Research Activity Identifier (RAiD)

The Persistent Identifier (PID) for Research Projects

Chorus Forum

06 June 2023

PRESENTED BY

Shawn Ross
Natasha Simons
Siobhann McCafferty
Matthias Liffers

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

ACCESS
Australian Earth System Simulator
National Research Infrastructure

ANFF

ANSTO

APPF
Australian Plant Phenomics Facility

ARDC
Australian Research Data Commons

Astronomy Australia Ltd.

Atlas of Living Australia
ala.org.au

AURIN
aurin.org.au

AuScope

BIOPLATFORMS AUSTRALIA

CSIRO

HIA
HEAVY ION ACCELERATORS

IMOS
Integrated Marine Observing System

CSIRO Marine National Facility

MICROSCOPY AUSTRALIA

National Imaging Facility

NCI AUSTRALIA

Phenomics Australia
Enabling precision medicine

PHRN
Population Health Research Network

pawsey

tern
Ecosystem Research Infrastructure

THERAPEUTIC INNOVATION AUSTRALIA

Supporting
NCRIS Capabilities

aarnet
Australia's Academic and Research Network

AUSTRALIAN ACCESS FEDERATION AAF.EDU.AU

ARDC
Australian Research Data Commons

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

ARDC is enabled by NCRIS

ARDC STRATEGY

PURPOSE

To provide Australian researchers with competitive advantage through data.

MISSION

To accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets.

PEOPLE & POLICY

Connecting the ARDC

- Communications
- Engagements
- Skills & Workforce Development
- Data Policies

PLATFORMS & SOFTWARE

Accelerating research insights and supporting collaboration

- Platforms for Analysis & Curation
- Research Software

DATA & SERVICES

Maximising the value of Australia's data assets

- Data Assets
- Information Infrastructure
- Data Capability

STORAGE & COMPUTE

Providing foundation infrastructure

- Research Computing Cloud
- Data Retention

AUSTRALIA'S NATIONAL RESEARCH DATA COMMONS

What is RAiD?

A RAiD is a persistent identifier for…

- Research projects and activities, linking organisations, people, inputs, and outputs to a project and providing key project information found nowhere else
- RAiD is governed by ISO Standard 23527:2022
  - The ARDC is the international Registration Authority
  - The ARDC is also a Registration Agency focusing on Australasia

A RAiD is not for…

- Grants
- Researchers
- Durable organisations or organisational units (teams, centres, groups, departments, etc.)
- Documents, papers, articles, books, recordings, or other digital objects
- Software or datasets
- Instruments
- Samples or specimens

How does RAiD work?

A RAiD has two parts: The **RAiD identifier** and the **RAiD metadata record**.

RAiD uses the Handle system to create **globally unique, persistent identifiers**. The Handle is like the address on an envelope, while the metadata record makes up its contents.

The **metadata record** includes other PIDs for various project inputs and outputs like:

- Collaborators (people)
- Organisations (institutions)
- Grants, awards, and investments
- Infrastructure, tools, instruments, and services
- Data and software
- Publications, reports, and events

Where necessary, the **RAiD metadata record** captures project information found nowhere else, such as a project's title, description, and subject. Many RAiD elements can be time-bound. Relationships between elements can be described or qualified.

Example RAiD metadata record

https://raid.org/13.1010/463UQDMR

Title (primary) Lorem Ipsum Project
Title (acronym) LIP
Start date 2023-01-30
Description (short) Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas vitae condimentum nisl, eget ornare felis. Morbi pretium erat eu ultrices interdum.

Principle investigator orcid.org/0000-0002-3843-0000
Role https://credit.niso.org/contributor-roles/conceptualization/
Co-investigator orcid.org/0000-0002-3639-2080
Role https://credit.niso.org/contributor-roles/data-curation/

Lead organisation ror.org/00rqy9422
Partner organisation ror.org/03b94tp07

Grant doi.org/10.8948/908234D93EAF
Dataset doi.org/10.1594/PANGAEA.726855
Article doi.org/10.1038/nphys1170
Instrument doi.org/10.1337/jdlc-tima

Sample XXAB000IH

Local storage uq.edu.au/114/32
Cloud storage 79.152.127.243

Sub-project https://raid.org/13.1010/401XQPOI

Why a Project ID?

