

# HASS Community Data Lab (HCDL)- Phase 1 Project Plan

## Revision history

Version	Date	Editor	Summary of changes
0.1	31/10/2022	Shubhra Dargar	Initial Project Plan setup
0.2	16/01/2023	Shubhra Dargar	Revised Project plan setup

## 1. Project details

<b>Project title</b>	HASS Community Data Lab Project
Contracting Organisation	Internal
<b>Project lead</b>	
Name	Jenny Fewster
Email	jenny.fewster@ardc.edu.au
<b>Project Manager</b>	
Name	TBD
Phone	TBD
Email	TBD

## 2. Project Aims and Outputs

### 2.2 Project Overview

With this project delivered, the Community Data Lab could assist with the generation of focused platform(s) tools and collection subsets for open groups of researchers. To provide an example of how the 'collaboration layer' might actually work in practice, researchers with a shared interest or focus could curate/create/collect a set of tools and datasets within the larger platform and embed them within their own set of discussions, documentation and annotations.

These tools or services could have their own portal but still be part of the broader HCDL platform. Linked access to a more limited pool of resources might enhance search functions and skills for a broad cross-section of researchers, while also developing depth in some topics or approaches or tools. One means of exploring the extent of the collections in the Australian context, while also producing significant new research, is to help generate several mini-platforms that group certain types of collections and tools for investigation in multiple ways by researchers from many disciplines.

This solution could incorporate sophisticated search tools, language interrogation, visualisation, mapping and place names and geolocation, and images. It could provide linked data for visualisation via knowledge graphs and network analysis that would provide new insights into Australian society. It would need to be open to others complementing data, while not compromising what has already been completed. It would house datasets which could be created, developed, returned to, and used by others, whether to test, interrogate differently, or to add to with new data. It could have, in conjunction, a form of exhibition or teaching component to it as well. It would be a series of sets and subsets of data that still could include integration with resources beyond itself. The mini-platforms could themselves be linked and could spark new examples.

### 2.2 Project Aim

The aim of the project is to build a HASS Community Data Lab (HCDL) that, over time, can be phased and extended based on the needs of the HASS research community. This Community Data Lab will

enable the sharing of an increasing range of tools and datasets, provide environments for running the tools, and options for researchers to analyse and annotate datasets. It is a space that is not just a number of instruments and infrastructures, but also governance and procedures that have protocols, and practices for managing and recording research as a process.

The project will be delivered in two phases over the span of 1.5 years (2023-2024) however this project plan only details information in regard to Phase 1. Phase 2 will create a separate project plan.

Phase 1 will focus on Trove as the central data source and provide documentation and examples of how to process Trove data and other data using existing research infrastructure available to Australian researchers, including work packages such as: Nectar computing services, publicly available data repositories such as Zenodo and Figshare and workspace services such as github and gitlab. The defined work packages can be found in the Section 6 (Milestone and Deliverables) and will be prioritised dynamically during this project to maximise the impact and interoperability of the growing HCDL.

### 3. Project Team Roles and Responsibilities

The following table defines the roles and responsibilities of key stakeholders and staff throughout the implementation of the project.

Project role	Organisation/ Individuals	Responsibility
HASS Director	Jenny Fewster	Overall program oversight including leading stakeholder discussions
Solution Architect	Peter Sefton, consultant	Overall solution oversight and guidance on various work packages
Project Manager	Owen O'Neill, (Proposed contractor)	Project management, oversight and direction
<b>PHASE 1</b>		

<b>(1) WP NSI- Nectar Service Integration</b>		
Technical Lead	ARDC	Technical lead and overseeing technical design of the Nectar Service Integration
DevOps lead	ARDC	Lead dev ops for developing the Binderhub Services
<b>(2) WP TDG- Trove Data Guide &amp; Tools Catalog</b>		
Technical Lead	Tim Sherratt, (Proposed contractor)	Technical lead and overseeing technical design of the Trove Data Guide & Tools Catalog
<b>(3&amp;4) WP SIA &amp; IAW- Stylometrics tool (Intelligent Archive) &amp; Image Annotation Workbench</b>		
Research Consultant	Systemik Solutions, (Proposed external organisation)	Management, oversight and direction of solution
Technical Analyst	Systemik Solutions, (Proposed external organisation)	Lead analyst to develop and establish the Stylometrics tool (Intelligent Archive) & Image Annotation Workbench.
Technical Developer	Systemik Solutions, (Proposed external organisation)	Lead developer to develop and establish the Stylometrics tool (Intelligent Archive) & Image Annotation Workbench.
Content Designer	Systemik Solutions, (Proposed external organisation)	Design and develop workflows and procedures
<b>(5) WP DAT- Data transformation and movement – supporting the integration of tools and data</b>		
Technical Lead	Conal Tuohy, (Proposed contractor)	Technical lead and overseeing technical design of the Data transformation and movement – supporting the integration of tools and data
<b>(6) WP ENG- Engagement and Skills Development</b>		

