



Australian Research Data Commons

A thematic approach for research data infrastructure: ARDC, HASS and Indigenous RDC, and ADA

May, 2024

Andrew Treloar, Director International Strategy

Jenny Fewster, Director HASS and Indigenous RDC

Ryan Perry, Australian Data Archive



ARDC is enabled by NCRIS



Australian Government



Australian Centre for
Disease Preparedness



Southern Coastal
Research Vessel Fleet



Supporting
NCRIS Capability



Our Audiences



Researchers

Researchers from
any discipline
(academic, but also
government and industry)



People Who Work With Researchers

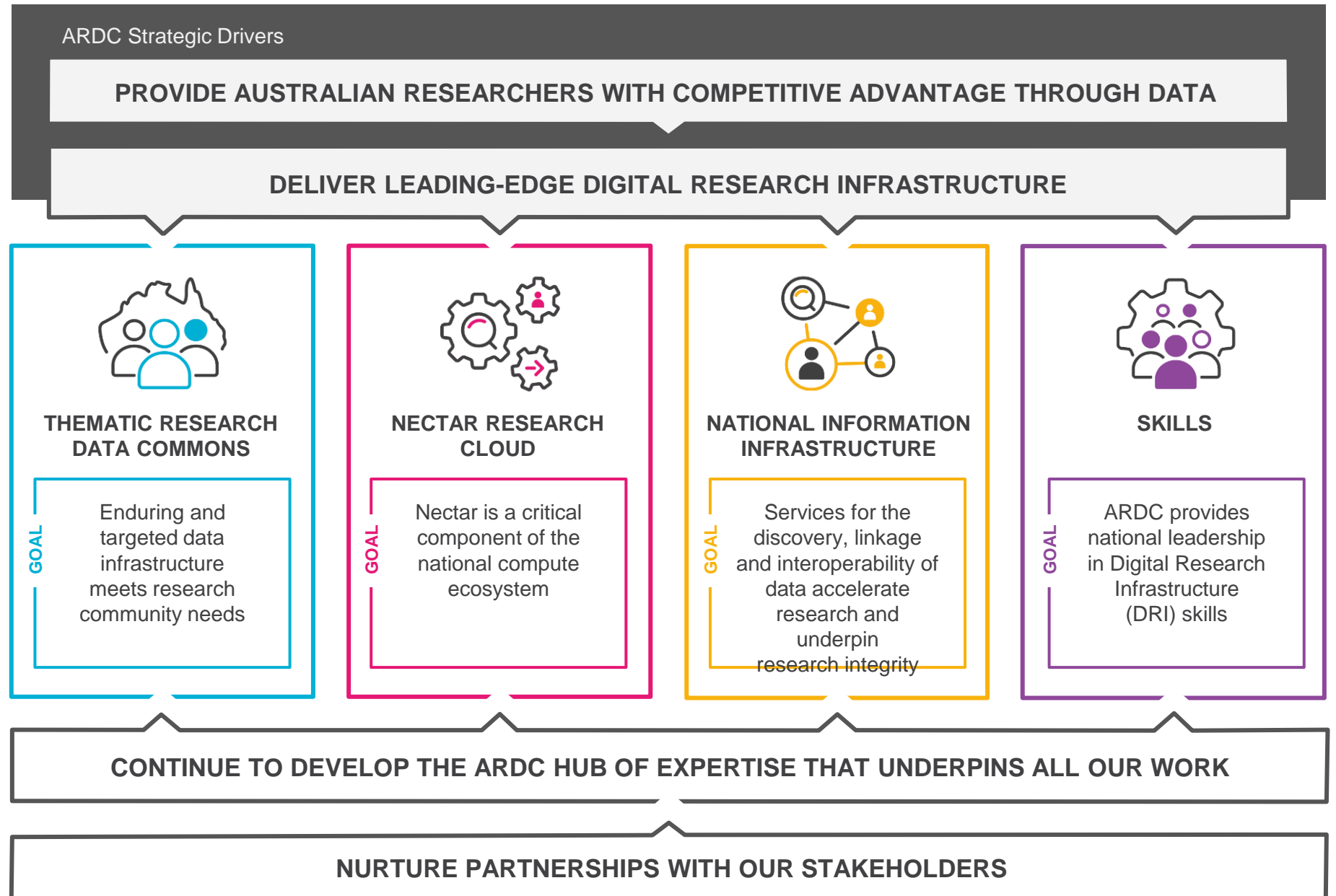
Infrastructure providers
Universities / Research Institutions
Data Managers
Trainers
Librarians
eResearch / Research
Support Offices