Projects are where research happens

- Reflects collaborative practices while accommodating sole researchers
- Time-limited but identifiable and meaningful ‘container’

Projects are not grants

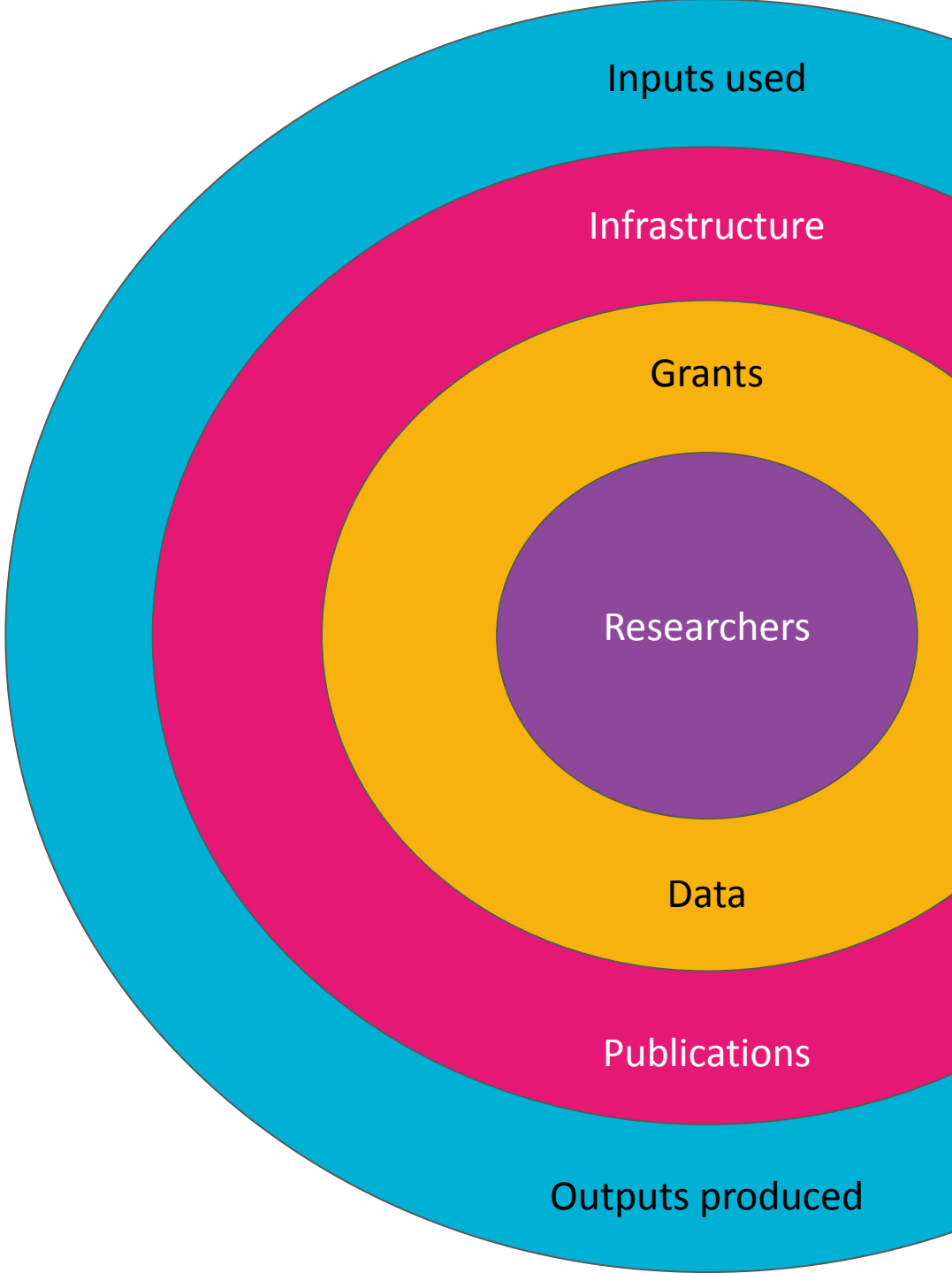
- Not 1:1 - some projects never grants, others have many grants
- RAiD captures longer-term outputs, outcomes, and impacts - grants close, but projects may produce outputs for many years

