

FAIR Principles for Research data



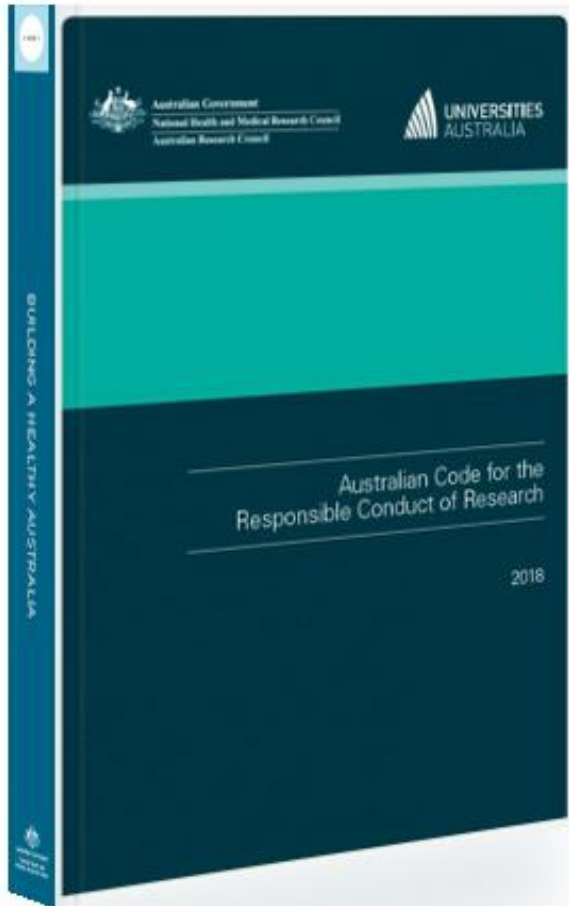
Flinders
UNIVERSITY

FEARLESS

Open Research Data Management

- **Research Data Management**
 - Overview of managing research data
 - Beginning with the end in mind
- **FAIR Data**
- **ROADS**
 - Repository of Open Access Data Sets
 - Demonstration
- **Good data practice**
- **Australian Research Data Commons (ARDC)**

Australian Code for the Responsible Conduct of Research



Responsibility 22 Researchers will:

- Retain clear, accurate, secure and complete records of all research including research data and primary materials.
- Where possible and appropriate, allow access and reference to these by interested parties

Balance

The EU's H2020 Program Guidelines on FAIR Data, describes this balance as:

“as open as possible and as closed as necessary”

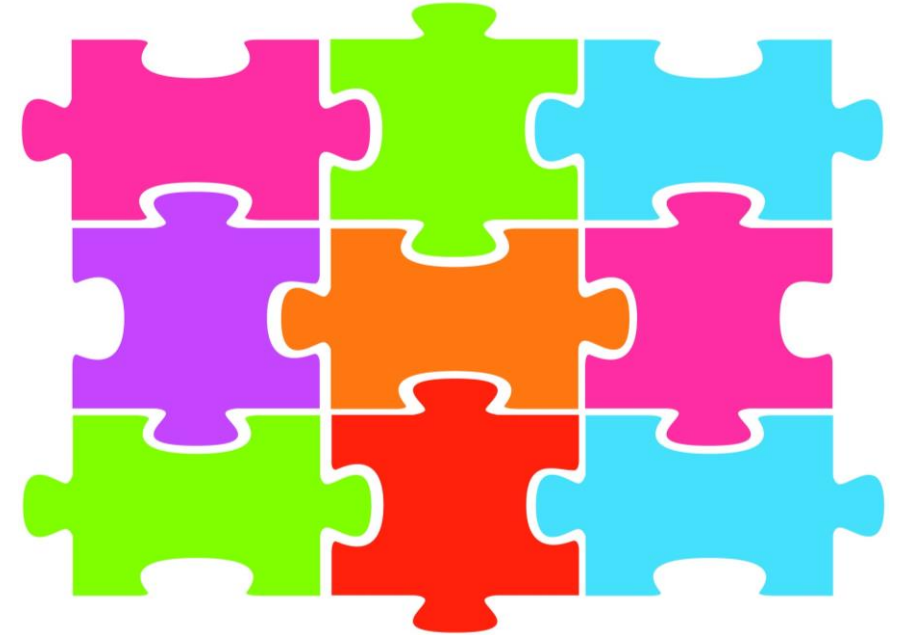
Open to allow access and reference to these by interested parties

Closed to safeguard the privacy of research subjects and maintaining security for sensitive data



Research Data Management Core Elements

- **Data Collection**
- **Documentation and Metadata**
- **Ethics and Legal Compliance**
- **Storage and Backup**
- **Selection and Preservation**
- **Responsibilities and Resources**
- **Data Sharing**



Data Collection

- What data will you collect or create? &
- How will the data be collected or created?
 - Type, format and volume of data?
 - Will your format and software/equipment enable sharing or long-term access?
 - Does your discipline have data standards or 'norms'?
 - How will you structure and name your folders and files?
 - How will you handle versioning?
 - What quality assurance processes will you adopt?
- Is there any existing data that can be reused?