Key Stakeholders

Australian Government (NCRIS)
Research Funders
Government / Policy
International Data Organisations
National Data Organisations
ARDC Members

Strategy on a Page



Thematic Research Data Commons

- “Enduring and targeted data infrastructure that meets research community needs”
- Three thematic areas
 - Planet (earth and environment)
 - People (health and medical)
 - HASS (humanities, arts and social sciences)&Indigenous

What is a Research Data Commons?

A research data commons brings together people, skills, data, and related resources such as storage, compute, software, and models to enable researchers to conduct world class data-intensive research.



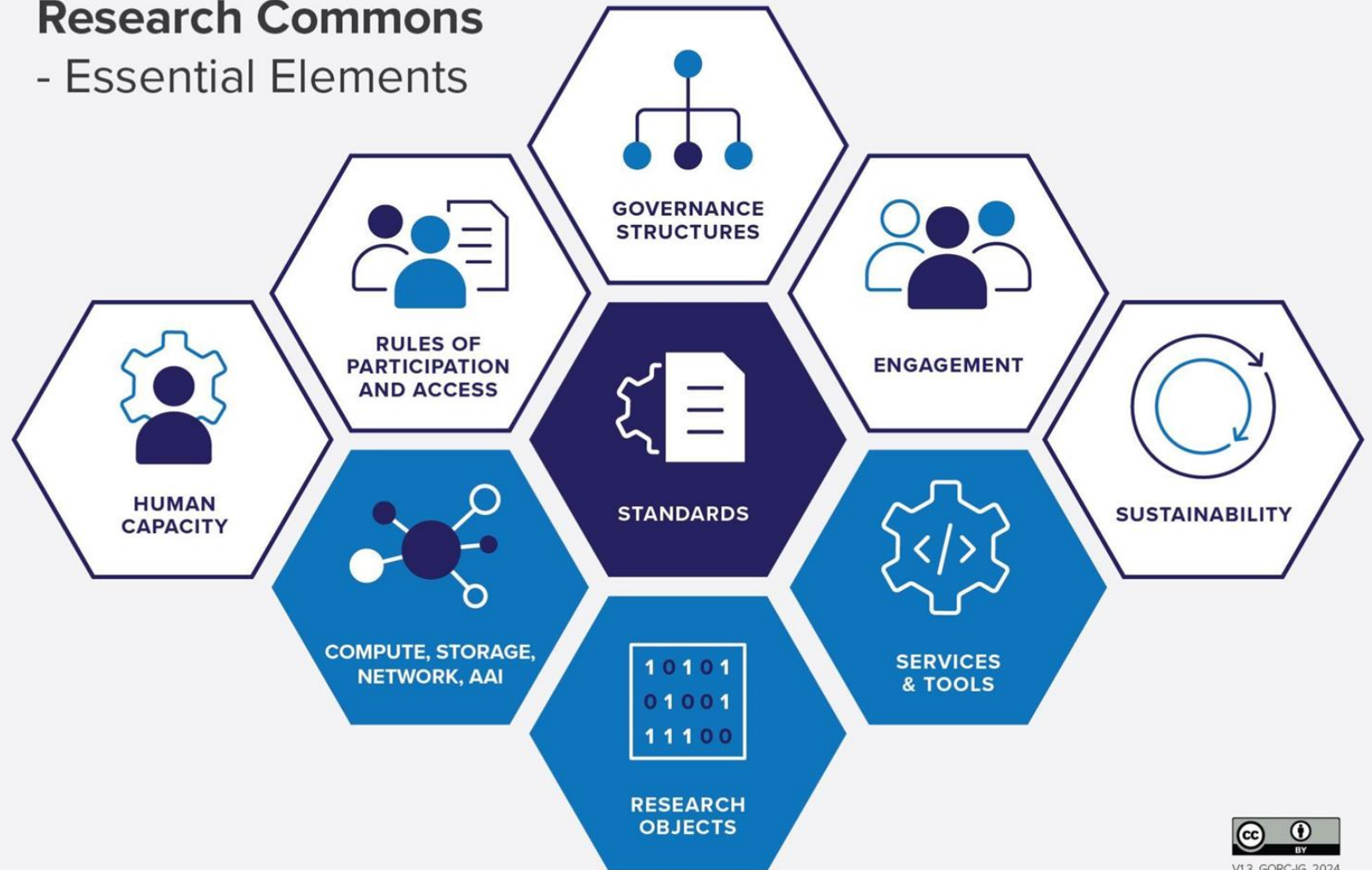
Global Open Research Commons

- Research Data Alliance Interest Group
- Definitions and diagrams [here](#)

Can be used:

- Retrospectively (SURF)
- Prospectively (REASON, BioFair)
- Analytically (interop between commons)

Global Open Research Commons - Essential Elements



World Leading Digital Infrastructure Enables Australian Researchers to Transform Our Lives



PEOPLE

Research Data Commons



PLANET

Research Data Commons



HASS & INDIGENOUS

Research Data Commons



Australian Research Data Commons

PEOPLE

Research Data Commons

National-scale data infrastructure for health research and research translation.

2023 – 2028
≈ AUD 30M
= SEK 206M
= EUR 18M



Find Resources for Researchers

Explore datasets, free tools and upskilling materials we've curated for data-intensive research in health research and research translation

Get Involved

Learn more about and register your interest in the People Research Data Commons.

<https://ardc.edu.au/people-research-data-commons/>



Australian Research Data Commons

PLANET

Research Data Commons

National-scale data infrastructure for environmental research and decision making

5 yr program, but only first 3 years are funded at present



Similar funding to other RDCs is what we are seeking



Find Resources for Researchers

Explore datasets, free tools and upskilling materials we've curated for data-intensive research in earth and environmental science

Get Involved

Learn more about and register your interest in the Planet Research Data Commons.

<https://ardc.edu.au/planet-research-data-commons/>



Australian Research Data Commons

HASS & INDIGENOUS Research Data Commons

National-scale data infrastructure for HASS
and Indigenous Research Data Commons.

2023 – 2028
≈ AUD 35M
= SEK 215M
= EUR 20M



Find Resources for Researchers

Explore datasets, free tools and
upskilling materials we've curated
for data-intensive HASS and
Indigenous research

Get Involved

Learn more about and register
your interest in the HASS and
Indigenous Research Data Commons.

