

Persistent Identifier-Driven Infrastructure for National Research Discovery

Behind the Scenes of Research Link Australia and Research Data Australia

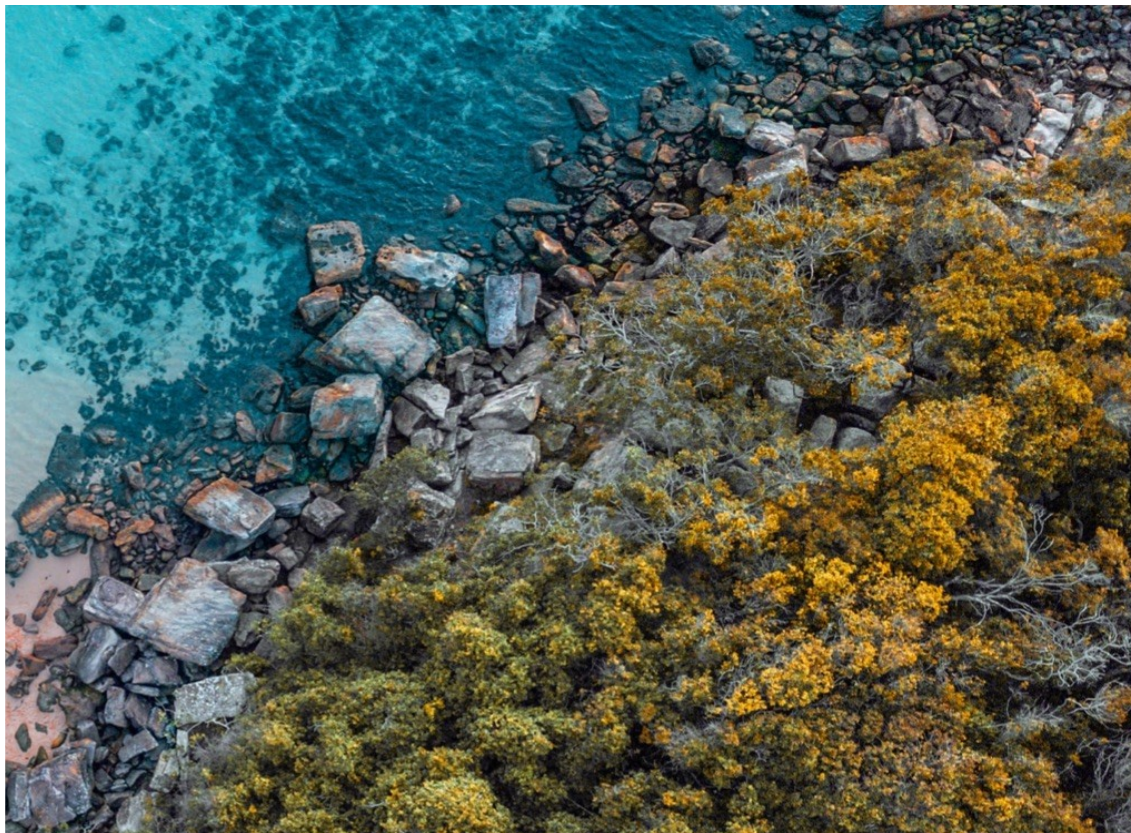
21 Oct 2025

PRESENTED BY

Enming Zhang, ARDC
Shikha Chawla, ARDC

ACKNOWLEDGEMENT OF COUNTRY

We acknowledge and celebrate the First Australians on whose traditional lands we meet, and we pay our respect to their elders past, present and emerging.





Product Overview

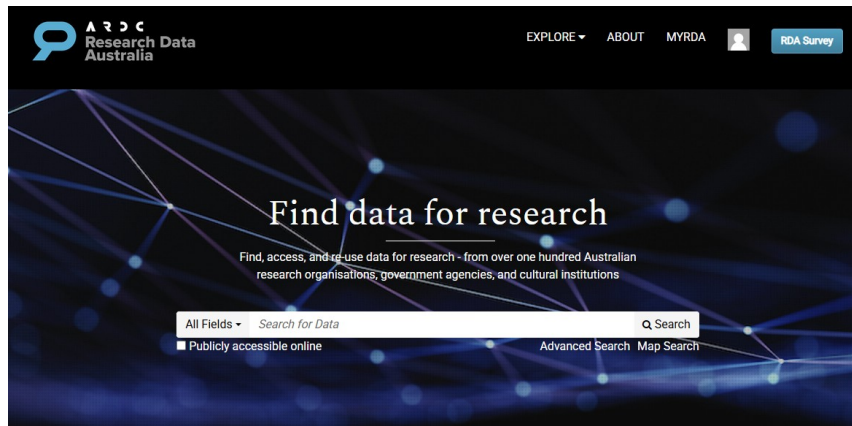
Image — Photo by Surface on Unsplash

What is Research Data Australia?