Looking for Research Data

Google Dataset Search

<https://datasetsearch.research.google.com/>



<https://researchdata.edu.au/>



<https://data.gov.au/>



<https://www.abs.gov.au/>



<https://data.sa.gov.au/>

Documentation and Metadata

- What documentation and metadata will accompany the data?
 - What information is needed for the data to be to be read and interpreted in the future?
 - How will you capture / create this documentation and metadata?
 - What metadata standards will you use and why?

Ethics and Legal Compliance

- How will you manage ethical issues?
 - Will data sharing be included in your ethics application?
 - Have you gained consent for data preservation and sharing?
 - How will you protect the identity of participants if required?
 - How will sensitive data be handled to ensure it is stored and transferred securely?
- Management of copyright and Intellectual Property Rights (IP) issues?
 - Who owns the data?
 - Sponsor
 - University
 - Student
 - Are there any restrictions on the reuse of third-party data?

What does the National Statement say?



- Data or information may be deposited in an open or mediated access repository or data warehouse, similar to an archive or library, and aggregated over time. Archived data or information can then be made available for later analysis, unless access is constrained by restrictions imposed by the depositor/s, the original data custodian/s or the ethics review body.
- **When planning to share data or information** with other researchers or to establish or add them to a databank, **researchers must develop data management plans** in accordance with the guidance provided in 3.1.45. This plan should enable the sharing of data and information and propose appropriate conditions on the sharing of data and information.
- In any proposals to share or disclose research data or information, **researchers should distinguish between disclosure to specific third parties, sharing with other researchers and disclosure to the public** and clarify whether the sharing or disclosure of data or **information is subject to participant consent, other voluntary agreements or mandatory requirements.**
- **Participants should be advised** that publication or funding requirements may require **submission of data or information to controlled access repositories that meet international security and safety standards** for sharing with researchers globally.
- **Sharing information must be consistent with the consent that has been obtained for the research project.**

What does the Australian Code for the care and use of animals for scientific purposes say?



Institutions must promote compliance with the Code by:

- promoting and facilitating adoption of the governing principles of the Code in all aspects of animal care and use, including coordinating planning and operations, **and sharing resources and information**, to facilitate the application of Replacement, Reduction and Refinement (the 3Rs).

Information must be provided on:

- **opportunities for sharing of tissues and other biological material from animals**

Why should you share your data?

- **Over-researched and vulnerable research populations**
 - Has this type of research been done before? If so, do you really need to repeat the research and ask the same questions again or can you use an already existing dataset?
- **The 3Rs**
 - Replacement, Reduction and Refinement – requirement under the Code!
- **Data availability**
 - Time limited research projects (e.g., Honours, Masters by Research) may not be able to be conducted without access to available datasets.
- **Research Transparency & Public Funding**
 - Most research is funded by the public so the public should learn and benefit from the work of researchers. Improves trust in research – very important!

Ethical issues that must be considered

- **Are there any laws/contractual obligations which will prevent you from sharing your data?**
 - e.g. Defence Trade Control Act, Autonomous Sanctions, Foreign Interference, Confidentiality Agreements, Mandatory Reporting, Illegal Behaviour
- **What about personal and sensitive data?**
 - Do you really need to collect this data? Is it essential to your research project? How will you anonymise the data? How will you protect the identity of your participants?
- **Data storage arrangements**
 - Where will the data be stored? Is it stored securely and in accordance with accepted safety standards? Who will have access to the data?
- **Consent**
 - Will you obtain consent from your participants? Will the data sharing be consistent with the consent arrangements? Will your participants know where the data will be stored and who will have access to it? Will you share data collected from minors? Are their parents/guardians aware of this?
- **Risk**
 - Can the release of the data harm you, your participants or others not involved in the research project?