<https://ardc.edu.au/hass-and-indigenous-research-data-commons/>

HUMANITIES, ARTS, SOCIAL SCIENCES (HASS) AND INDIGENOUS RESEARCH DATA COMMONS



**Improving Indigenous Research
Capabilities**



**Trove Enhancements /
Community Data Lab**



**Integrated Research
Infrastructure for Social Sciences**

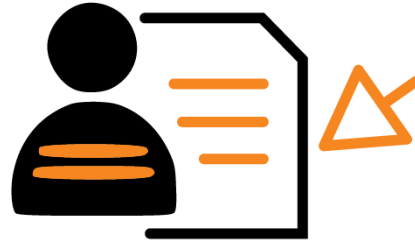


**Language Data Commons of
Australia**

HUMANITIES, ARTS, SOCIAL SCIENCES (HASS) AND INDIGENOUS RESEARCH DATA COMMONS



**Improving Indigenous
Research Capabilities**



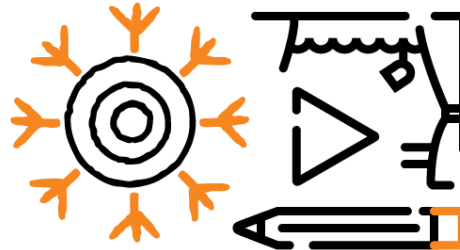
Connections



**Language Data
Commons of Australia**



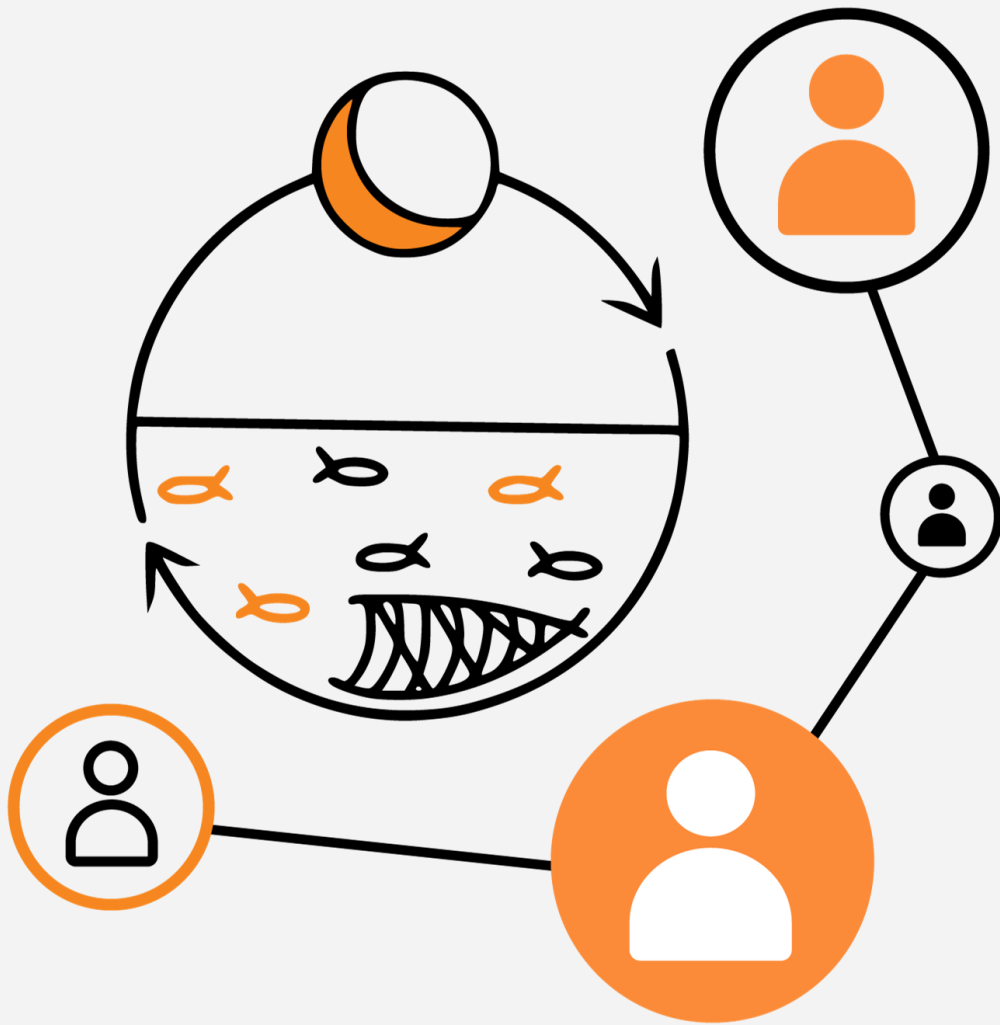
**Social Science Research
Infrastructure Network**