Projects evolve

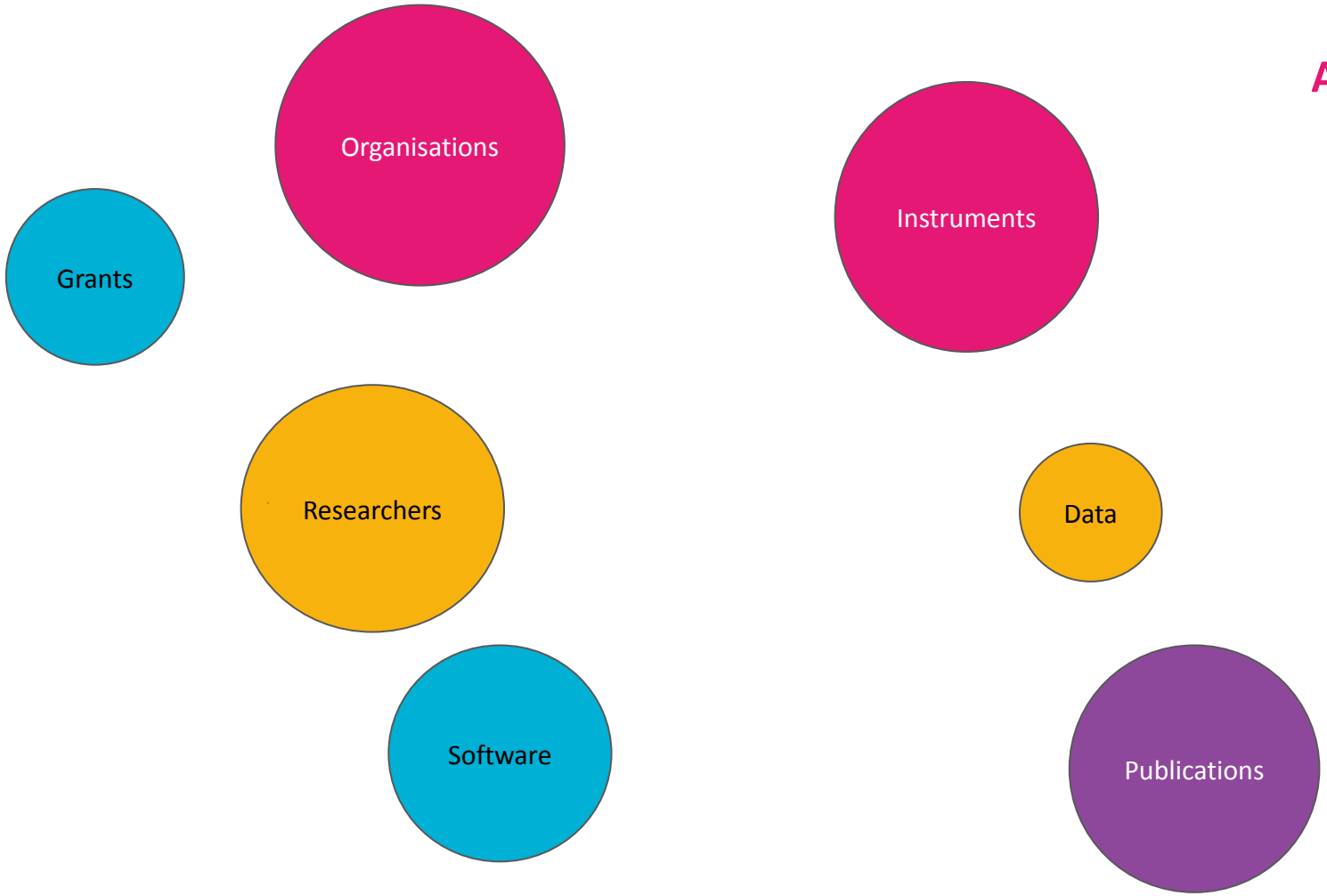
- PIDs aimed at stable digital artifacts provide ‘snapshots’ but projects change continuously
- RAiD metadata is designed to evolve over time as contributors, organisations, etc., change, producing a history of the project