Storage and Backup

- How will the data be stored and backed up during the research?
 - Support from IDS
 - Can you use existing University infrastructure
 - Do you have sufficient storage or will you need to include charges for additional services?
 - How will the data be backed up?
 - Who will be responsible for backup and recovery?
- How will you manage access and security?
 - How will you control access to keep the data secure?
 - How will you ensure that collaborators can access your data securely?
 - If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?

Storage Option	Recommended for Research			Enterprise Storage Solutions	
	R Drive	Cloudstor	ROADS	OneDrive	Teams / SharePoint
Recommended	✓	✓	✓	✓	✓
Data classification	Restricted	Internal Use/Public	Public	Internal Use	Internal Use
Stored in Australia	✓	✓	✓	✓	✓
Accessible off campus	✗*	✓	✓	✓	✓
BackUp & disaster recovery	✓	✗	✗	✗	✗
External collaboration	✗	✓	✗	✓	✓
Available to DeepThought HPC	✓	✗	✗	✗	✗
Stage of data capture**	Active & Storage	Active	Open Data	Active	Active
Total max size	by approval	1TB	25GB	1TB	1TB
Max file size	16TB	100GB	5GB	15GB	15GB
Version control	✓	✓	✓	✓	✓

* R drive is only accessible off campus by using the Flinders remote network access (VPN)

** Stage of data capture - **Active** is for research projects that are still gathering/processing data, **Storage** is for finalised data that needs to be held but not necessarily published in an open access repository, **Open Data** is for data that can/should be made public in an open access repository.

Selection and Preservation

- Which data should be retained, shared, and/or preserved?
 - Research Data must be kept for minimum of 5 years
 - From publishing date
 - What data must be retained/destroyed for contractual, legal, or regulatory purposes?
 - How will you decide what other data to keep?
 - What future uses are there for this research data?
 - How long will the data be retained and preserved?
- What is the long-term preservation plan for the dataset?
 - Where e.g. in which repository or archive will the data be held?
 - What format? (e.g. .xlsx vs .csv)
 - What costs if any will your selected data repository or archive charge?

Responsibilities and Resources

- Who will be responsible for data management?
 - Who is responsible for managing data?
 - How will responsibilities be split across partners/collaborators?
- What resources will you require to deliver your data plan?
 - Is additional specialist expertise (or training for existing staff) required?
 - Do you require hardware or software which is additional or exceptional to existing institutional provision?
 - Will charges be applied by data repositories (e.g. for large datasets)?

Data Sharing

- How will you share the data?
 - What platform will you use
 - ROADS
 - Third party repository
 - Discipline specific
 - Figshare+ for extra large datasets
 - When will you make the data available?
 - Will you pursue getting a persistent identifier for your data?
 - ROADS – All datasets get a DOI



Data Sharing



- Are any restrictions on data sharing required?
 - Be mindful of:
 - Contractual obligations
 - Export controls (e.g. defence related areas)
 - Ethics (did you obtain consent? Have you deidentified your data)
 - Collaborations - Data Sharing Agreements
 - Prevent inadvertent misuse/mistaken use of material
 - What action can you take to overcome or minimise restrictions?
 - For how long do you need exclusive use of the data and why?
 - Publishing
 - Patents

Open Research Data Management



- **Open Science**
 - Demonstrates transparency in scientific enquiry
- **Reproducibility**
 - Addresses (some) reproducibility issues in academia
- **Grant funders**
 - Including ARC & NHMRC pushing for funded research to be made available
- **Publishers**
 - Requiring data to be made available
- **Reuse of data**
 - Reduces the cost of producing/recreating data

