



Australian Research Data Commons

# HASS Research Data Commons and Indigenous Research Capability project plan

Activities are required to submit a project plan as the first deliverable in their contract. It should outline the high level aims, detail the milestones and deliverables, define the project team, budget breakdown and assess risks. The format and layout of the plan is up to individual projects - you may use your own template - but please ensure you have covered all the areas listed below.

## Revision History

Version	Date	Editor	Summary of changes
1.0	08/09/2021	Michael Haugh	Draft project plan

## 1. Project title

HASS RDC and Indigenous Research Capability: Language Data Commons of Australia (LDaCA)

## 2. Lead contact

- First name: Michael
- Last name: Haugh
- Organisation: The University of Queensland
- Group/Department: School of Languages and Cultures
- Job title: Professor of Linguistics and Applied Linguistics
- State: Queensland
- Email: michael.haugh@uq.edu.au
- Phone number: (07) 3365-7221
- ORCID: 0000-0003-4870-0850

## 3. Proposal summary

Australia is a massively multilingual country, in one of the world's most linguistically diverse regions. Significant collections of this intangible cultural heritage have been amassed, including collections of Australian Indigenous languages and regional languages of the Pacific, and of Australian English, as well as collections important for cyber-security (AusTalk, Australian National Corpus, corpora of regional languages), for gauging popular sentiment (Australian Twitter Corpus), and for emergency communication (languages of the region and some Indigenous languages). Many collections remain under-utilised or at risk, and researchers lack the tools and skills to exploit their research potential. The Language Data Commons of Australia (LDaCA) will capitalise on existing infrastructure to secure vulnerable and dispersed language collections and link these with improved analysis environments. Establishing an integrated national infrastructure that supports language research will enable researchers and communities to access and use nationally significant collections of written, spoken, multimodal and signed text. The project will improve researchers' digital skills and raise awareness of best practice in digital research; render valuable collections of national significance more FAIR while adhering to CARE principles; and develop the integrated national technical infrastructure to analyse language collections at scale. It will support researchers to deliver innovative research outcomes, and will open up the social and economic possibilities of Australia's language data for translational research in the national interest. The project will build connections to other HASS RDC projects by: developing APIs and text analytics tools that can be applied to the Trove collections; facilitating text analysis of aggregated administrative data collections from AusIRISS; and developing a community-driven approach for governance of Indigenous language collections. We will address the challenge of balancing research needs while respecting community rights for language/cultural collections; highlight contributions that language research and HASS disciplines can make to STEM research and non-academic applications; and position Australia internationally as a leading contributor of language collections and digital infrastructure.

## 4. Project partners

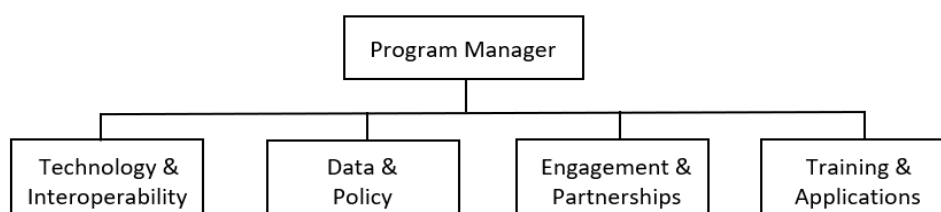
Provide a short profile of the project partners that are collaborating on the proposal (this may be used in comms generated by ARDC).

- The University of Queensland
- Australian National University
- Monash University
- The University of Melbourne
- The University of Sydney
- AARNet
- First Languages Australia (FLA)
- Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS)
- PARADISEC
- ARC Centre of Excellence for the Dynamics of Language (CoEDL)
- CLARIN

## 5. Project team roles and responsibilities

Stakeholders in the project are represented through the Steering Committee, Technical Advisory Group, Researcher Advisory Group and the Indigenous Languages Working Group (for further details see Section 10: Governance).