**Australian Creative
Histories and Futures**



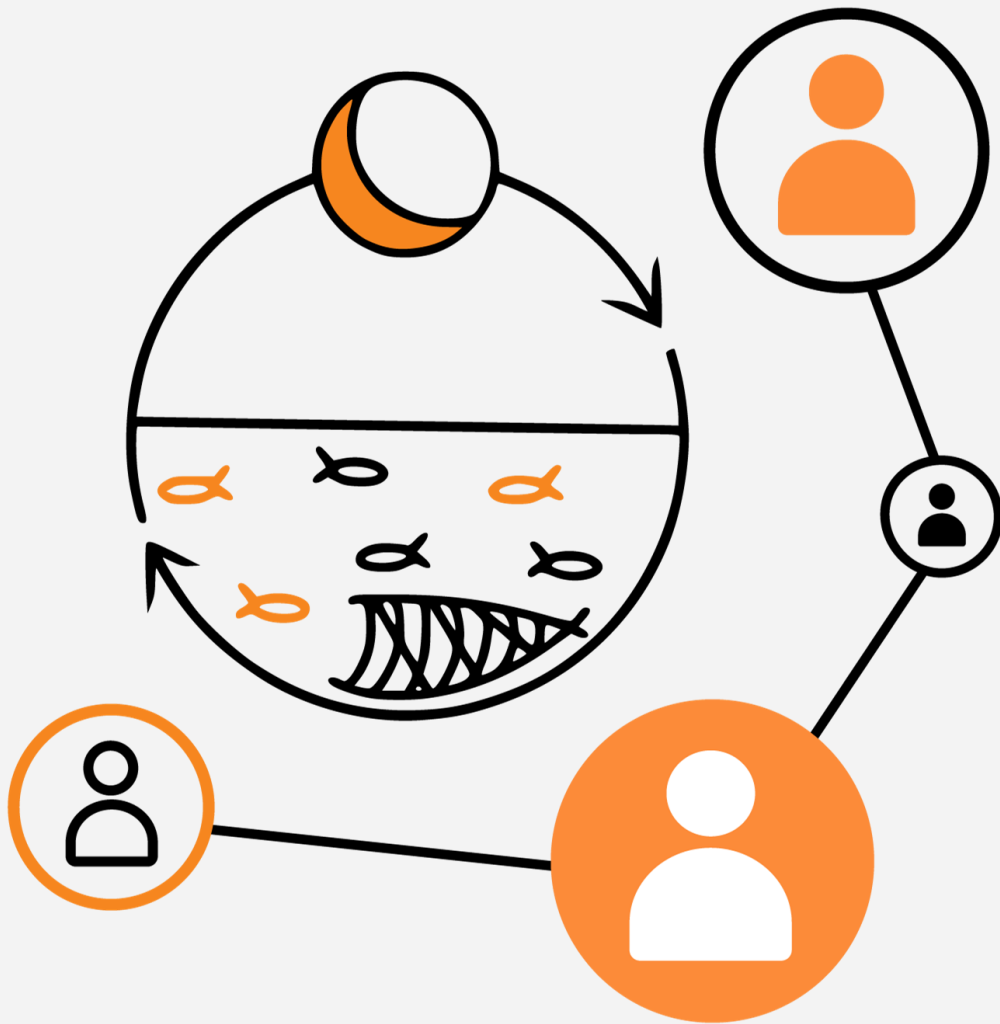
**Australian Internet
Observatory**



Enhancing Metadata for Inclusive Research on Entrenched Disadvantage: pilot project

Lead: Prof. Wojtek Tomaszewski
University of Queensland

- Review of best practice in metadata documentation
- Review of Higher Education metadata in PLIDA and its provenance
- User experience consultations
- Solution design and forward plan
- Developing a Monitoring and Evaluation plan to capture long-term outcomes



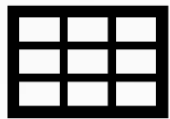
Social Science Research Infrastructure Network

Lead: Prof. Wojtek Tomaszewski
University of Queensland

- Establishment activities, partner & community engagement, evaluation & innovation
- Improving integrated data usability
- Health and social data integration
- Public social science data: Improving discoverability and building on existing data assets
- Guidelines for users of government administrative data on Indigenous people
- Training and capacity building