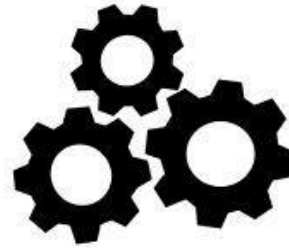
F
Findable



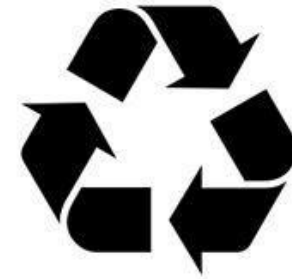
A
Accessible



I
Interoperable

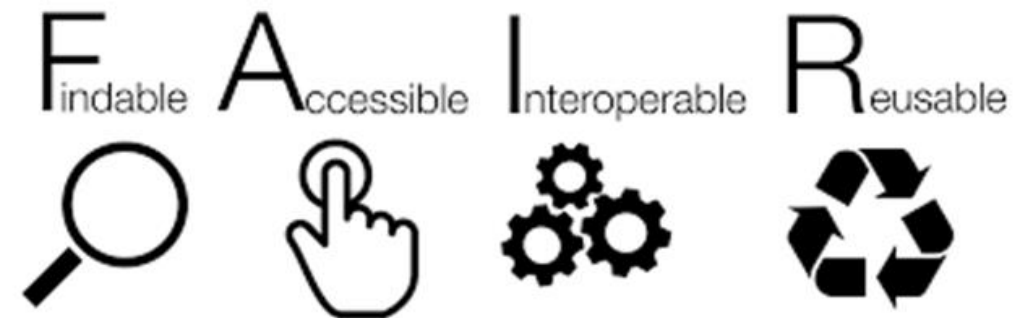


R
Reusable

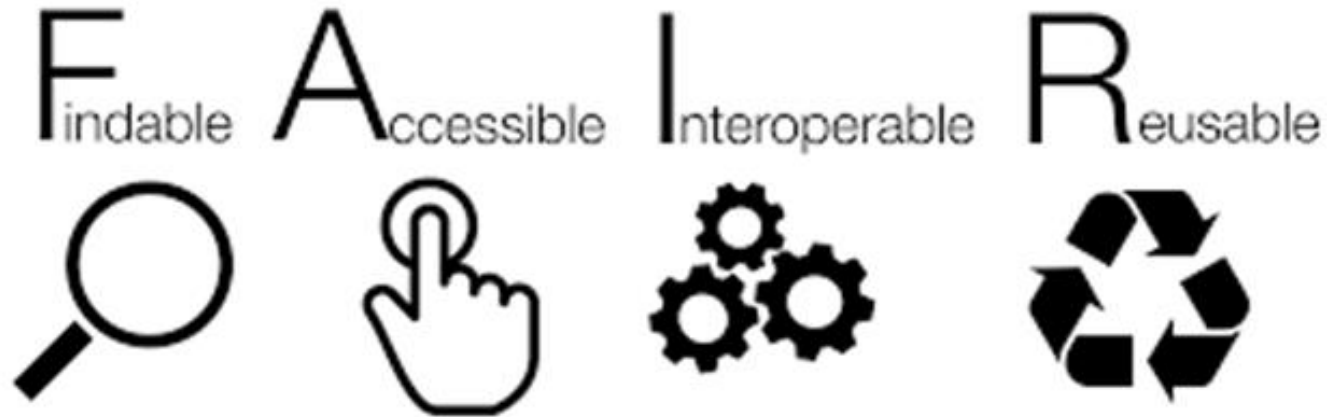


The FAIR Data Principles

- Drafted at the [Lorenz Centre Workshop in 2014](#)
- Published in [Scientific Data \(Nature\)](#) article in 2016
- Recognised by [NHMRC](#), European Commission, NIH, [OECD](#), Nature, PLoS, [NCRIS](#) and many others
- Making data usable by humans and machines
- FAIR is not necessarily Open
- Technology agnostic
- Discipline independent
- Apply to both data and metadata



Why be FAIR?



Open (as possible) and accessible research inputs and outputs drive research impact, quality, collaboration, translation and innovation

“data generated, created, captured or stored by NCRIS funded projects will be made available to the wider research community based on the F.A.I.R. principles, appropriately implemented for individual research communities.” (NCRIS Guidelines)

<https://ardc.edu.au/resources/working-with-data/fair-data>

Findable



Metadata and data should be findable by humans and computers

To be Findable:

- F1. (Meta)data are assigned a globally unique and persistent identifier
 - F2. Data are described with rich metadata (defined by R1 below)
 - F3. Metadata clearly and explicitly include the identifier of the data they describe
 - F4. (Meta)data are registered or indexed in a searchable resource
-
- Data discovery portals, local and international
 - Research Data Australia: <https://researchdata.edu.au>
 - Registry of Research Data Repositories: <https://www.re3data.org>

Accessible



Once found, users need to know how the data can be accessed

To be Accessible:

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol**
 - A1.1 The protocol is open, free, and universally implementable**
 - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary**
- A2. Metadata are accessible, even when the data are no longer available**
 - **As Open as possible, as restricted as necessary**
 - **Access ideally using an open documented API interface (machine executable)**

Interoperable



Data needs to work with applications or workflows for analysis, storage and processing

To be Interoperable:

1. (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation.
 2. (Meta)data use vocabularies that follow FAIR principles
 3. (Meta)data include qualified references to other (meta)data
- Vocabularies: Research Vocabularies Australia
<https://ardc.edu.au/services/research-vocabularies-australia/>