National Research Data Catalog

An online platform providing metadata and access links to datasets from universities, government agencies, and cultural institutions across the nation.

Over **247,000** data collections searchable across diverse fields including health, environment, business, and humanities.



Browse By Subjects



2024 - 2025

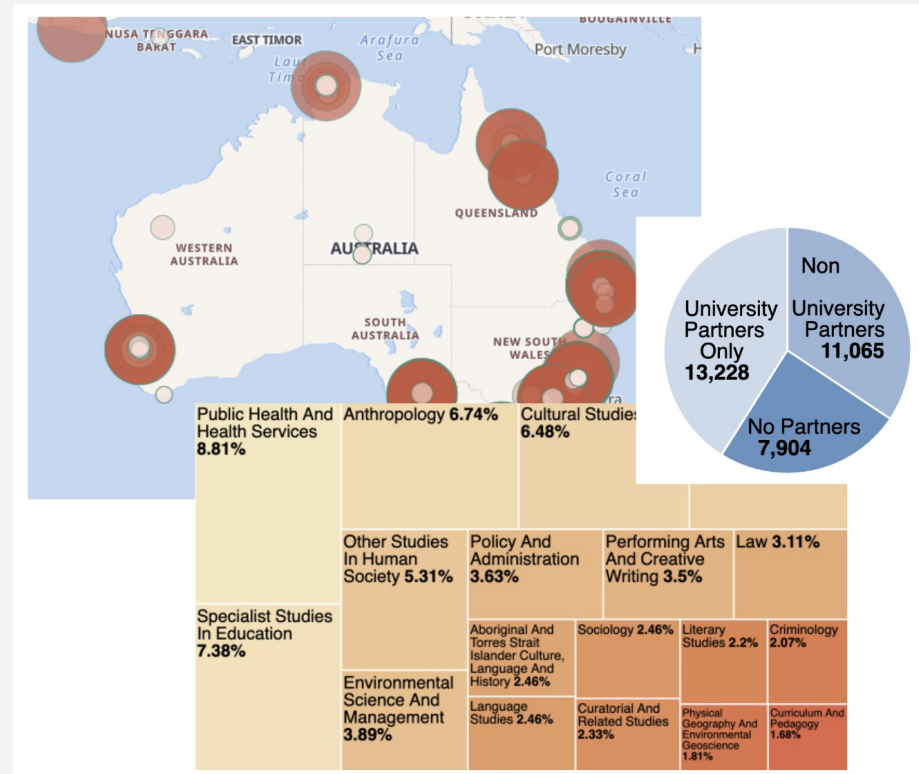
247,965
data collections

869
data services
and tools listed

717,892
page views

Research Link Australia (RLA)

- Platform designed to connect publicly funded researchers with industry collaborators, facilitating the translation of discoveries into products and services.
- Discover information about research collaboration and research capabilities.
- Provide dynamic visualisations of research connections and impact through interactive dashboards.



Records in RLA:



805,500+

Researchers



310,000+

Organisations



88,000+

Funded Activities



1,800,000+

Linked Publications

Collaborations in all research fields

The dashboard displays ARC-funded activities categorised by Collaboration Type. All partners are included, even if some partners departed before the grant period ended. The data source is ARC grants, and is reproduced without change.

A Non-University Partner is defined as any organisation that is not an Australian University

Optional Filters (Case Sensitive)

Start Date Range

2000 2000 2024 2024

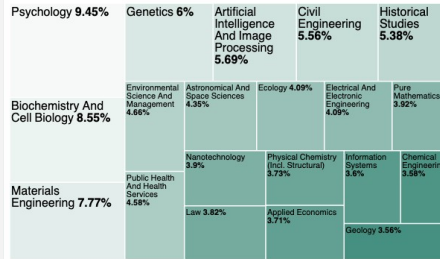
Grants with these Non-University Partners

Select...