The Australian Data Archive

National data repository for the social sciences



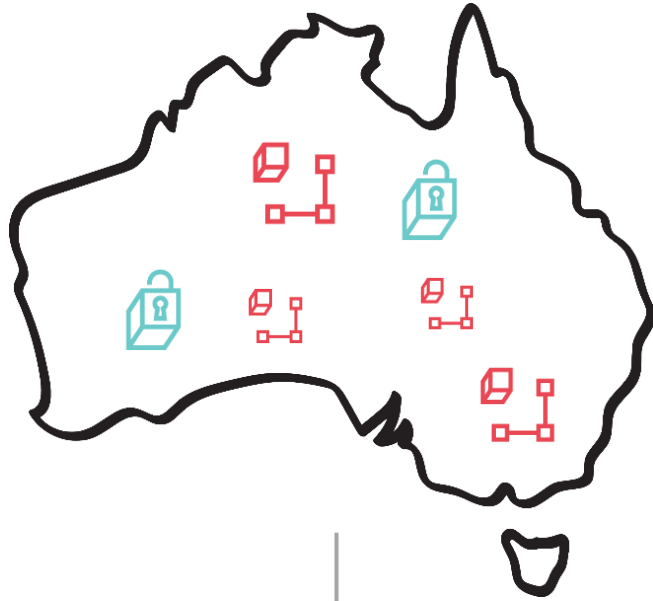
1,700 datasets



4,000+ access requests p/a

- ❖ Est. 1981: Australia's national social-science data repository, housed at *The Australian National University*
- ❖ Hosts major Australian Government cohort surveys
- ❖ 3 data archivists • user-support officer • dedicated dev team.
- ❖ Operates on the OAIS model; services delivered via Harvard Dataverse on the National Computational Infrastructure for secure, scalable storage.
- ❖ CoreTrustSeal certification - pending review (submitted 2025).
- ❖ Active member of standards communities (CODATA, WDS, DDI Alliance). Aligned with international interoperability efforts.





Data integration



Data analysis



Data Platforms

Problem

Australian social science data infrastructure is fragmented with inconsistent standards and vocabularies.

Solution

Establish a foundation for integrating data, analysis, and platforms.

The **Integrated Research Infrastructure for the Social Sciences (IRISS)** project built a coordinated set of tools to support interoperability:

- Metadata hosting and integration with
 - Vocabularies (via [ARDC Research Vocabs Australia](#))
 - Questionnaire generation (via LimeSurvey)
- GeoSocial (service integrating spatial aggregate data with individual unit record data)
- Survey data curation tools...

ADA Automation Suite

Curation and Risk Assessment Tool (CARAT)

R-based data management tool

- Self-assess data on a series of quality and privacy criteria

Functions include:

- Data cleaning
- Flagging direct and indirect identifiers
- Recording metadata edits
- Exporting dissemination files

Archival Data Preparation Tool (ADaPT)

A web-based ingest utility that

- packages datasets (files + metadata)
- moves them between Dataverse instances and ADA preservation storage
- Applies OAIS-compliant checks and workflow reporting.

Project Consonance

Prototype AI data harmonisation tool for time-series data.

- Detects and compares variable labels, categories, and code-schemes across longitudinal waves
- Recommends re-labelling or recoding to create clean, time-series-ready datasets.

Capability is being extended to embedding-based search enabling variable-level, cross-study discovery within the entire ADA collection.