Reusable



Goal of FAIR is to optimise data reuse, while maintaining the initial richness of the data

To be Reusable:

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes**
- R1.1 (Meta)data are released with a clear and accessible data usage license**
- R1.2 (Meta)data are associated with detailed provenance**
- R1.3 (Meta)data meet domain-relevant community guidelines**

FAIR Assessment tool for Projects

		Answer options: Increasingly FAIR -->				
FINDABLE						
Q1	Does the dataset have any identifiers assigned?	No identifier	Local identifier	Web address (URL)	Globally unique, citable and persistent identifier (e.g. DOI, PURL, or Handle)	
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
Q3	Is the data described by a metadata record?	The data is not described	Brief title and description	Brief title and description, and multiple other fields filled out, albeit briefly.	Comprehensively (a min metadata template will be provided) using a formal machine-readable metadata schema.	
Q4	What type of repository or registry is the metadata record in?	The data is not described in any registry or repository	Local institutional repository	Domain-specific repository	Generalist public repository	Data is in one place but discoverable through several places (i.e. other registries, RDA, Google Data Search)
ACCESSIBLE						
Q5	How accessible is the data? Note: The access method(s) must be explicitly stated in the metadata record, e.g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	Embargoed access after a specified date; or A deidentified version of the data is publicly accessible	Fully accessible public, or to persons who meet and follow explicitly stated conditions and processes, e.g. ethics approval for sensitive data
Q6	Is the data available online without requiring specialised protocols or tools once access has been approved?	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. OpenAPI/Swagger/informal API)	Standard web service API (e.g. OGC)
Q7	Does the repository/registry agree to maintain the persistence of the metadata record, even if the data product is no longer available?	No (or not applicable, if no metadata record exists)	Unsure	Yes		
INTEROPERABLE						
Q8	Are the data available in (an) open (file) format(s)?	Data are mostly available only in a proprietary format	Data are available in an open format	Data are available in an open, documented, widely-used standard format (i.e. NetCDF, CSV, JSON, XML, etc)		
Q9	Are the data machine readable?	The data are unstructured	The data are structured and machine-readable (i.e. csv, JSON, XML, RDF, database files, etc)			
Q10	What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?	Data elements are not described (i.e. fields or objects are labelled with codes or not at all)	Data elements are described (so that a human user can correctly interpret the data), but no standards have been used in the description	Recognised standards have been used in the description of data elements, but no published vocabularies with resolvable URIs are used	Published vocabularies using resolvable identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.	Published vocabularies using persistent resolvable identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.
Q11	How is the relationship to other data and resources (e.g. related datasets, services, publications, etc) described in the metadata, to provide context around the data?	There are no links to other metadata or data	The metadata record includes URI links to related metadata, data and definitions	Qualified links to other resources are recorded in a machine readable format, e.g. a linked data format such as RDF		
REUSABLE						
Q12	Which of the following best describes the license (usage rights) attached to the data?	No license is applied	Non-standard license applied, without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Non-standard license applied, WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record
Q13	How much provenance information has been captured to facilitate data reuse? i.e. project objectives, data generation/collection (including from external sources) and processing workflows.	No provenance information is recorded	Partially recorded	Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively recorded in a machine readable format (i.e. in metadata record's schema or PROV, or in RDF, JSON, NetCDF, XML, etc)	

Good data practice

- Avoid: Embargos
- Avoid: Closed datasets
- Data description
- Links, links, links
- Versioning
- Collections
- Multiple records
- Citing data

Embargos

(avoid where possible)

- On Files only
 - The metadata will be made available.
 - Flags to the research community this content will be available soon.
 - Reason will be publicly displayed
 - Can be permanent if required.
- On the entire content
 - Nothing is made publicly available.
 - Must have an end date.
- Embargos expire automatically and content becomes available
 - Can be extended if needed.

🕒 Apply embargo



Embargo period

The embargoed content will become public on the date you select.

Select period... ▼

or

until specific date... 📅

Embargo type

On files only ▼

Why are the files under embargo?

Useful for people viewing the public metadata record.

Title

File(s) under embargo

Reason

Write the reason why the files are under embargo...

Closed Datasets

Not recommended



Metadata record only

Reason:

- ROADS is designed to be an open data sharing platform
 - Choose Metadata record only in place of uploading content
 - The description of your data becomes critical
 - Need to describe exactly what your data is as it won't be visible
- Licence: Reuse under agreement
- Primary Contact: Name and email
- Access Rights: Who and How

Licence [\(what's this?\)](#)

Reuse under agreement (Restrictive Licence Template) ▼

Primary contact

Sam Researcher (sam.researcher@flinders.edu.au)

Access Rights

Conditions apply to the use of the data. Contact the above ...

