Tracking research projects: the heart of open research infrastructure

eResearch NZ 2022

11 February 2022

PRESENTED BY
Natasha Simons (ARDC) & Yvette Wharton (The University of Auckland)

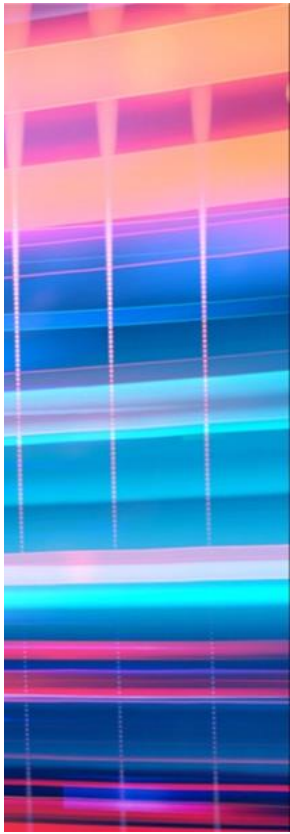
Australian Research Data Commons

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.

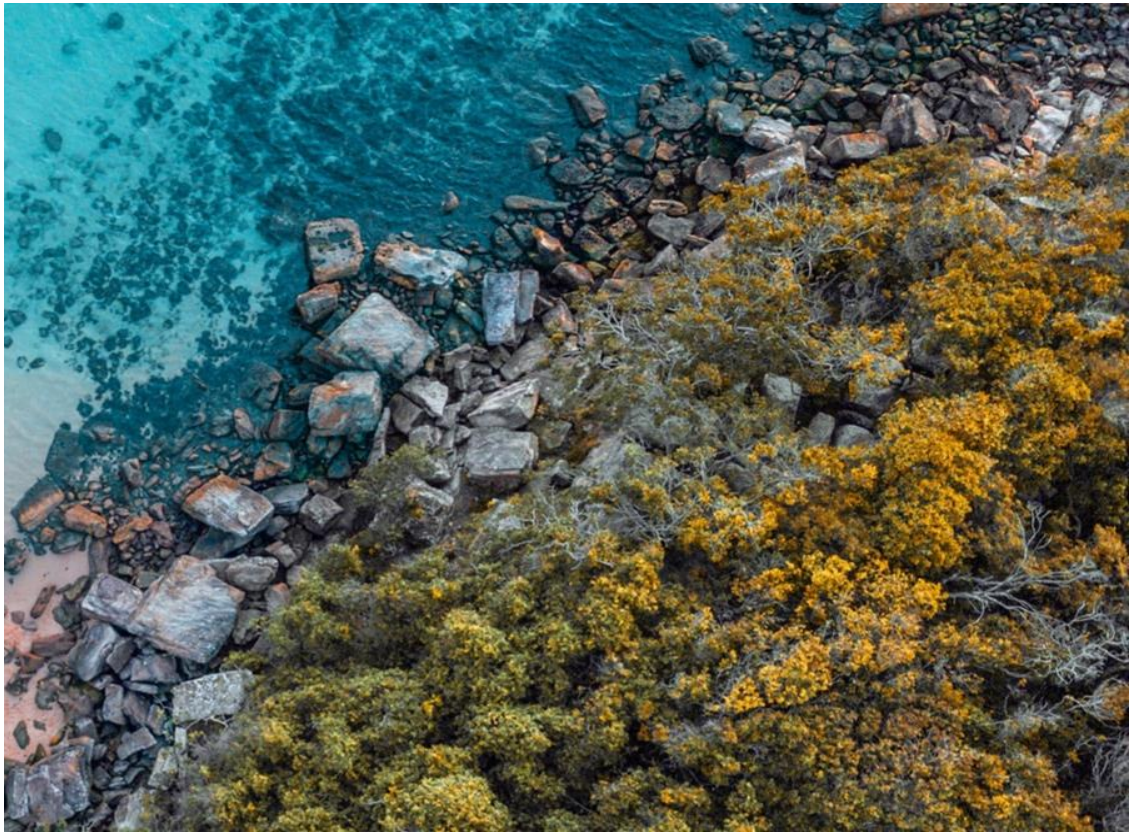


NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

ARDC is enabled by NCRIS

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.