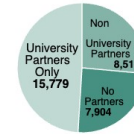
Grants with these Australian University Partners

Select...

Top Primary Fields of Research ranked by # of Grants



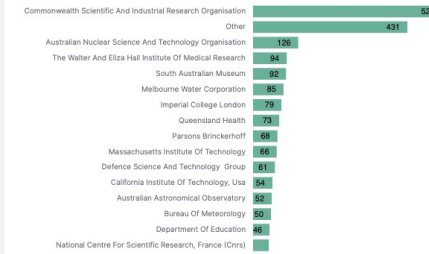
Collaboration Type by # of Grants. Click to filter



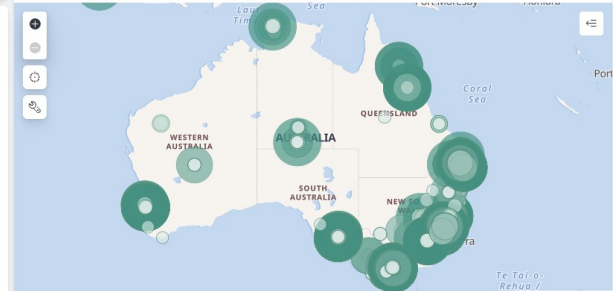
32,197
of Grants

Collaboration Type	Funding
University Partners Only	\$6,828,115,674
Non University Partners	\$5,833,267,283
No Partners	\$3,167,801,797
Sum:	\$15,829,184,754

Non-University Grant Partners ranked by # of Grants



Locations of Organisations ranked by Most Recent Grants (Shows all organisations linked to a grant)



Our Core Mission: Building Connections



Making research outcomes findable through contextual relationships



Showing how funded activities, publications, researchers, and organisations connect



Bring metadata discovery to life using visualisation and interaction

Key Challenges

Implicit Relationships

Many linkages are inferred via shared PIDs rather than explicit relationships. These linkages are crucial but remain hidden from standard discovery methods

Change Detection and Management

Full metadata refreshes strain bandwidth, while complex reindexing, record removal, and on-time synchronisation across systems are difficult to coordinate.

Duplicates

Inconsistent identifiers and manual errors create duplicate records across sources

"Everyone is becoming an aggregator, but no one tracks changes"



Our Approaches



PID-based Graph Database

Link indirectly related entities based on PIDs in a graph database.

Visualise relationship paths and context to enhance FAIRness across records.



DELTA Updates

Ingest only changes via provider-supplied deltas or differential change modeling.

Coordinate incremental reindexing and safe record removals to keep indexes current without full re-harvests.



De-Duplication and Normalisation

Apply entity resolution across different entities to merge true duplicates.

Normalise identifiers, casing, and formats while preserving source provenance.