Closed Datasets

Not recommended



Journal of Clinical Epidemiology

Available online 30 May 2022

In Press, Journal Pre-proof ?



Original Article

Many researchers were not compliant with their published data sharing statement: mixed-methods study

Mirko Gabelica¹✉, Ružica Bojčić²✉, Livia Puljak³✉

A study published in 2022 found:

- 1,792 papers where authors indicated they were willing to share their data
- 1,670 (93%) authors either did not respond or declined to share their data.
- Only 122 (6.8%) authors provided the requested data

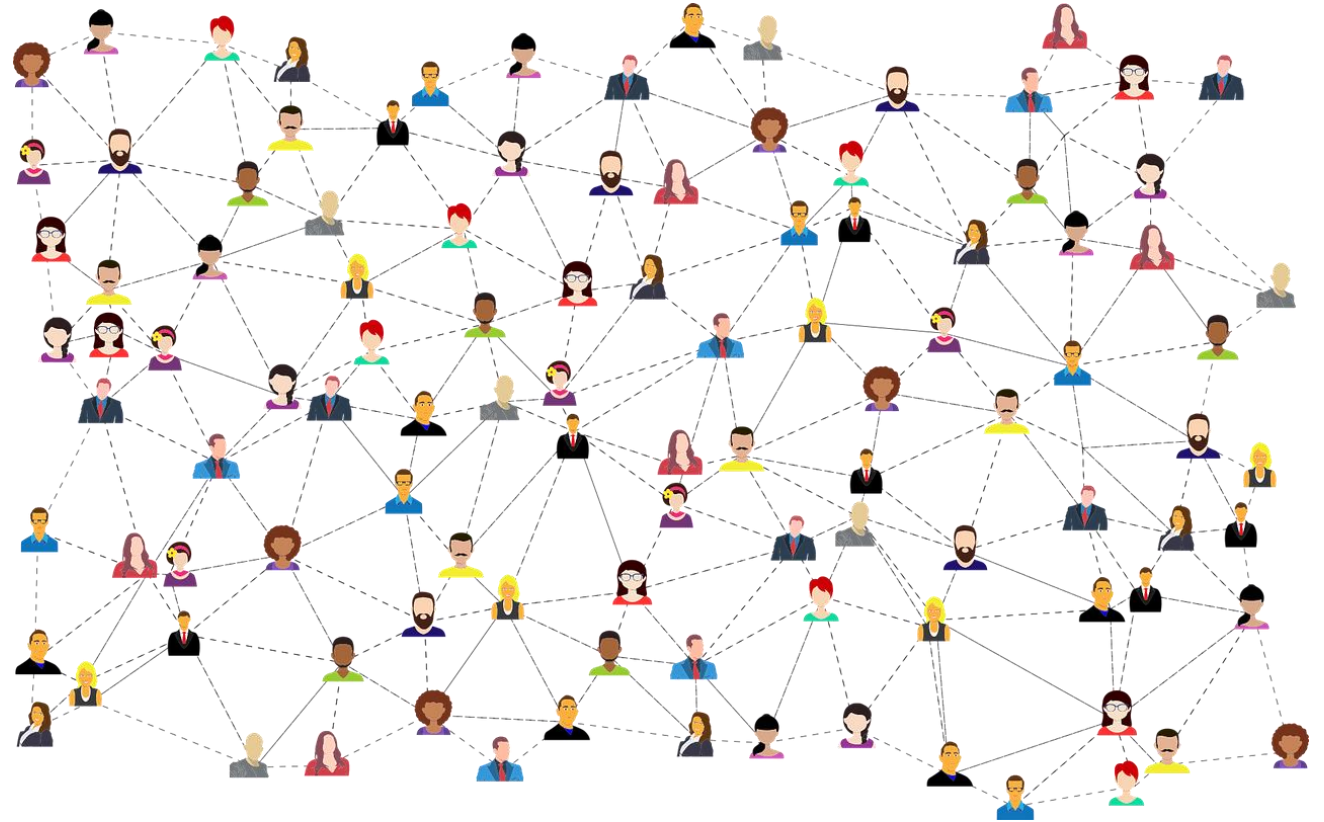


Fully described

- **Full and Rich description of your data**
 - Describe what your data is not why it's available
 - Don't use: *Dataset for XYZ journal article* (as either title or description)
 - Don't limit your data to only supporting your publication
 - Let other researchers know what they need to use it
 - Specialised software, equipment etc.
 - What standards your data uses
 - Consider creating a 'ReadMe.txt' file to accompany your data
 - Particularly with complex data or multiple files