Persistent Identifiers (PIDs) as National Infrastructure for Research

ARDC provides PID services free to charge to the Australian research sector:

- DataCite Consortium lead - **DOI** for research data, software, grey literature, instruments
- **RAID** - for research projects
- **IGSN** Allocating Agency - for physical samples
- **PURL** - for research grants (hopefully move to DOIs in the future)
- **Handle** - for data that is not suitable for a DOI

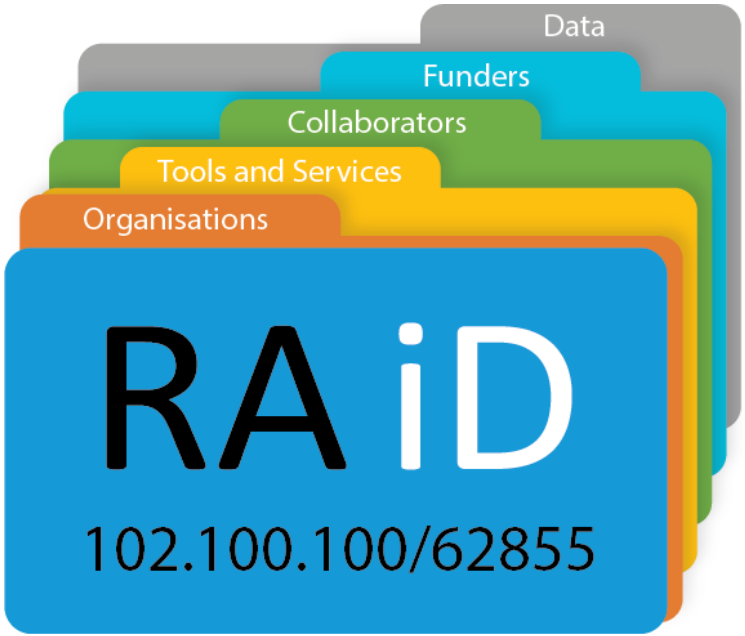
ORCID - via membership of the Australian ORCID Consortium led by Australian Access Federation

The ARDC's suite of PID services, combined with those available through the AAF led Australian ORCID Consortium, form the backbone of enabling FAIR research outputs via PIDs in Australia.

ARDC PIDs Policy - https://ardc.edu.au/about_us/policies-and-guidelines/persistent-identifiers-policy/

Globalization of Information Infrastructure: Opportunity

The RAiD (Research Activity Identifier) is a persistent identifier for research projects



Grants vs Projects

A grant is something you **get**

A project is something you **do**

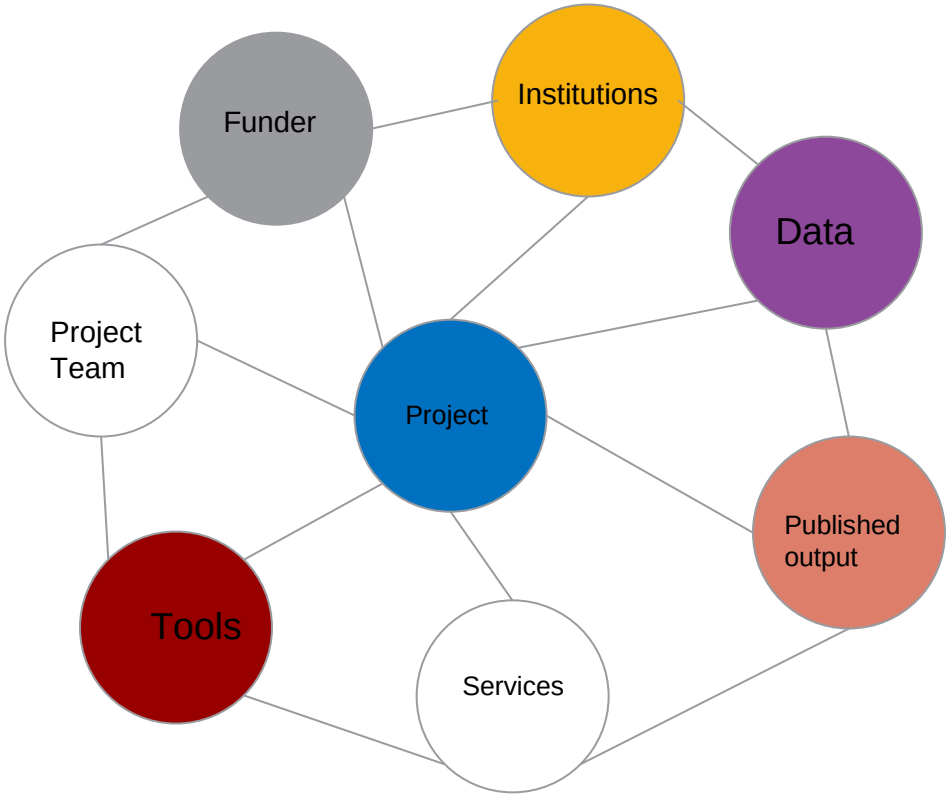
Project activities include:

- Assign researchers to project
- Write data management plan
- Use research infrastructure
- Process and analyse data
- Write reports
- Publish articles
- Publish data and software
- etc

A project changes over time so how do we track?

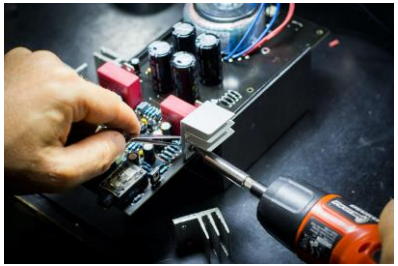
RAiD puts the Project at the centre

- Project has Persistent ID
- Entities recorded in Metadata
- Research actions reflected in project timeline
- Related PIDs recorded in metadata



RAiD metadata envelope

RAiD on an international stage



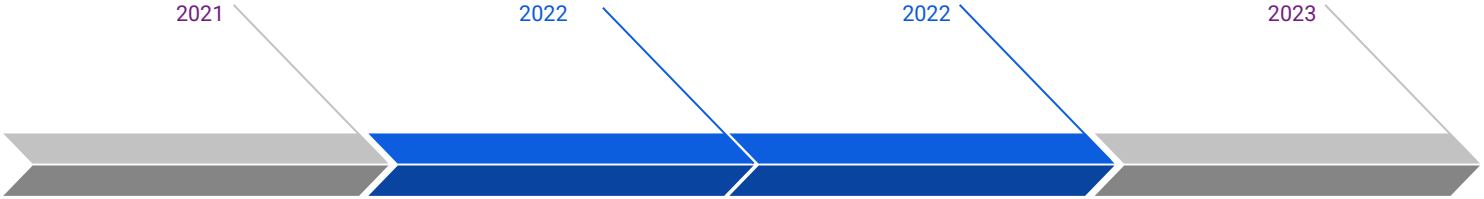
Technical

Community

Sustainability

Image source: World Economic Forum

RAiD Roadmap



Gearing up for accelerated development

- Staff recruitment
- Meetings with partners and potential partners
- Identify or kick off pathfinder projects with partners e.g. FAIR workflows project with DataCite
- ISO Certification round (Dec)

Rapid technical development

- Extend API
- Increase integrations with PID systems and with research information management systems (CRIS/RIMS)
- Further develop code base

Community & governance focus

- Establish RAiD Advisory Group (international)
- Develop international governance model
- Pathfinder projects with partners identified or in progress
- Metadata schema development