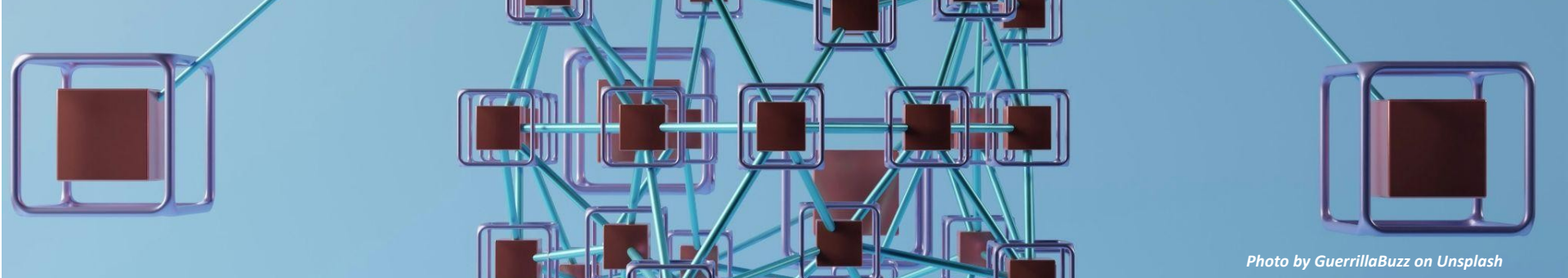


Photo by GuerrillaBuzz on Unsplash

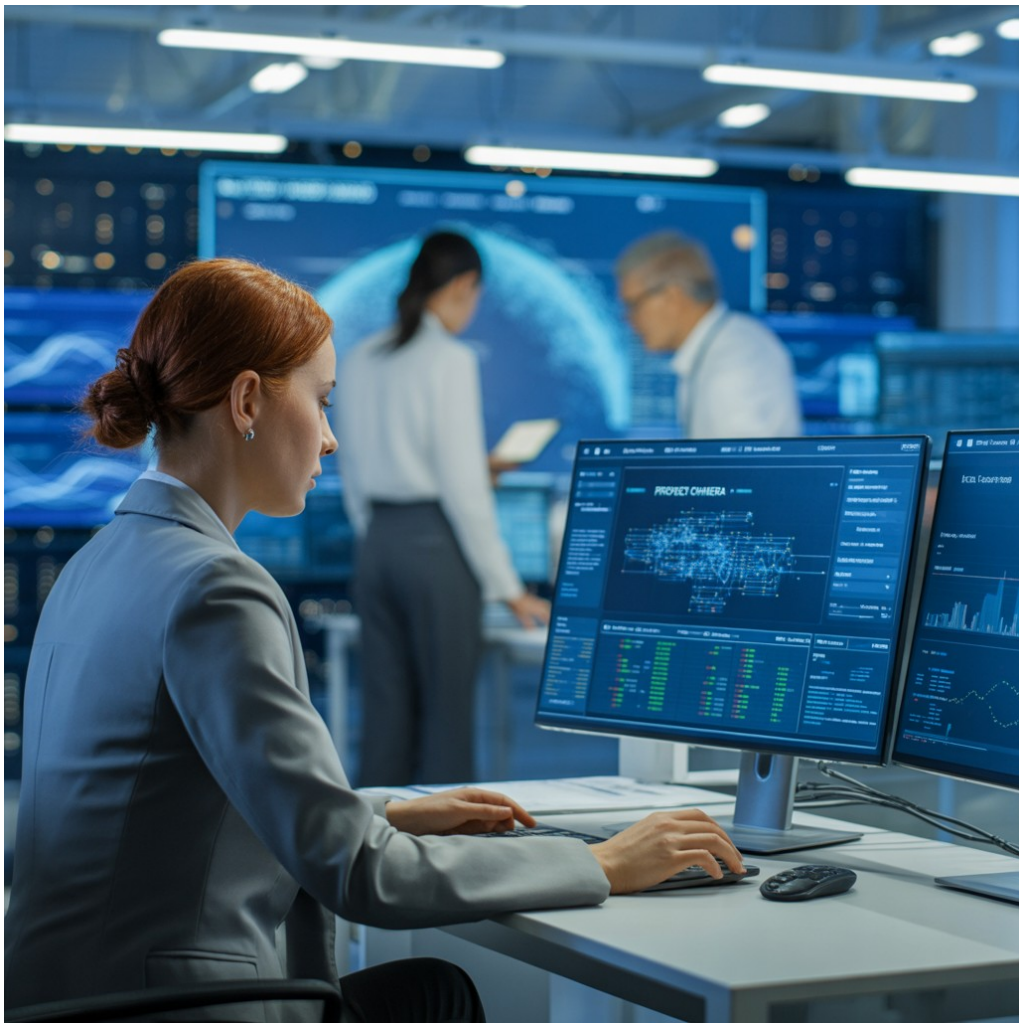
Identifier Normalisation

- Goal: Make entities resolvable and findable across sources via persistent identifiers (PIDs).
- Normalisation rules:
 - Lower casing
 - Trimming
 - Format normalisation per scheme
 - De-duplication: Cluster by normalised IDs to merge true duplicates and keep provenance.
- Outcomes: Higher match rates, fewer duplicates, and consistent linking across the graph.