The Project Team consists of four work groups overseen by the Program Manager:



The project Stakeholder and Team roles are summarized in the table below:

Name	Project role	Organisation	Responsibility
Michael Haugh	Director	UQ	Lead the project; Chair steering committee; Work packages 2.1, 2.3, 3.1, 3.2, 4.1, 4.2
Catherine Travis	ANU Lead	ANU	Work packages 1.3, 2.2, 4.2
Louisa Willoughby	Monash Lead	Monash	Work packages 1.3, 1.4
Nick Thieberger	Melbourne Lead	Melbourne	Work packages 1.1, 1.2
Monika Bednarek	Sydney Lead	Sydney	Work packages 3.2, 4.4
Faith Baisden	FLA Lead	FLA	Work package 1.1, 2.1, 4.3
Carina Kemp	AARNet Lead	AARNet	Work packages 2.3, 3.1
tbc	AIATSIS Consultant	AIATSIS	Indigenous Languages Working Group
Franciska de Jong	CLARIN Consultant	CLARIN	CLARIN knowledge exchange
Marco Fahmi	Manager	UQ	Project management; oversight and control
Peter Sefton	Technology &	UQ	Technical infrastructure and information

	Interoperability Group Lead		exchange; Work packages 2.1, 2.2
tbd	Software Developer	Melbourne	Software development and operation of infrastructure; Work packages 1.1, 1.2
tbd	Software Developer	ANU	Software development and operation of infrastructure; Work package 2.2
tbd	Software Developer	Monash	Software development and operation of infrastructure; Work package 2.2
tbd	Software Developer	UQ	Software development and operation of infrastructure; Work packages 2.3, 3.1, 3.2
tbd	Software Developer	AARNet	Software development and operation of infrastructure; Work packages 2.3, 3.1
Kathrin Kaiser	Data and Policy Group Lead	UQ	Data availability, standards, quality and accessibility; Work packages 1.3, 2.1, 2.2, 4.2
tbd	Data Collections Specialist	Melbourne	Data quality and accessibility; Work packages 1.1, 1.2
tbd	Data Collections Specialist	Sydney	Data quality and accessibility; Work package 1.2
tbd	Data Specialist	ANU	Data quality and accessibility; Work packages 1.2, 1.3, 2.2, 4.2
tbd	Data Specialist	Monash	Data quality and accessibility; Work packages 1.3, 1.4
Simon Musgrave	Engagement & Partnerships Group Lead	UQ	Community engagement, outreach and education; Work packages 4.1, 4.2
Leah Gustafson	Communication Officer	UQ	Communication and partner liaison; Work packages 4.1, 4.4
Sara King	Engagement lead	AARNet	Community engagement, outreach and education; Work package 4.4
Des Crump	Indigenous Industry Fellow	UQ	Indigenous community partnerships and projects; Work packages 1.1, 2.1, 4.3
Robert McLellan	Indigenous Industry Fellow	UQ	Indigenous community partnerships and projects; Work packages 1.1, 2.1, 4.3
tbd	Indigenous Community Consultants	FLA	Indigenous community partnerships and projects; Work packages 1.1, 2.1, 4.3
Mel Mistica	Training and Applications Group Lead	UQ	Research analytics and training; Work packages 3.2, 4.4
tbd	Data Scientist	UQ	Research applications and training using language collections; Work packages 3.2, 4.3, 4.4
tbd	Data Scientist	Sydney	Research applications and training using language collections; Work packages 3.2, 4.4

## 6. Project objectives

Australia has many large collections of language data, but many remain under-utilised or at risk. This project capitalises on existing infrastructure to secure vulnerable and dispersed language collections of written, spoken, multimodal and signed text, and to link these with improved analysis environments for new research outcomes. The Languages Data Commons of Australia (LDAcA) aims to establish a sustainable long-term repository for ingesting and curating language data collections of national significance: to democratise access to Australia's rich linguistic heritage through enabling those collections to become more FAIR while following the CARE principles; to develop the computational capabilities, technical infrastructure and support services to analyse language collections at scale; to increase the awareness and skills of researchers in applying digital methods; and to open up the social and economic possibilities of Australia's language data for impactful research with significant benefits to the nation. Three key principles shape LDAcA's activities and determine its success:

### 1. Capitalising on existing infrastructure

Significant collections of language data amassed through past ARC, ANDS and NeCTAR investments will be leveraged, alongside current investments by ARDC (Australian Text Analytics Platform, PL074; Language Data Commons of Australia Data Partnerships, DP768) and the ARC (Nyingarn Platform, LE2001000006), to establish a robust, distributed national infrastructure for language-based Humanities research. The project will build on existing NCRIS capabilities by working with established research cloud providers to ensure continuity of service, using industry-standard open-source tools and libraries to avoid software vendor lock-in, partnering with long term custodians of data to ensure ongoing hosting of collections, and reusing standard architectures and software stacks of proven technology solutions.

### 2. Securing vulnerable and dispersed collections (see Overview of Technical Architecture of LDAcA).

It is hard for researchers and communities to access many of Australia's trove of internationally significant collections of Indigenous and non-Indigenous languages of Australia and its region. Collections need to be hosted on durable infrastructure that ensures that those collections are given perennity and security. LDAcA will work with key groups to secure vulnerable and dispersed collections, and to demonstrate how to balance research needs with preserving community rights.

### 3. Linking with improved analysis environments for new research outcomes

The significant potential of language collections in Australia remains largely untapped because of limited skills and use of tools by Australian researchers and communities. New language data sources such as the web and social media are available, but they require large scale analytics capability. LDAcA will develop the technical infrastructure to analyse language collections at scale, and highlight the contributions that language research and HASS disciplines can make to STEM and the health sciences, and to translational research.

The LDAcA HASS RDC project objectives will be implemented through thirteen work packages that fall into four inter-related activity streams. These work packages build on and complement the six work packages delivered through the LDAcA Data Partnerships project (see LDAcA DP Project Plan), and the five work packages delivered through the Australian Text Analytics Platform (ATAP) project (see ATAP Project Plan).

The work packages in Stream 1, Securing Language Data Collections, will focus on securing at-risk or under-utilised language data collections, through securing data, and associated metadata, as preservable digital objects, and providing access to that data in culturally, legally and ethically appropriate ways.

The work packages in Stream 2, Aggregating Language Data Collections, make language data in different institutional collections and repositories more FAIR. They enable researchers to aggregate language data dispersed across institutional repositories, and make those collections more findable and accessible, adhering to CARE principles..

The work packages in Stream 3, Improving Text Data Analysis Environments, develop specialist tools for text analytics, and extend the text analytics workbench to enable large-scale computational analysis of written, spoken, multimodal and signed language data, and to share those workflows with other researchers.

The work packages in Stream 4, Strategic Partnerships, Engagement and Training, build national and international partnerships for long-term sustainability of LDAcA, engage researchers and communities in LDAcA, and increase the capacity of Australian researchers to access and analyse different forms of language data.

## 7. Project activities

### ***Activity Stream 1: Securing language data collections***

#### **WP1.1: Secure nationally significant Aboriginal and Torres Strait Islander language data collections**

Through consultation with communities, vulnerable collections of Aboriginal and Torres Strait Islander language data not currently secured will be selected and secured as preservable digital objects using the Research Object Crate (RO-Crate) and the Oxford Common File System Layout (OCFL) implemented in PARADISEC. Access to these collections will be secured through the development of cultural access protocols for Indigenous communities and researchers. Deliverables include a report on at-risk collections of Aboriginal and Torres Strait Islander languages, migration of community-identified vulnerable collections of Aboriginal and Torres Strait Islander language data into RO-Crate/OCFL formats, and an access framework for Indigenous researchers and communities and non-Indigenous researchers. This work will build on the same infrastructure being developed for Nyingarn, an ARC LIEF project in 2021-2024.

This work package will be led by First Languages Australia, the University of Melbourne and UQ in consultation with AIATSIS.