Grow RAiD use and adoption

- International collaboration

Our journey at the University of Auckland

Uni

	Level 1 **INITIAL** *Process disorganised & ad hoc*	Level 2 **DEVELOPMENT** *Process is under development*	Level 3 **DEFINED** *Process is standardised and communicated*	Level 4 **MANAGED** *Process is managed and measured*	Level 5 **OPTIMISED** *Focus is on continuous improvement*
Strategy, Governance, & Policies	- Strategy, policies & procedures may be underdeveloped, not up to date, and/or inconsistent - Institutional awareness is low - Institutional governance of research data are ad hoc or absent	- Strategy, policies & procedures, governance mechanisms are developing & harmonised for specific tasks - Institutional awareness of roles and responsibilities is increasing	- Strategy, policies & procedures, and governance are defined and operationalised into research practice - Widespread institutional awareness of roles and responsibilities	- Strategy, policies & procedure, and governance are integrated and accepted as part of culture & subject to audit	- Institution is a recognised leader following international best practice in capability maturity roadmapping - Strategy, policies & procedures, and governance are accepted as BAU, audited and regularly reviewed to align with current best practice
Māori Data Sovereignty	- There is little or no recognition of Māori data - Māori data stored without consideration of sovereignty issues	- Recognition of Māori interests in data - Māori participation in a data access committee - Some integration of Māori principles (e.g. kaitiakitanga)	- Māori governance of data & protocols for data access - Māori data definition is consistent - Māori can access raw data about their collective on request - Full range of storage options, including NZ based, as required	- Māori ownership of data - Māori data access committee- - Māori have full access to data about their collective -Widespread use of Māori principles and DM protocols	- Māori Data sovereignty principles are intrinsic to RDM at an institutional level - RDM is aligned with the principles of rangatiratanga, whakapapa, whanaungatanga, kotahitanga, manaakitanga and kaitiakitanga
Data Management	- Metadata management is chaotic & understood by only a few - Data quality measures are ad hoc or absent	- Responsibilities are defined & skills are developed - Simple DMPs available - Data collection guidelines in development	- Processes are standardised and integrated - All data are assigned an appropriate globally unique persistent identifier (e.g.DOI)	- All datasets described & metadata shared - DMPS used for provisioning and active management of data	- Continuous improvement applied to processes & capabilities - Integrated research ecosystem (e.g leveraging multiple persistent identifiers to enable seamless RDM)
Data Operations	- Simple data sharing can be a challenge - Curation & preservation services absent or disorganised	- Project-based data sharing services become available - Curation practices are developing and awareness of the value of curation is increasing - Value and practice of preservation is not recognised	- Project-based data management practices are developed and documented - Data sharing is straightforward - Curation practices are standardised and widely understood - Preservation practices for selected data	- Data are FAIR: sharing becomes commonplace and embedded in practice; curation understood as critical; and, data are being preserved in alignment with policy	- In addition to data being FAIR, data management operations (sharing, curation and preservation) are fully automated and machine accessible throughout the organisation
Infrastructure, Platforms, & Architecture	- IT infrastructure is acquired, deployed, and managed inconsistently, not easily discoverable, supportable, nor documented as services	- Responsibilities for provisioning, maintaining, and lifecycle management are defined - Beginning to integrate instruments,, storage & transfer services, and research compute	- Widespread availability of data platforms and tools, including analysis, visualisation - Facilities are well-defined, standardised, and integrated - Automated provisioning of project infrastructure in development	- Funding adapts to platform needs. - Platforms are well-managed within a defined research-delivery architecture - Auditing of platforms and architecture in place	- IT infrastructure management optimises the IT infrastructure evolved in previous levels through continuous focus on management and improvement of data assets
Skills & Support	- Data management planning is unsupported - Training is ad hoc or missing - QA is ad-hoc or absent - Staff are unable to locate required support or documents	- Investment in skills and processes - Data management planning is used on projects, documentation & training are developed	- Widespread availability and uptake of training and skills development in data management - QA becomes feasible on training and support services	- QA is routinely applied to processes, results feed into future planning	- Support services and training processes are optimised and periodically refined

Platform, Infrastructure, & Architecture

Encompassing data services and platforms, infrastructure, architecture, and the resources that enable them.
1. Data platforms (a set of services, tools and technologies that enable a community e.g. clinical research to meet their specific needs) & services (what researchers consume, the 'front-door' to infrastructure)
2. Infrastructure, including storage, research compute - hardware and software
3. Architecture (the plan, criteria and connectors by which the research delivery ecosystem is developed and integrated)
4. Roles and responsibilities

Level 1 **INITIAL**

Process disorganised & ad hoc

1.1 Provision of platforms and services for data capture, organisation, analysis and visualisation is ad hoc

1.2 Data sharing and transfer processes are ad hoc, including email, portable drives, personal cloud-based storage and institutional offerings