Records with the following identifiers:

- <https://doi.org/10.1234/12345>
- doi.org/10.1234/12345
- DOI: 10.1234/12345

Will be merged together after normalisation



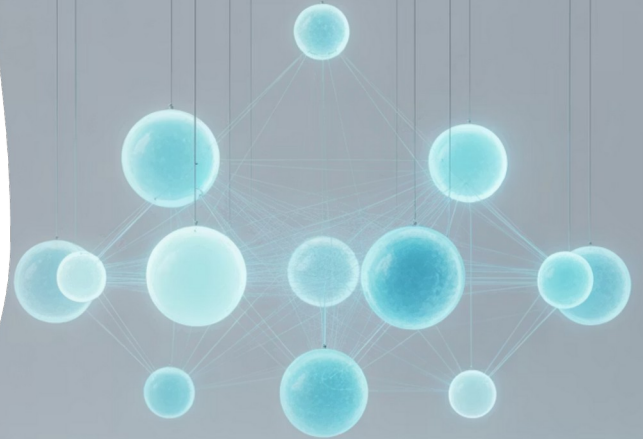
Infrastructures Behind the Scenes

Technical Infrastructure Evolution

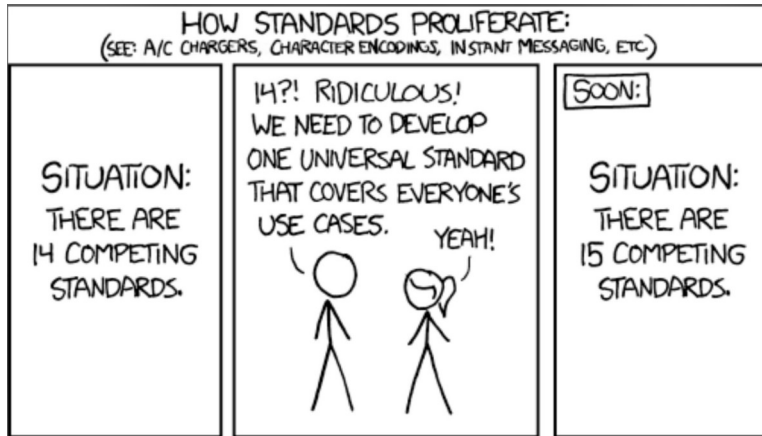
In the digital world, everything can be thought of as an object.

Every entity: publication, researcher, grant, organisation, Patent and Instruments is an object with relationships.

Challenge: Model the **entire Australian research ecosystem** with millions of complex relationships.



The Technical Challenge



Scalable Solution

Software that handles objects, relationships, and metadata efficiently

Data Diversity

Each source describes entities differently—needed flexible system

Standards Problem

15 years with RIF-CS taught us: unwinnable battle

Solution: Support **all standards** that our data providers use.

Our Three-Pillar Approach

01

Data Registry

Store all metadata as source of truth

02

Neo4j PID Graph

Strong model relationships, enable fast traversal

03

ElasticSearch

Fast, scalable architecture supporting nested documents and queries and AI-ready data structures

ElasticSearch contains **everything we know** about any researcher as one complex document.



Technical Infrastructure Evolution

Smart Services & Indexing

The evolution of our technical infrastructure has focused on implementing robust smart services and efficient indexing mechanisms to enhance performance and data retrieval capabilities.

Real-time Reindexing

- Instant data updates
- Zero downtime operations
- Always current search results

Structured Queries

- Complex nested search capabilities
- Precision-based search
- Advanced filters & facets
- Context-aware discovery

Version Support

- Track metadata changes
- Compare record versions
- Ensure data integrity



Lesson Learned: Choose widely supported technology, not just the "best" – community adoption matters more than technical superiority

Elastic Search Index

- Powers RLA searches and dashboards
- High-speed, scalable search engine
- Supports complex and nested queries
- Enables faceted and contextual exploration
- Fuels dynamic, data-driven dashboards

The screenshot displays the Research Link Australia search interface. At the top, there are navigation links for Home, About, and Feedback, along with a purple button for Explore Collaborations. A search bar is prominently featured, with a dropdown menu set to 'Researcher' and a search button. Below the search bar, there are statistics for the current selection: Researchers (809130), Funded Activities (83831), and Organisations (310842). A note explains that the sequence of results does not indicate ranking but reflects grant and publication activity. The search results are displayed in a list format, showing researcher profiles for Kayleigh Scotcher, Adam Cobb, and Honghua Hu. Each profile includes their name, affiliation, ORCID profile link, and the number of publications.