#### **WP1.2: Secure internationally significant Indigenous language collections in Australia's Pacific region**

Building on established relationships with agencies in the Pacific, vulnerable collections of language data not currently secured by PARADISEC will be selected and secured as preservable digital objects using RO-Crate and OCFL. This will involve working with relevant individual researchers and regional agencies to catalogue and store audio and video records, and textual collections (for example, microfilm and paper collections). Deliverables include securing vulnerable collections of Indigenous language data collections of the, accession into PARADISEC and then into RO-Crate/OCFL formats.

This work package will be led by the PARADISEC team at the University of Melbourne, ANU and University of Sydney.

#### **WP1.3: Secure nationally significant Australian English and migrant language data collections**

This Work package focuses on the large number of collections of Australian English and migrant language data, existing since the 1960s, not all of which are currently accessible, and which are highly dispersed. This includes collections compiled for linguistic analysis, as well as oral histories, and other language data that captures a social history of Australia. Nationally significant collections of Australian English and migrant language data will be secured as preservable digital objects using RO-Crate and OCFL, and access protocols will be developed.

Deliverables include a report on at-risk collections of Australian English and migrant language data, migration of Australian English and migrant language data into RO-Crate/OCFL formats, and an access framework for Australian researchers and communities.

This work package will be led by ANU and Monash with support from UQ.

#### **WP1.4: Secure nationally significant sign language data collections in Australia and its region**

Nationally significant collections of sign language data of Australia and its region will be secured as preservable digital objects using RO-Crate and OCFL, and access protocols for Australian researchers and communities will be developed. This will involve using Signbank – the online dictionary of Auslan – and the Auslan Corpus, and working with sign language and gesture researchers, to deposit annotated multimodal video data and dictionary resources. Deliverables include migration of selected sign language data collections into RO-Crate/OCFL formats, the development of access protocols and resources for Australian researchers and communities, including Auslan teachers and Deaf Australians.

This work package will be led by Monash University.

### ***Activity Stream 2: Aggregating language data collections***

#### **WP2.1: Aggregated discovery of Aboriginal and Torres Strait language collections in institutional repositories**

Develop a portal to make Aboriginal and Torres Strait language data findable across selected libraries, archives and museums in Queensland through a web interface. Deliverables include increased accessibility to collections of Aboriginal and Torres Strait language data held in Queensland institutional repositories for Australian researchers and communities, and a worked framework for aggregated discovery of Aboriginal and Torres Strait Islander language data in GLAM institutions in other jurisdictions in Australia.

This work package will be led by UQ in consultation with AIATSIS and FLA.

#### **WP2.2. Aggregated discovery of specialist language data collections in institutional repositories**

Develop a portal to make Australian English, migrant language and sign language collections held in institutional repositories findable and accessible through a web interface. Deliverables include increased accessibility to Australian English, migrant languages, Auslan and other sign language collections for Australian researchers and communities.

This work package will be led by ANU and Monash with support from UQ.

#### **WP2.3: Enable machine-access to national language data collections**

Enable access to national significant collections of language data through the use of application programming interfaces (APIs) that will allow users to analyse and aggregate language data collections. Deliverables include creating APIs for nationally significant collections within LDaCA, and enabling access to language data collections in institutional repositories for researchers through the Australian Text Analytics Platform (ATAP).

This work package will be led by UQ and AARNet.

### ***Activity Stream 3: Improving text data analysis environments***

#### **WP3.1: Improve analytical capabilities of the text analytics workbench**

Create the tools and infrastructure necessary to support analysis at scale of language collections, including unstructured text and multimodal collections, and for researchers to be able to share those workflows through BinderHub. Deliverables include improvements in the usability of data in language collections, including making them more machine-readable, support for the application of machine learning techniques, and the integration of BinderHub into the Australian Text Analytics Platform.

This work package will be led by UQ and AARNet.

#### **WP3.2: Specialist language data processing and mining tools**

Extend ATAP through adapting existing tools for quantitative and qualitative text analytics for domain-specific applications. Deliverables will include open-access tools and resources to enable preparation of language data collections for analysis, and demonstrator examples that can be used by the research community.

This work package will be led by UQ and University of Sydney.