2.1 IT infrastructure for RDM is acquired, deployed, and managed inconsistently, not discoverable, not supportable, and not documented as services

2.2 IT infrastructure is inadequate to support targeted research data needs

2.3 Research data storage facilities may be rudimentary such as shared drives or free web services

3.1 Lack of consideration of IT research delivery architecture

3.2. Capability and processes to support minting of persistent identifiers are non-existent or limited

4. IT/research data platforms, services, and,infrastructure expertise is held by only a few individuals who may be unable to cope with the demand

Level 2 **DEVELOPMENT**

Process is under development

1.1 Some provision of standard platforms and services for data capture, organisation, analysis and visualisation but no defined process to access/acquire non-standard offerings

1.2 Some provision of standard data sharing and transfer platform, services and tools, including internal and external data sharing processes, to meet most needs

1.3 Developing integration of research compute, storage & transfer platforms and services

1.4 Institutional Data Management Planning tool available to some

1.5 Provision of an data publishing service (institutional data repository/repositories) with limited integration

1.6 Services to support institutional, and community specific data registries are accessible to very few

2.1 Institutional needs for storage of research data are recognised and procedures in place to ensure persistent storage for data, whether created in-house or imported

2.3 Data storage may be in-house, shared with another institution or outsourced.

3.1a IT infrastructure provisioning and IT equipment are recognised in research delivery architecture

3.1b Research compute service available but processes not integrated within research delivery architecture

3.2 Increasing institutional ability to mint persistent identifiers across the research lifecycle

Research compute service available but processes not integrated within research delivery architecture

Level 3 **DEFINED**

Process is standardised and communicated

1.1 Widespread availability of data platforms and services to meet different research community needs

1.2 Awareness and access to a range of data sharing and transfer platforms and services, including sharing clinical data with other organisation

1.3 Automated provisioning of project infrastructure in development

1.4 Machine actionable DMP tool integrated with systems and researcher workflow

1.5 Provision of an institutional data repository integrated with systems and researcher workflow

1.6 Services to support institutional, and community specific data registries are accessible are widely accessible

2.1 IT infrastructure needs for RDM are coordinated with capital expenditure planning and with technology transition planning

2.2 Some integration with national infrastructure services (e.g. HPC, storage, repositories and transfer)

2.3 Well-defined internal and external data storage facilities are available & in the process of being integrated with other data management services

2.4 Research compute service fully integrated within institution architecture

3.1 Facilities are well defined, standardised and integrated as part of architecture for research delivery

3.2 Ability to mint range of persistent identifiers across the research lifecycle

3.3 Identity, authentication and access management are standardised

3.4 Architecture supports agreed data classification

4. Responsibilities for IT infrastructure,

Level 4 **MANAGED**

Process is managed and measured

1.1 Data management platforms and services are fit for purpose, managed and auditable

1.2 Data sharing and transfer platforms and services, including sharing clinical data with other organisation, are managed

1.3 Automated provisioning of project infrastructure becomes standard practice

1.4 DMP adherence is measured

1.5 Institutional and national platforms and services fully integrated and available to all researchers

1.6 Automated testing and auditing of platforms and services to support institutional and community specific data needs registries are accessible are widely accessible

2.1 Scalable computational infrastructure available to support a range of targeted research compute services, including HPC, AI, machine learning, etc.

2.2 Data treated as key organisational asset and supported by targeted IT infrastructure

2.3 Well-defined matrix of internal & external data storage facilities options to suit different needs and data classifications

3.1 RDM platforms and services well managed within a defined architecture that incorporates University and external ecosystems

3.2 Managed consumption, linking, validation and curation of persistent identifiers (including for organisations, people, projects, instruments, research outputs) that are minted internally and externally

Level 5 **OPTIMISED**

Focus is on continuous improvement

1.1 Data platforms are continuously developed with user community and support international best practices

1.2 Data services/platforms automatically capture metrics and audit findings to inform service quality and drive change in rapidly changing technology support areas

1.3 Use of data platforms, services and tools is driven by measures of effectiveness

2.1 IT infrastructure management optimises the IT infrastructure evolved in previous Levels through continuous focus on management and improvement of data assets