Research Link Australia
Australian Research Data Commons

ARDC Newsletter Subscribe Contact Us

Home About Feedback Explore Collaborations

Researcher Search by researcher's name or research expertise Search

Need help searching? View our Search Guide Advanced Search

Current Selection
Showing all data.

Researchers (809130) Funded Activities (83831) Organisations (310842)

The sequence of this search result does not indicate the ranking of researchers in the searched topic area. However, it does reflect those who have received grants or published manuscripts related to the searched area.

Filter by Field of Research

- Psychology (797)
- Biochemistry and Cell Biology (783)
- Nanotechnology (773)
- Materials Engineering (725)
- Environmental Science and Management (716)
- Public Health and Health Services (649)
- Functional Materials (634)
- Nanomaterials (577)
- Genetics (576)
- Artificial Intelligence and Image Processing (501)
- Sociology (497)

Filter by Socio-Economic Objective

- Expanding Knowledge in the Biological Sciences (2305)
- Expanding Knowledge in Technology (1435)
- Expanding Knowledge in the Physical Sciences (1403)
- Expanding Knowledge in Engineering (1354)
- Expanding Knowledge in the Chemical

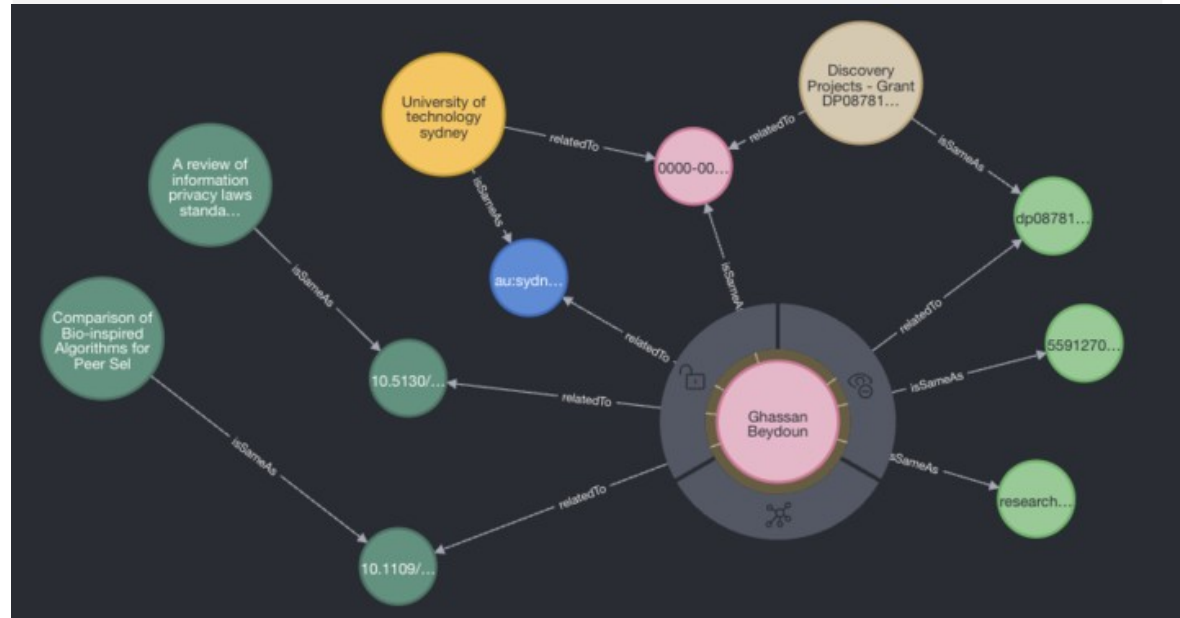
Kayleigh Scotcher Researcher
University of Melbourne
ORCID Profile [0000-0002-3748-6564](#)
Publications 3
More information >

Adam Cobb Researcher
ORCID Profile [0000-0002-3748-4892](#)
More information >

Honghua Hu Researcher
Macquarie University
ORCID Profile [0000-0002-3749-4089](#)
Publications 85

PID graph

- Relationship engine connecting research objects
- Built on persistent identifiers (PIDs)
- Links datasets, publications, researchers, grants, and Organisations
- Supports RLA search enhancement & analytics
- Provides contextual and trustworthy connections