Links, links, links

- **Authors**
- **Publications**
- **Other datasets**
- **Vocabularies**
- **Standards**



Versioning

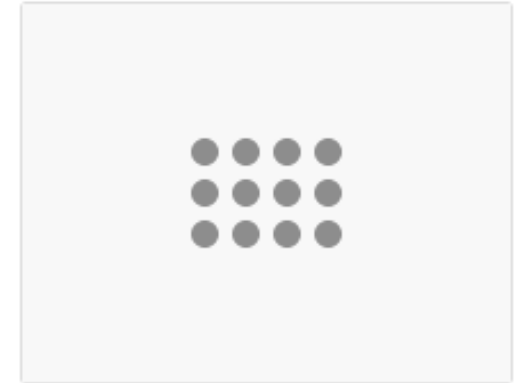
- A **new version** is automatically created in ROADS when changes are made to the dataset
 - Title
 - Authors
 - Files (added, removed, replaced)
 - Links (added, removed, modified)
- BOTH versions remain available
- The DOI will be incremented



<https://doi.org/10.25451/flinders.19770742.v3>

Collections

- Collections:
 - Enables you to create a *container* to hold individual datasets.
 - Collection does not directly hold data
 - Collection is publicly visible
 - Collection has its own DOI & suggested citation
 - Collection has its own metadata (eg. description)
- You can collect datasets from ROADS or Figshare.com
- No previous relationship needs to exist
 - They don't need to share authors etc.



COLLECTION:

Peter's map collection

Posted on 29.04.2022

Peter Mason

Multiple records

- Datasets with multiple files should be kept on a single record.
 - Don't create a separate record for each file
 - Maximum 500 files
 - Where a dataset shares the same authorship keep on a single record
- Datasets with different authorship should be on separate records
 - Use a collection to group together if required.



Citing data

Cite

- *Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.*
- ROADS , Figshare, and Research Data Australia all provide suggested citations to help you cite research data

Benchmark wind farm noise data set

Cite

Download all (723.18 MB)

Share

Embed

+ Collect

...

DataCite

Nguyen, Duc Phuc (2022): Benchmark wind farm noise data set. Flinders University. Dataset. <https://doi.org/10.25451/flinders.19618023.v1> Copy citation

<https://doi.org/10.25451/flinders.19618023.v1> Copy DOI



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.

Add your project name

Enabled by NCRIS



ARDC Services

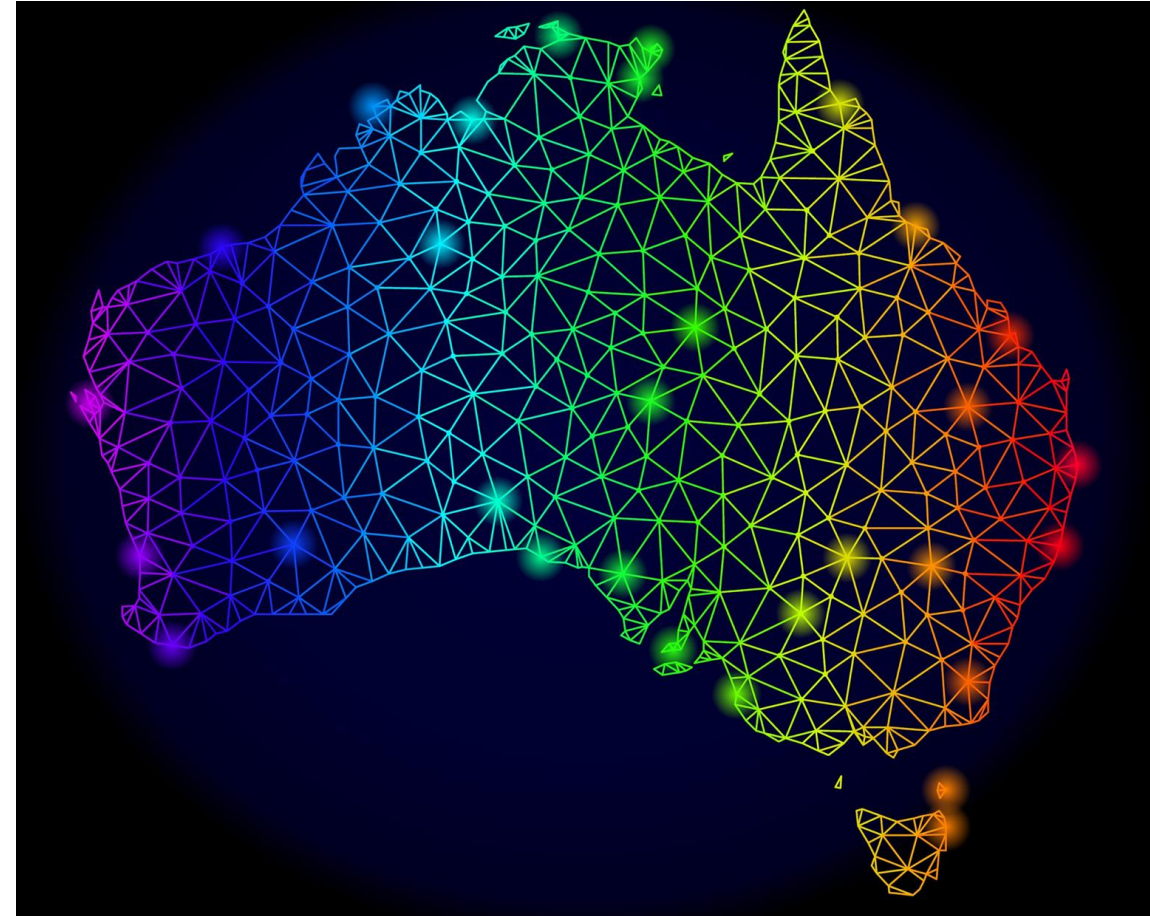
- Research Data Australia
- Research Vocabularies Australia
- NeCTAR Research Cloud
- Research Platforms
- Identifier Services