Skills and engagement Manager	Owen O’Neil, (Proposed contractor)	Leading engagement and skills development for the project
-------------------------------	---------------------------------------	---

\* Contracting will follow ARDC’s procurement policy

## 4. Governance and Project Management

For the purpose of this project, a Steering Committee will be established to provide governance for assessing the project’s performance. The Steering Committee will consist of relevant ARDC representatives, Key Project members, selected HASS research researchers/representatives and technical community members. The Terms of Reference for the Steering Committee will be attached to this project once the Committee is formed.

As a broad group of individuals will be working together on this project to deliver the various work packages, a defined project governance framework (such as ARDC Project Governance Framework) will be used by the Project Manager for this team in order to ensure timely decisions are taken for the project, timelines are managed and effective communication is emphasised throughout the project.

## 5. Milestones and Deliverables

The below table represents the high level milestones to be delivered for Phase 1 of this project.

Work packages/Milestone/Deliverables	Responsibility (org) / Contractor	Due date
<b>Milestone: Pre-Planning</b>		
Architecture Document drafted and socialised with Project Group	Peter Sefton	16 Jan 2023
Architecture Document updated with feedback	Peter Sefton	25 Jan 2023
Project Plan drafted	ARDC	1 Feb 2023
Project Plan reviewed by the PMO	ARDC	9 Feb 2023
Project Plan approved by the CEO	ARDC	17 Feb 2023

Procurement exceptions drafted and submitted for approval to the COO	ARDC	22 Feb 2023
Contracts drafted for external project team members	ARDC	28 Feb 2023
Contracts approved and signed by the parties	ARDC	7 March 2023
<b>Milestone: PHASE 1</b>		
<b>WP- Nectar Service Integration (NSI)</b>	ARDC	
Develop a HCDL Binderhub service in Nectar	ARDC	27 June 2023
Provide Nectar resources to Systemik for hosting Intelligent Archive (IA) and Image Annotation Workbench for work developed in WP-SIA and WP-IAW	ARDC	27 June 2023
Provide Nectar resources (Nectar Allocation) hosting the following Python apps to support GLAM Workbench: <ul style="list-style-type: none"> <li>• <a href="#">Trove API Console</a></li> <li>• <a href="#">PDF proxy for Zotero</a></li> <li>• <a href="#">Trove newspaper places</a></li> </ul>	ARDC	27 June 2023
Develop a set of <b>recommendation for Phase 2</b> regarding long-term service requirements particularly: <ul style="list-style-type: none"> <li>• Virtual desktop deployment of Intelligent Archive (IA) and other HASS tools building on Neurodesk (VDI)</li> <li>• Binderhub style Server on demand (using IA) with preloaded data</li> <li>• Test installation of ATAP-style data archive and portal/API with harvested tool metadata to provide a <a href="#">directory</a></li> <li>• Determine effort and approach to support proxies and caches in front of Trove</li> <li>• Support CILogon logon, building on current integration activity (see this eResearch <a href="#">presentation from LDaCA</a>)</li> </ul>	ARDC	27 June 2023
<b>WP-Trove Data Guide &amp; Tools catalogue (TDG)</b>		

<p>Developing the Trove Guide (Jupyter notebook) will (More <a href="#">info</a>):</p> <ul style="list-style-type: none"> <li>● Provide an entry point for researchers</li> <li>● Create detailed documentation of Trove data to support critical understanding of what is, and isn't, in Trove and the sorts of research questions it supports</li> <li>● Collection of pathways to show how Trove data can be analysed and visualised using existing tools</li> <li>● Links to other tools and services being developed as part of the project</li> </ul>	Tim Sherratt (Proposed Contractor)	June 2023
Enhance metadata on existing Jupyter notebooks and tools	Tim Sherratt (Proposed Contractor)	June 2023
Integrate RO-Crate packaging tool for Trove API data	Tim Sherratt (Proposed Contractor)	June 2023
<b>WP-Stylometrics tool (Intelligent Archive) (SIA)</b>		
Establish Intelligent Archive (IA) API	Systemik Solutions (Proposed external organisation)	June 2023
Develop the IA platform and IA site on ARDC server	Systemik Solutions (Proposed external organisation)	June 2023
Establish the Pilot Trove and ATAP (Australian Text Analytics Platform) collection(s)	Systemik Solutions (Proposed external organisation)	June 2023
Export of analysis results as RO-Crate to Zenodo	Systemik Solutions (Proposed external organisation)	June 2023
Conduct skill transfer workshops	Systemik Solutions (Proposed external organisation)	June 2023
<b>WP- Image Annotation Workbench (IAW)</b>		