### ***Activity Stream 4: Strategic partnerships, engagement and training***

#### **WP4.1: Develop strategic partnerships with key international stakeholders**

Develop a working relationship with CLARIN and other selected international university partners to ensure best practice. Deliverables include participation of CLARIN in the technical development of LDaCA, a series of joint CLARIN-LDaCA knowledge exchange workshops, and the development of roadmap plans to join CLARIN as a Third Party Member.

This work package will be led by UQ with support from CLARIN.

#### **WP4.2: Develop strategic partnerships with key national stakeholders**

Develop a working relationship with key custodians of nationally significant language data collections, particularly in the GLAM sector, to develop a joint roadmap for the long-term preservation and sustainability of those

collections. Deliverables include the formation of a working group with key stakeholder institutions, and the development of roadmap plans for long-term preservation and sustainability of language data collections in Australia.

This work package will be led by UQ and ANU.

#### **WP4.3: Develop training resources for Aboriginal and Torres Strait Islander language researchers and communities**

Engage with Aboriginal and Torres Strait Islander language researchers and communities to understand how they wish to access and use language collections and develop training that meets those needs. Deliverables include a series of engagement and training workshops, online training modules and research support tailored to the needs of Aboriginal and Torres Strait Islander language researchers and communities.

This work package will be led by First Languages Australia and UQ in consultation with AIATSIS.

#### **WP4.4: Stepwise capacity building of Australian language researchers**

Increase capacity of Australian researchers to leverage national language data collections and produce cutting edge research that also benefits communities through better access to information, education and training in the use of digital tools. Deliverables include a series of engagement and training workshops, online training modules and research support for researchers working with text.

This work package will be led by AARNet, Monash University, University of Sydney and UQ.

## 8. Integration Component

Describe, in detail, elements of your proposal that will utilise RDC wide developments to be supported by the ARDC Integration Component.

Working with the Indigenous Data Network, Trove, and the Integrated Social Sciences Infrastructure, the project will identify common infrastructure needs, including data storage, compute and access infrastructure.

There are significant synergies between LDaCA and the other three HASS RDC projects, including a commitment to a community-driven approach to access and governance of Indigenous and non-Indigenous language and cultural collections, the need for authentication and authorization of research groups and communities, and the pressing need for tools and skills to enable data-intensive humanities research.

Possible shared work packages include:

#### 1. Authentication and authorization

Enable research groups and communities that do not currently have AAF credentials to access LDaCA.

#### 2. Compute: HPC and GPUs

To support use of machine learning/AI in text analytics.

#### 3. Coordination of HASS RDC community consultation, engagement and training across IDN, Trove and the Integrated Social Sciences Infrastructure

NB. We believe there may be some place for engagement with the GLAM workbench in this shared work package

#### 4. HASS RDC Roadmap: (1) Governance framework; (2) Legal and cultural data access framework; (3) Technical architecture; (4) Building communities.



## 9. Outcome and Impact

Provide use cases which demonstrate how the deliverables will support research programs. (1400 characters<sup>1</sup>)

### **Use Case 1: Closing the Gap through Supporting Australian Indigenous Languages**

The role that the maintenance and reawakening of Indigenous languages can play in improving well-being in Australian Indigenous communities has been recognised in the recent National Agreement on Closing the Gap. Outcome 16 of the Agreement is that “by 2031, there is a sustained increase in number and strength of Aboriginal and Torres Strait Islander languages being spoken”. One bottleneck in the maintenance or reclamation of Indigenous languages in Australia is identifying relevant collections and obtaining access to that material, as data from the same language may be held across different institutions, such as the AIATSIS collections, National Library of Australia, State Libraries, Australian university libraries, and other language archives, including PARADISEC and CoEDL collections. This project will enable community researchers to more readily identify language data held across different archives and will facilitate access to that data. A key aspect of outreach activities will be to make community researchers aware of this capability and to provide training in using it effectively, ensuring timely access – as is critical if community researchers are to take advantage of momentum around support for Indigenous languages. This project will involve an interdisciplinary team including Indigenous community language professionals, data scientists, linguists, librarians and information specialists.