Projects are where research can be administered

- Common concept in Research Information Systems
- Appear in other PIDs
- Appear in domain-specific metadata standards



Research components



Actions

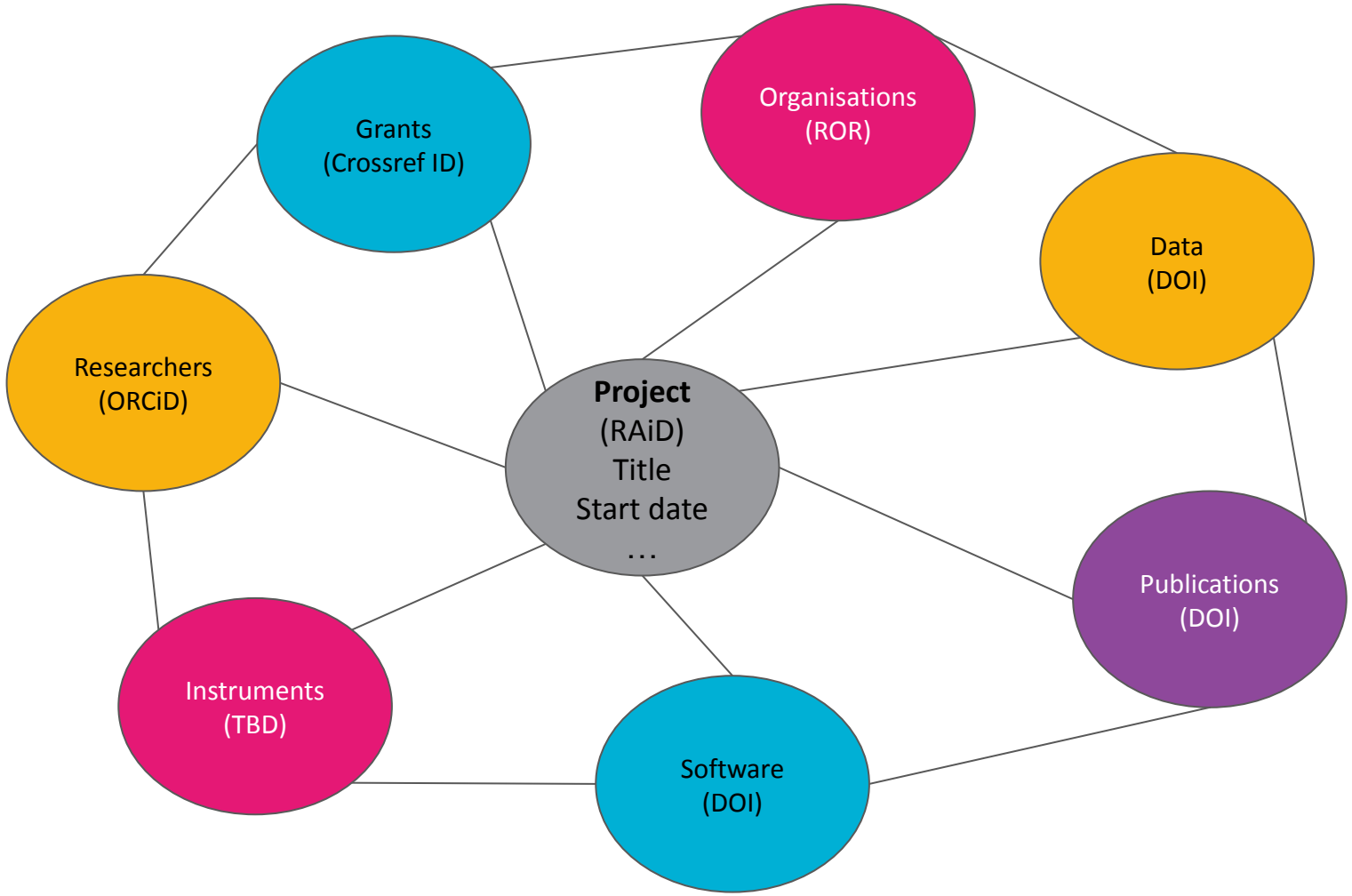
- Uses (infrastructure)
- Is funded by (grant)
- Creates (dataset)
- Hosted by (organisation)

