

Response to feedback to LDaCA's Project Plan 2024-2028

We received three submissions from the public providing feedback on the LDaCA project plan. We are grateful for the supportive comments from the respondents (indicated as R1, R2 and R3), as well as their views on potential issues or gaps with respect to the project and the broader HASS RDC and Indigenous Research Capability Program.

In terms of the overall project, R2 noted the potential benefit to the translation and interpretation sector, for example, sharing of data to help with evaluating outcomes of initiatives. WP2.3 includes 'Identification of significant collections ... of migrant language data through consultation with relevant linguistic and cultural groups'. We will aim to include representatives of the translation and interpretation sector in those consultations.

In terms of the overall communication of the plan, R3 thought the language used was highly technical and needed to be made accessible for a broader target audience. Training activities under WP3.1.02, WP5.1 and WP5.2 will address these issues.

The majority of the comments related to implementation of existing work packages within the plan.

1. (R2) Collection and/or maintenance of Auslan (Australian sign language) data: This is in the scope of Activity Stream 2.
2. (R2) Cross analysis of data sets, e.g. Indigenous and migrant data: LDaCA applies standards at the level of data packaging but not at the level of data structure. Therefore comparing different datasets is a problem for researchers to address. However, where there is potential for reuse, notebook-based workflows could be developed in collaboration with researchers.
3. (R3) For WP2.3, consultation with university-based language archives and GLAM organisations: These bodies are relevant. These institutions are included in the work and deliverables for WP2.3.
4. (R3) For WP3.2, careful consideration of ethics and consent for social media data collections: This is also covered by the work in Activity Stream 1.
5. (R3) For WP4.2, focus on workbooks in this work package compared to the discussion of moving away from notebooks in the Creative Arts co-design workshop: Notebooks are a sustainable way to enable innovative globally competitive research – they allow for low cost development of research tools that allow researchers to engage at a variety of levels from simply running an existing tool to learning advanced research tools. In line with the Community Data Lab principles, commonly used tooling can be turned into reusable libraries and packaged for broader use without the need for any interaction with code. We understand, however, through ongoing consultations with our stakeholders that notebooks are not the only way to support the analysis of text and language. Other types of tools for analysis will also be developed over the next four years.
6. (R3) For WP1.1.02, accuracy of dates; WP1.1.03 precedes WP1.1.02: The dates are correct. These two deliverables are not dependent on each other - working with individual collections as they are on boarded will inform the overall governance model which will abstract from what we learn and provide a framework for ongoing management

7. (R3) For WP1.2.01, meeting accessibility needs when producing written licensing guides: The guides will be written for the website using accessibility standards, ensuring screen readers can be used.
8. (R3) For WP1.5 and WP1.6, general applicability of these work packages, which are specific to institutions or collections: These work packages have the status of proofs of concept and pilot studies which will create clear use-cases and technology. Various institutions will be involved (UQ, QUT Digital Observatory, UniMelb and Batchelor) and the results will inform more general work in the future.
9. (R3) For WP2.3, identification of all relevant GLAM collections: This will be determined by the initial scoping at WP2.3.01. A national call-out will certainly be considered as an option.
10. (R3) WP3.1 addresses only annotation, whereas WP2.3 addresses discovery: This is an indicator of the different stages of work in these two areas. WP2 generally is concerned with securing language data collections (i.e. bringing material into the LDaCA ecosystem), while WP3 is concerned with data already available through the LDaCA portal.

There were also suggestions for additional work packages or integrations within the project. R1 provided a few possibilities.

1. For WP3.1, integration with Image Annotation Workbench (IAW): Nyingarn already exists as a tool for working with images of language materials, making them searchable and providing annotation possibilities. A scoping study would be required to decide whether integration with IAW would offer additional capability.
2. For WP3.3, enhancement of existing LDaCA portals based on requirements identified by the user testing phase of the ARDC Heurist 2024 sustainability project: Partnership with the Sydney Informatics Hub (The University of Sydney) continues in the new project plan and these issues will be addressed as part of that collaboration.
3. For WP4.2, expansion of notebooks to include mapping and annotation research use cases using the TLCMap API and the IAW API: Recent enhancements to the Oni and Crate-O tools have allowed incorporation of geolocation data and development of tools which exploit these possibilities will follow.

Additionally, R3 wanted to know if the project could provide support for the challenges associated with digitisation and transcription, and whether it was possible to include speech-to-text alongside text-focused analytics. LDaCA will work with data stewards to package data in standard formats and to ensure that metadata supports FAIR and CARE use. Carrying out the work of preparing raw data for archiving is beyond the scope of the project, but LDaCA can offer advice on tools and workflows for data processing.