### **Use Case 2: News coverage of obesity**

It is difficult for researchers as well as NGOs/government organisations in public health to analyse how health conditions and diseases (e.g. obesity, diabetes, covid-19) are covered in the media on a large scale basis. Relevant research tends to be qualitative and focus on small datasets, or be quantitative and time-consuming because of manual content analysis. Alternatively, institutions may rely on costly outsourcing of such quantitative analysis to experts. Despite the clear potential for text analytics to contribute to big data analysis of public health communication, its application has thus been limited. The text analytics platform in LDaCA will enable researchers to import newspaper articles and policy or other documents on health, and will provide the tools needed to undertake analysis of a set of health topics of their own choosing. Tools such as named entity recognition and automatic identification of quotation will allow people and organisations that are most influential in health coverage to be traced, and will also allow biased or misleading frames of communication to be identified. This will enable stakeholders to examine large datasets on how health issues are communicated to the public, from which qualitative follow-up studies can be undertaken. This project will involve an interdisciplinary team including data scientists, public health scholars and linguists.

### **Use Case 3: Alternative discourses of memorialisation**

There is a well-developed official discourse tradition in Australia which is used to memorialise the experience of war and of Australians who have been members of the armed forces. Australia also has a strong subversive tradition of language use, often given the label 'Australian Slang'. Such language was a feature of Australian life from early in the period of European settlement and has been maintained and renewed consistently. It is not hard to find sources which present a counter-narrative to the official account of war, but do these contribute to an alternative discourse? Is there a distinct set of linguistic resources which have been used in such an alternative discourse, and if it exists, who participated in it? These questions can be addressed by applying machine learning methods to Australian textual material from the time of the First World War on. For example, training classifiers on key terms such as 'the fallen' will enable identification of other expressions used in similar contexts, enabling researchers to identify various discourses around the topic of war. These methods will provide important knowledge about what war has meant to Australians across time, how different social groups have experienced war and its effects, and the set of Australian linguistic resources deployed for these purposes. This project will involve an interdisciplinary team including data scientists, historians, linguists and cultural studies scholars.

**Use Case 4: Research translation for the Australian deaf community**

An online, interlinked Signbank and Auslan Corpus will be a boon for Auslan researchers and communities. It will allow people accessing Signbank (the dictionary) to see examples of the sign being used in a sentence; or conversely for those viewing the corpus to view the citation form of a sign (i.e. the 'careful pronunciation'). These are extremely valuable resources for language learners to see how words/signs are used in context and to understand how grammatical features of sign languages – such as facial expression or the movement path of the sign – can modify the meaning of a particular sign. It also allows the demonstration of subtle shades of meaning between two signs that may have similar meanings in English by seeing the different environments in which they are used for research translation applications. Auslan textbooks and classroom resources could be linked to Signbank and the Auslan Corpus to provide vocabulary lists and resources for community and industry. This project will involve an interdisciplinary team including data scientists, linguists, and educators.

**Use Case 5: Australian voices on pandemics**

From the 1960s, researchers have been recording Australian English, and vast collections exist across the country, representing an invaluable source of information not only about Australian English, but about Australian history, culture and society, as captured in the stories that are recorded. Such information is complemented by more recent recordings from other groups shaping Australian society, in particular newer arrivals to Australia who have been recorded in migrant languages. The voices captured across those collections form a record of Australian views on different social issues over more than a century, but the potential to explore changes in social attitudes through these collections has not been realised to date because researchers have not been able to locate and aggregate such data. This project will explore commentary on the Spanish flu in older collections of oral history (including the Bicentennial Oral History Collection) and compare this with contemporary discourse on the COVID-19 pandemic in recently collected data. It will involve an interdisciplinary team including data scientists, cultural studies scholars, historians, and linguists.