Entities

- Researchers
- Organisations
- Data
- Software
- Publications
- Grants
- Samples
- Instruments
- Services

The ‘project’ as nexus of research inputs and outputs

- Project has Persistent ID
- Project components (people, organisations, inputs, outputs) in metadata record using PIDs
- Additional project information captured when necessary
- Research actions and project changes reflected in history



What are the problems with current practice?

Information about projects is distributed and siloed

- Partly in institutional Current Research Information Systems (CRISes)
- Sometimes in other university systems (e.g., finance)
- Other information on project websites or staff profiles
- Lots of double-entry of data

Output, outcome, and impact tracking is hard

- Longer timeframes not accommodated by grant reporting
- Project outcomes not fully captured by individual ORCID records

Project metadata not routinely or comprehensively captured

- Project-level metadata often required by data repositories, but may be incomplete
- Project information often maintained in an ad-hoc manner, sometimes lost
- Important for output provenance (e.g., datasets, publications)

No standardisation

- Information about projects, where it exists, is non-standard
- Often not machine readable

What are the benefits of using RAiD?

Provides a ‘single source of truth’

- Reduces double-entry of data
- Ensures coordination across organisations
- Saves time on administration and reporting

Supports reporting and impact measurement

- Facilitates tracking and reporting of inputs and outputs
- Grants insights into investments and outputs
- Collects evidence for understanding impact
- Facilitates better strategic intelligence on outcomes
- Supports better tools for analysis and decision-making

Captures research provenance

- Captures the evolution and history of a project
- Create a timeline of (inter)actions
- Comprehensive record of project make-up

Standardising project identification

- Governed by an ISO standard

What is the potential efficiency impact?

Estimate of active research projects in any given year

- 50k projects in the UK
- 21k projects in Australia
- 625k projects in the US
- 1.5M projects in the OECD