Develop the Image annotation Workbench (IAW) platform and IAW site on ARDC server	Systemik Solutions (Proposed external organisation)	June 2023
Implementation of persistent ID's for IAW	Systemik Solutions (Proposed external organisation)	June 2023
Establish International Image Interoperability Framework (IIIF) Servers	Systemik Solutions (Proposed external organisation)	June 2023
Implementation of persistent ID's for IIIF images	Systemik Solutions (Proposed external organisation)	June 2023
Establish the Pilot Trove collection(s)	Systemik Solutions (Proposed external organisation)	June 2023
Conduct skill transfer workshops	Systemik Solutions (Proposed external organisation)	June 2023
<b>WP- Data transformation and movement – supporting the integration of tools and data (DAT)</b>		
Consider best implementation path for the Trove Interfaces(as a proxy, code library etc) : <ul style="list-style-type: none"> <li>● IIIF Proxy for NLA and harvested sites PILOT (dependency for WP-IAW)</li> <li>● In browser Trove API client</li> <li>● Other formats if there is demand (TEI, CSV, RDF, OAI-PMH, ATOM etc)</li> </ul>	Conal Tuohy (Proposed Contractor)	June 2023
Conduct data movement (pilot / test and report): Torrents / test ATAP Authorised API	Conal Tuohy (Proposed Contractor)	June 2023
Explore options for caching proxies (to possibly reduce load on NLA servers and provide a foundation for a versioned API with a “research grade” versioned cache of NLA and other open data)	Conal Tuohy (Proposed Contractor)	June 2023



<b>WP- Integration Activities (INT)</b>		
Evaluate priorities to integration scenarios that demonstrate value and validate the architecture. [NOTE- This work package can be augmented as opportunities present themselves]	ALL	June 2023
<b>WP- Engagement and Skills Development (ENG)</b>		
Provide grants to researchers for running workshops - eg Hugh Craig to promote WP-SIA	Owen O’Neil (Proposed Contractor)	June 2023
Conduct Engagement opportunities and skills workshops. E.g.: a. <a href="#">Australian Historical Association conference</a> (July 2023) b. <a href="#">A Zotero translator bounty</a> for GLAM collections would help a lot of HASS researchers	Owen O’Neil (Proposed Contractor)	June 2023
Provide user support	Owen O’Neil (Proposed Contractor)	June 2023

A detailed HASS Community Data Lab Architecture document explaining the overall systems architecture can be found linked [here](#).

## 6. Assumptions

List any assumptions that are being made in order to deliver successful project outcomes.

Item #	Category (Scope/cost/quality)	Description
1	Scope	That the National Library will make the APIs accessible to the ARDC.
2	Quality	That the National Library will make agreed improvements to the APIs in a guaranteed timeframe. <i>Note- NLA API improvements are not a part of this project</i>

3	Quality	That the National Library will continue in general to support API access to Trove.
4	Quality	The Trove Enhancements project includes some API validators or a testing framework to provide confidence that the APIs are stable and well-defined and enable tool builders to have confidence that the API isn't (incompatibly) changing underneath them.
5	Scope	The effort estimated for each work package proposed has a June 2023 delivery timeline as per phase 1 requirement.

## 7.(Inter)dependencies

Dependency	Relationship to / impact on project	How and who will manage the dependency
Contributing platforms - ATAP etc	Knowledge sharing and providing the information to replicate or enhance the solutions?	HCDL Project Manager will be able to manage this dependency by working closely with the ATAP representative to ensure knowledge sharing does happen and is stored in a form of a document for project reference.
Nectar services - cloud and analytics	Availability of the Nectar Core services team to carry out the delivery of a minimum viable service of Binderhub for Phase 1 and investigate possible solutions for potential Phase 2 scope listed in Section 2 'Project aims & Outcomes'	HCDL Project Manager to work with Nectar Services team representatives and incorporate the team's availability to provide the confirmed deliverables and requirements through agreed deliverables.
Proxy API development	Image Annotation Workbench- The success of the project in delivering outcomes to schedule is dependent upon Proxy API development.	Ensuring the Proxy API development is captured as a dependency in the implementation plan and working closely with Systemik solutions program manager to manage this dependency.