The Role of Persistent Identifiers (PIDs) System Backbone

DOI, ORCID, ROR, ABR, RAiD, etc serve as the foundational linking mechanisms across the entire research ecosystem



Linking Benefits

Seamless connections between datasets, publications, researchers, grants, and institutional affiliations



Current Problems

Duplication
Metadata Standards: sources providing metadata in different data formats
Incomplete/ inaccurate metadata capture

Goal: Widely adopted and internationally recognised identifiers with strict uniqueness enforcement

Schema-Agnostic Registry

Flexible Ingestion

Accepts data from any source without losing information

- Multiple identifiers
- Names and descriptions
- Spatial/temporal data
- Subject codes (ANZSRC, SEO)
- Object relationships

Nested Index Power

Researcher records contain:

- All publications
- All grants
- All affiliations

Everything relevant in one place





Connecting the

1

Raw Metadata

Multiple sources ingested

2

Schema-Agnostic Registry

Flexible storage solution

3

Neo4j Relationships

Graph database mapping

4

ElasticSearch Discovery

Searchable, discoverable data

Research Link Australia is about connecting the dots in research data, so knowledge can accelerate faster than ever before.

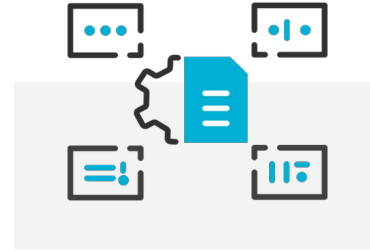
Impact and Outcomes



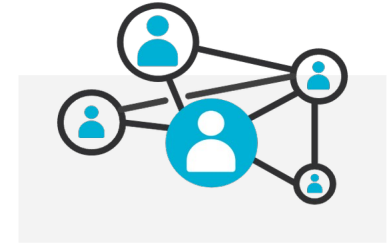
**Dynamic PID
resolution and
relationship building**



**Continuous metadata
enrichment from
multiple sources**



**Robust performance
with low latency in
searching.**



**Future-proof solution with the
potential to scale up, evolve,
and adapt to more robust tech
stack.**

Where to From Here?



Future Directions & Collaboration

1

AI-Assisted Normalisation

Advanced algorithms for handling typos, spelling variations, classify and catalogue metadata with incomplete and inconsistent data