**Use Case 6: Next generation sentiment analysis**

Current approaches to sentiment analysis rely primarily on analysing the valence of words. However, it is widely understood that we communicate through more than words and studies of dialogue have demonstrated that inferences about sentiment are expressed through recurrent combinations of cues with respect to the specific timing or position of words and grammatical structures. By aggregating Australian collections of interactive data and making it accessible through LDaCA, researchers will be able to work in this area on a scale which is not currently possible. Understanding the inferences that are made available in communication allows a range of research problems to be addressed; these range from going beyond searches for explicit instances of hate speech in order to identify people with malicious intent (e.g. terrorist organisations), through to systematically tapping into inferences that can be drawn from communications when assessing the well-being of individuals in vulnerable communities (e.g. farmers at risk of depression). Such research is also of relevance to business and industry, as predictive models of inference will underpin the next generation of speech recognition, search engines, machine translation. This project will involve an interdisciplinary team including linguists, and computer scientists.

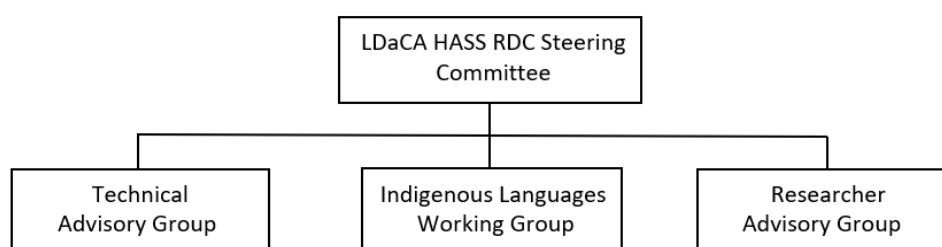
**Use Case 7: Offence and Freedom of Speech in Australia through the Lens of Big Data**

The taking of offence in workplaces and on social media can be devastating for individuals, undermine workplaces and polarise communities, and has implications for freedom of speech. There is, however, no clear and general understanding of what is meant by offence. Raising public awareness and developing policy in this area requires systematic examination of these phenomena in their social context. A data-driven approach to interrogating offence-terms across very large datasets facilitated by LDaCA would yield an understanding of how Australians conceptualise and talk about offence. This has significant implications for development of policy and laws in relation to freedom of speech in Australia, especially in light of the rapid shift to online communication driven by widespread uptake of social media. The project will aim to recruit an interdisciplinary group (including e.g. law, philosophy, politics, and linguistics) to work on this problem, providing an exemplar of the benefits of data-driven research crossing disciplinary boundaries.

## 10. Governance

The Program Manager and Project Group Leads (see Section 5) will report to a consolidating Project Steering Committee that will oversee the Language Data Commons of Australia Humanities, Arts and Social Sciences Research Data Commons (LDaCA HASS RDC) Project, the Language Data Commons of Australia Data Partnerships (LDaCA DP) Project, and the Australian Text Analytics Platform (ATAP) Project. The existing Technical Advisory Group and Researcher Advisory Group formed for the ATAP and LDaCA DP projects will be complemented by the formation of an Indigenous-led Indigenous Languages Working Group.

The Governance structure of the LDaCA HASS RDC project is summarized below:



The Project Steering Committee will be governed by its Terms of Reference in line with those previously developed for the ATAP and LDaCA DP projects. The Project Steering Committee will convene every two months to assess progress on project deliverables and milestones.

Members of Project Steering Committee:

- Michael Haugh – UQ (Chair)
- Faith Baisden – First Languages Australia
- Catherine Travis – ANU
- Jane Simpson – ANU
- Nick Thieberger – Melbourne
- Louisa Willoughby – Monash
- Monika Bednarek – Sydney
- Carina Kemp – AARNet
- Dan Angus – QUT (Digital Observatory)
- Invited representative from AIATSIS
- Invited representative from IDN
- Invited representative from Trove
- Jenny Fewster – ARDC (HASS RDC Program Manager)
- Marco Fahmi – UQ (Program Manager) (ex officio)

Members of Indigenous Languages Working Group (tbc)

Members of Technical Advisory Group (tbc)

Members of Researcher Advisory Group (tbc)