## 8. Risks

### Risk Rating Key

		Consequence				
		Insignificant (1)	Minor (2)	Moderate (3)	Major (4)	Significant (5)
Likelihood	Almost certain (5)	6	7	8	9	10
	Likely (4)	5	6	7	8	9
	Possible (3)	4	5	6	7	8
	Unlikely (2)	3	4	5	6	7
	Rare (1)	2	3	4	5	6

Risk	Impact type and how will impact project (Scope/cost/quality/schedule)	Controls/Mitigation Strategy	Residual Risk Rating (after controls are in place)	Risk Owner
Contracts execution (based on leads delivering the work packages) time frame: To ensure the project is able to start for Phase 1 of the project.	Schedule	In this case, it will be vital to track the project delivery closely to ensure work packages are evaluated and the scope can be delivered based on the timelines confirmed.	4	Jenny Fewster

<p>That the National Library will make agreed improvements to the APIs in a guaranteed timeframe</p>	<p>Quality</p>	<p>Project will utilise the APIs in its current state until the improvements are made.</p>	<p>5</p>	
--	----------------	--	----------	--

## 9. Outcomes

Outcome	Evaluation question	Indicator/s	Data source/s	Timeline for data collection	Responsibility	Baseline
Broader accessibility to tools and datasets in Trove to more HASS researchers	Does the Community Data Lab enable the sharing of an increasing range of tools and datasets, provide environments for running the tools, and options for researchers to analyse and annotate datasets?	# of users of the TCDL?  # increase in Trove data published in research outputs?	# platform metrics  Feedback form?	June 2024	Jenny Fewster	Currently, there is no one platform that provides all the required tools, datasets, services to HASS researchers.
It will facilitate a larger development of new tools and new ways of conducting	Has the HCDL facilitated better interaction between	Researcher satisfaction with HCDL outputs	Feedback form/survey	October 2023	Jenny Fewster	There is no model developed/running that supports researchers to

Outcome	Evaluation question	Indicator/s	Data source/s	Timeline for data collection	Responsibility	Baseline
<p>large-scale research projects with Trove and other institutions collecting data. It offers the capacity for the enhanced use of Trove data by researchers, a better development of user communities and user co-creation, and a better interaction between researchers and Trove data.</p>	<p>researchers and Trove data?</p>					<p>interact better with the Trove data.</p>
<p>It responds to users' stated need for more information on</p>	<p>Has the delivery of the Engagement and Skills Development work</p>	<p># ppl who attend training / skills workshops</p>	<p>Survey of participants following training.</p>	<p>October 2023</p>	<p>Jenny Fewster Project manager?</p>	<p>There is a gap currently in providing the right tutorials/training to</p>

Outcome	Evaluation question	Indicator/s	Data source/s	Timeline for data collection	Responsibility	Baseline
how to capture the data they seek, a wider range of tools and functionality, and a better comparability of skills among and between users.	package supported researchers to confidently use the tools and datasets within the HCDL?		Feedback from users on training and tutorials that help skills development to use a wide range of tools, data and services for research.			users seeking more information on how to capture the data, tools for and among researchers.

## 10. Communications and engagement

A communications and engagement plan will be developed for this project to cover the communication with the users, the user guides and their release along with an engagement plan for releasing the information about this project externally as well.

### Change Control (for ARDC information only)

Approval of this Project Plan will comprise the baseline for the project. Changes to any of the following are considered variances:

- Project details
- Project outcomes & aims
- Budget
- Project partners
- Project team roles and responsibilities
- Governance
- Milestones and deliverables
- (Inter)dependencies

Variances to the Project Plan require endorsement by the Steering Committee and then ARDC approval. If approved, the Project Plan will be revised and project reports from that point forward will report project progress against the revised Project Plan, not the original.

To request a variance:

1. The Steering Committee submits a request to ARDC for variance to the approved project plan.
2. ARDC reviews the changes and advises the project manager or project lead of the outcome.



# Appendices

Appendix A - [HASS Community Data Lab Architecture Document](#)

Appendix B - [Trove Researcher Consultation Report](#)