2.2 Concerted efforts to optimise platforms and architecture to fit emerging research needs

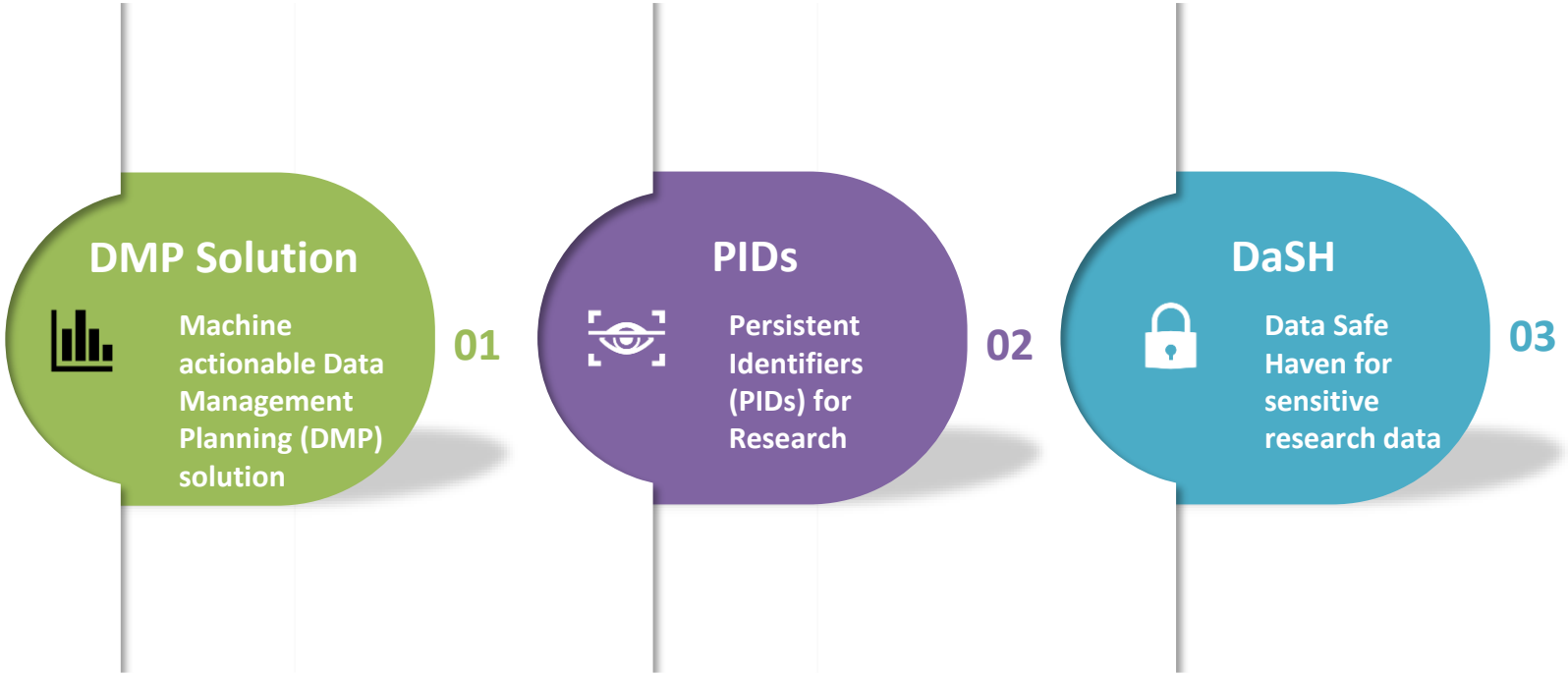
2.3 All aspects of data IT infrastructure facilities and services are monitored through highly visible automated testing and feedback systems

2.4 Automated provisioning of integrated research compute service with managed upgrades and migration

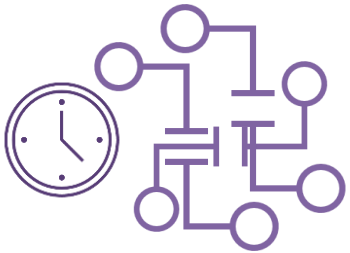
3. Ecosystem approach to architecture supported by use of persistent identifiers and systems integration for seamless experience, fit-for-purpose, user centric experience that is continuously adapting