CADRE: Coordinated Access for Data, Researchers and Environments

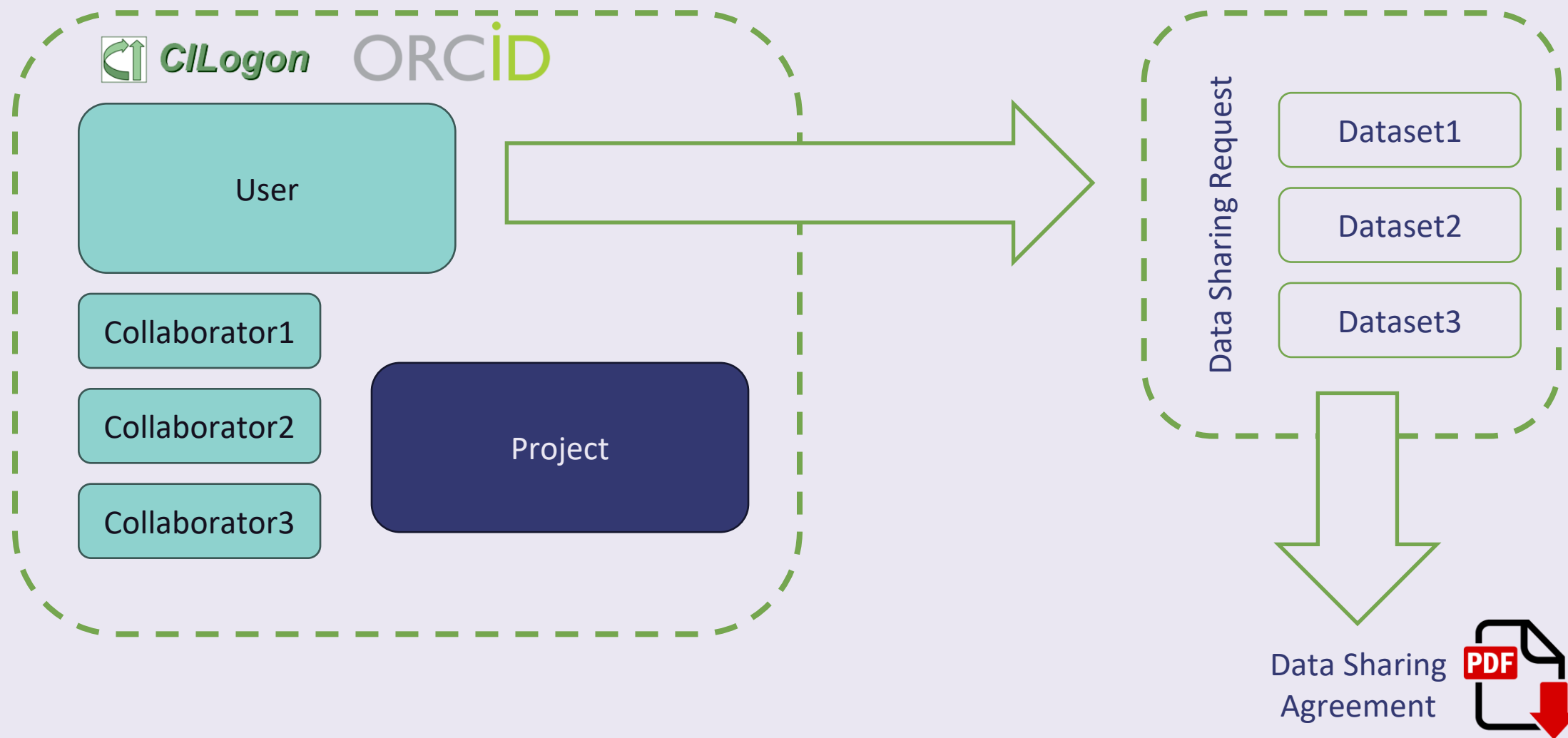


- Built for Australia's *Data Availability & Transparency Act (2022)*, embedding the **Five Safes** in data access provision.
- One platform handles applications and approvals for multiple providers of sensitive social-science data.
 - Pulls researcher credentials, project details, and attachments into a live dashboard to facilitate approvals and record keeping.

CADRE aims to operationalise the Five Safes framework and provides a sensitive data access management platform for the social sciences and related disciplines.

for the social sciences and related disciplines.

Coordinated requests





Where to next?

- Introducing data providers to CADRE
 - Scaling up capacity for data access requests at ADA
- Implementing curation and harmonization workflows
 - For archivists, depositors, and end users
 - Including ADAPT tool for archival/preservation workflows
- Expanding self-deposit and training services
 - CoreTrustSeal consistent documentation
 - Risk assessment and licensing



ryan.perry@ada.edu.au

<http://ada.edu.au>



INTERNATIONAL DATA WEEK

13-16 October 2025
Brisbane, Australia



HOSTED BY



Australian Research Data Commons



Australian Government



National Computational Research Infrastructure for Science

ARDC is enabled by NCRIS



OUR PARTNERS

BUSINESS EVENTS AUSTRALIA



internationaldataweek.org/idw-2025/