to find relevant data. It also recontextualises the data and was viewed by some participants as a decolonising practice. It is a step that can be taken to make archives and collections more trauma-informed and culturally safe by making sure community members do not view distressing records without warning. It was recognised that such work is labour-intensive and there is often limited funding to do it. Community members and other users were also identified as important in the data-enriching process, for example, through institution-organised or facilitated events (where community members can interact with and discuss objects) and through interactive options in online catalogues (where users can add information about an object to its record in the archive).

NEXT STEPS

A specific action was discussed: Moves should be made to encourage all Australian cultural institutions to use AustLang codes in their cataloguing and metadata to improve the discoverability of resources. This process might begin with a letter from the participants in the roundtable.

SUPPORTING MATERIALS

Watch the video recordings of the panel discussions:

- [Panel 1 recording](#)
- [Panel 2 recording](#)
- [Panel 3 recording](#)




Read a wrap up article and view images from the event on the [ARDC website](#).

Learn more about the [Language Data Commons of Australia \(LDaCA\)](#).






Australian Research Data Commons

CONTACT

-  ardc.edu.au
-  +61 3 9902 0585
-  contact@ardc.edu.au

FOLLOW

-  [@ardc_au](https://twitter.com/ardc_au)
-  [australian-research-data-commons](https://www.linkedin.com/company/ardc-au)
-  [subscribe to our newsletter](#)



The ARDC
is enabled
by NCRIS