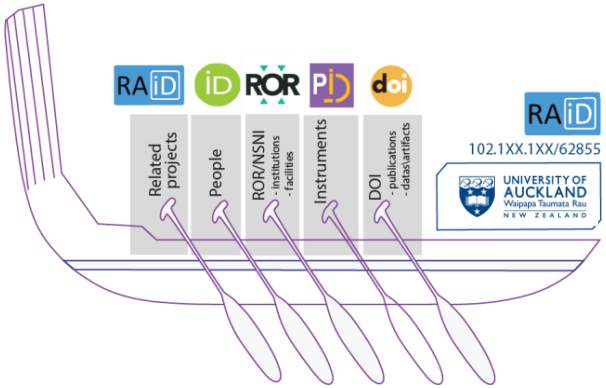
4. Individual IT support roles enact well-publicised technology transition and process improvement plans

RDM Programme



PIDs for Research





ISSUES

- Disconnection
- Duplication
- Difficulty
- Inaccurate reporting

BENEFITS

- Research time
- Reporting
- Reputation
- Planning

5 Primary PIDs for:

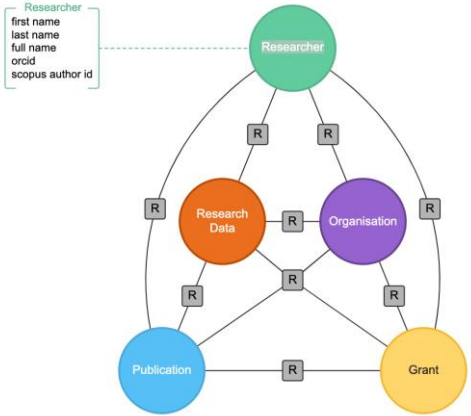
RAID - Projects

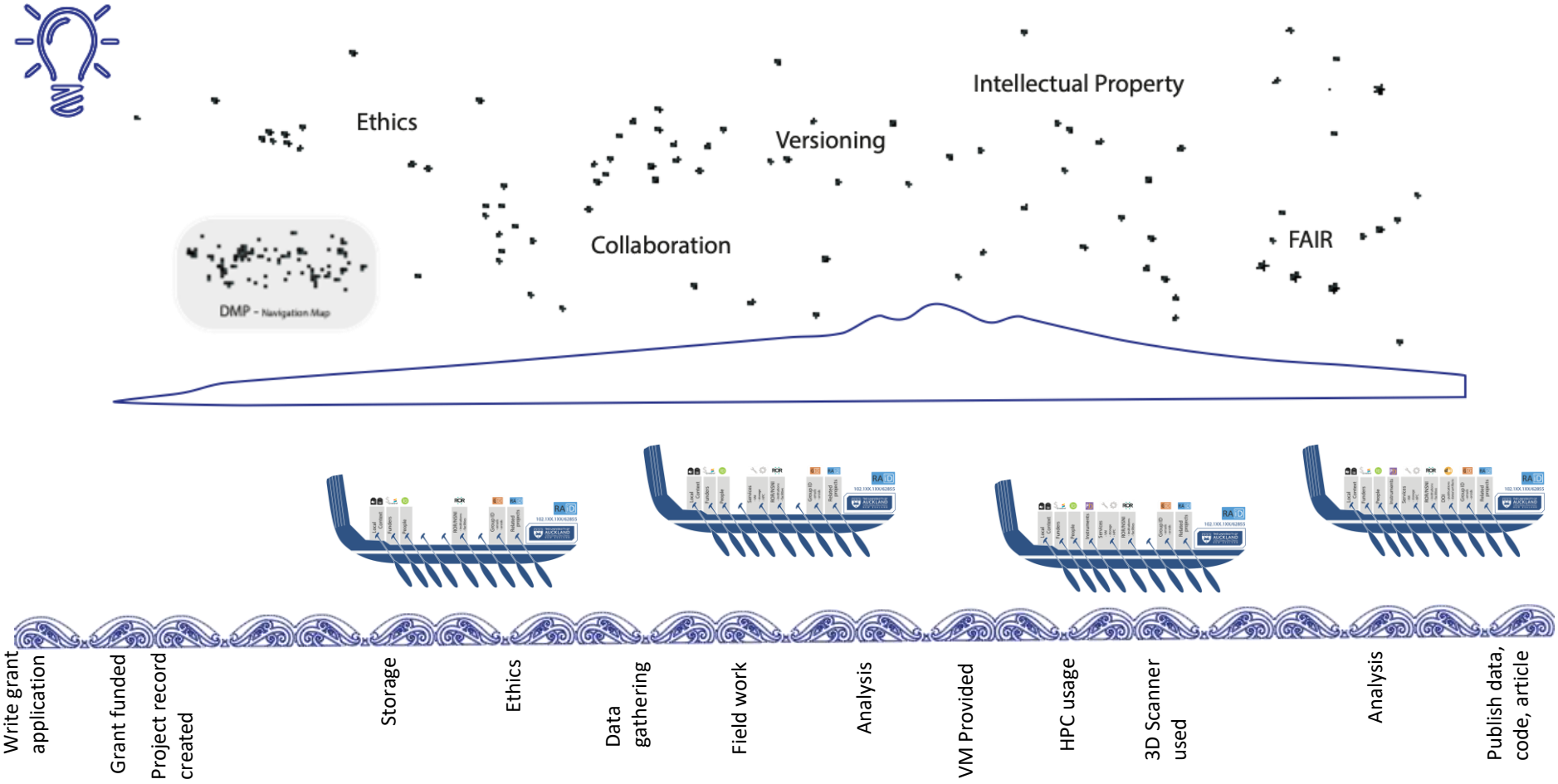
ROR -Facilities/Org

DOI - Data

PIDINST - Instruments

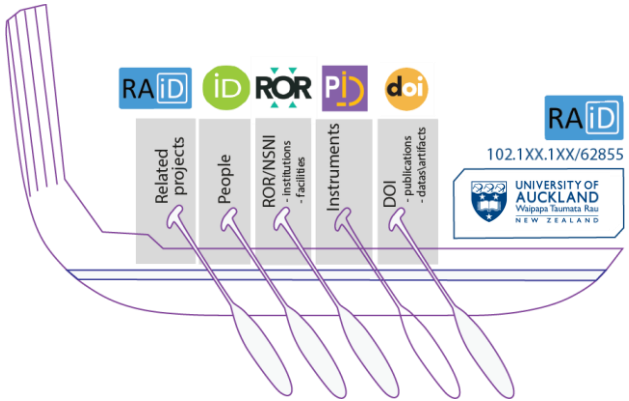
ORCID - *People




Ethics
Intellectual Property
Versioning
Collaboration
FAIR
DMP - Navigation Map
Write grant application
Grant funded
Project record created
Storage
Ethics
Data gathering
Field work
Analysis
VM Provided
HPC usage
3D Scanner used
Analysis
Publish data, code, article

Ko te whaea te takere o te waka

Mothers are like the hull of a canoe,
they are the heart of the family



PRESENTED BY

Natasha Simons

Associate Director, *Data & Services, ARDC*

ardc.edu.au

natasha.simons@ardc.edu.au

@n_simons

https://www.linkedin.com/in/natasha-simons-69814837/

https://orcid.org/0000-0003-0635-1998

Australian Research Data Commons

NCRIS National Research Infrastructure for Australia

ARDC is enabled by NCRIS

PRESENTED BY

Yvette Wharton

eResearch Solutions Lead, *Centre for eResearch, UoA*

eresearch.auckland.ac.nz

y.wharton@auckland.ac.nz

@y_vettles

https://nz.linkedin.com/in/yvette-wharton-4271abb4

https://orcid.org/0000-0002-6689-8840

THE UNIVERSITY OF AUCKLAND NEW ZEALAND

ARDC Australian Research Data Commons

NCRIS National Research Infrastructure for Australia An Australian Government Initiative — ARDC is enabled by NCRIS

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