2

Enhanced PID Governance

Improved adoption strategies and standardisation across research institutions

3

Real-time Supernode Management

Dynamic solutions for handling large-scale data clusters efficiently

4

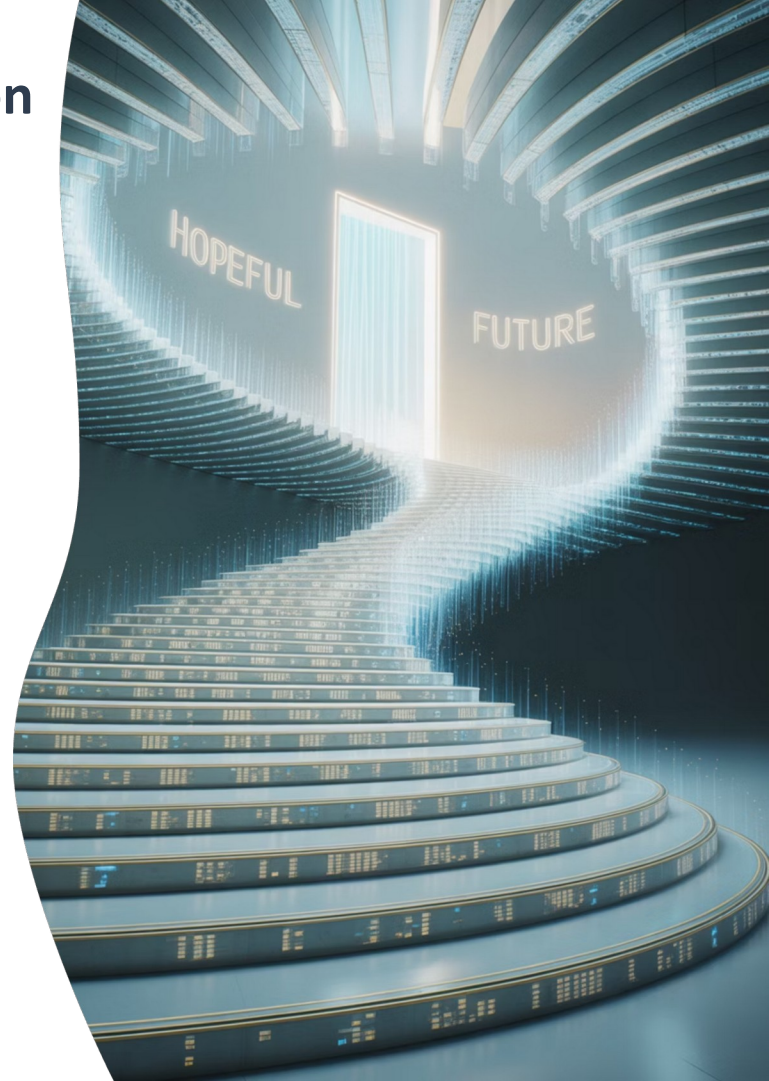
Community Standards

Collaborative engagement on data quality and interoperability standards

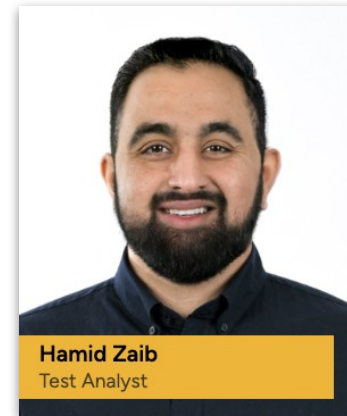
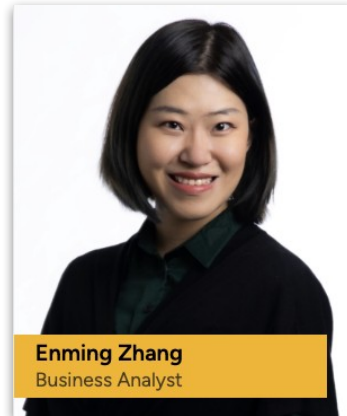
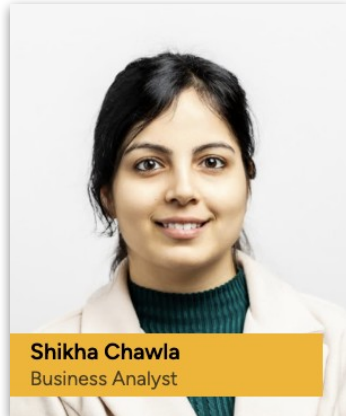
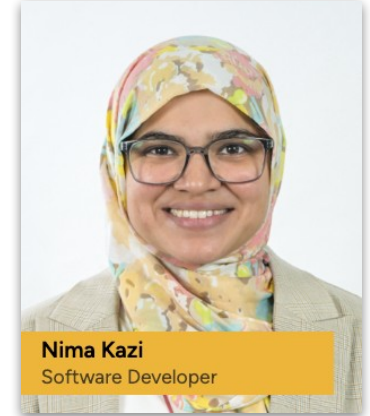
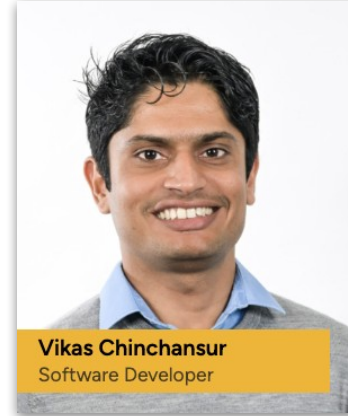
The Path Forward

RDA and RLA continue evolving through persistent identifiers, delta updates, and sophisticated normalisation techniques. Success depends on sustained collaboration and widespread adoption across Australia's research community.

Collaboration and adoption remain the key drivers for transforming research data discovery in Australia







Our Team





Subscribe to the
ARDC CONNECT
newsletter

THANK YOU

-  ardc.edu.au
-  contact@ardc.edu.au
-  [@ARDC_AU](https://twitter.com/ARDC_AU)
-  [Australian-Research-Data-Commons](https://www.linkedin.com/company/australian-research-data-commons)
-  [ARDC_AU](https://www.youtube.com/ARDC_AU)