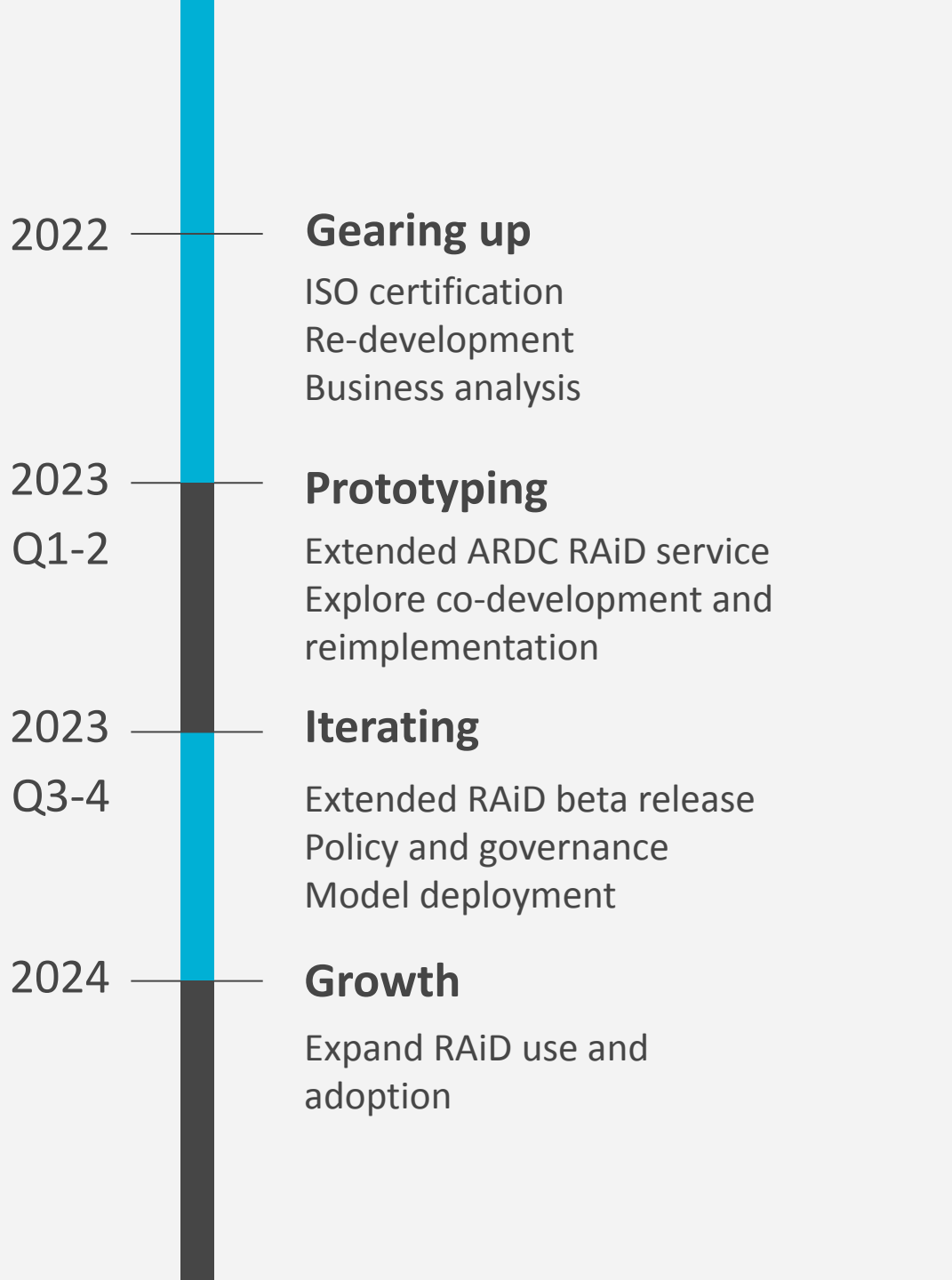
In Australia, elimination of double-entry of project metadata could save approximately:

- 2.9k person-days per year
- AUD $2.7M per year

Combined with publication and grant PIDs, could save approximately:

- 37.9k person-days per year
- AUD $23.8M per year

Source: Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). *Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems*. Zenodo. https://doi.org/10.5281/zenodo.7100578



Where are we now?

- **Done**: ISO certification with ARDC as Registration Authority and multiple Registration Agencies
- **Done**: Redevelop existing ARDC service while gathering requirements and consulting with stakeholders
- **In progress**: Extend service with new metadata schema, landing pages, updated API, new user interface, improved integration with other PIDs.
- **In progress**: RAiD Registration Agency handbook encapsulates policy and governance
- **Future**: International outreach and engagement to drive uptake

Subscribe to the **ARDC CONNECT** newsletter

THANK YOU

ardc.edu.au

contact@ardc.edu.au

+61 3 9902 0585

@ARDC_AU

Australian-Research- Data-Commons

Australian Research Data Commons

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

ARDC is enabled by NCRIS