Add your project name

Research Data Australia

- ARDC's Data Discovery Service
- Find, access and reuse data from >100 Australian research organisations, government agencies & cultural institutions
- Provides metadata and links to data held by publishing partners
- Tools and support for submitting metadata records to RDA

<https://researchdata.edu.au>



Research Data Australia

178,373 Datasets, 100 contributors



Humanities and Social Sciences



Business, Economics and Law



Medical and Health Sciences



Engineering, Computing and Technology



Built Environment and Design



Biological Sciences



Agricultural and Veterinary Sciences



Environmental Sciences



Earth Sciences



Physical, Chemical and Mathematical Sciences

Research Vocabularies Australia



Get Involved



Publish a vocabulary

Upload, describe and publish your vocabularies to Research Vocabularies Australia



Use a vocabulary

Understand how you can utilise Research Vocabulary Australia vocabularies



Explore widgetable vocabularies

Discover vocabularies that can be readily used in your system using our vocabulary widget



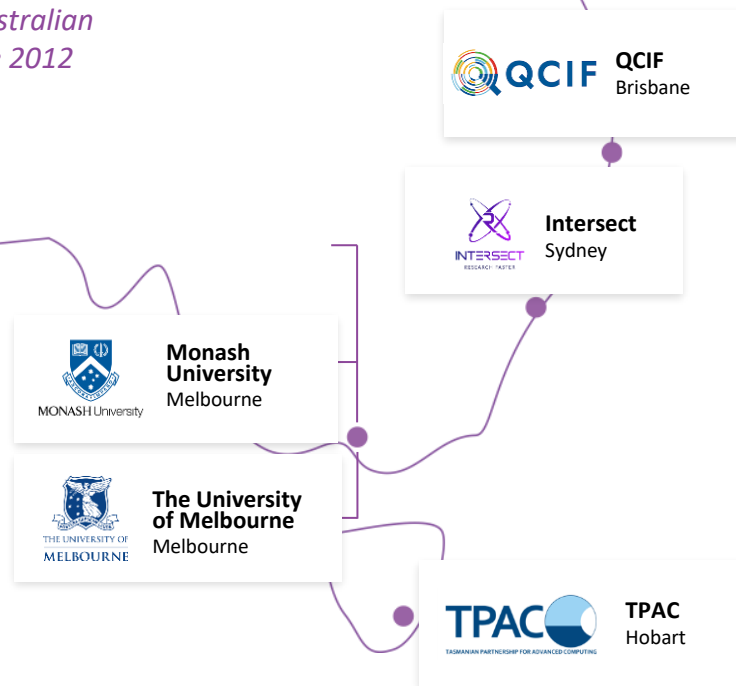
Provide feedback

Help Research Vocabularies Australia to grow into a comprehensive vocabulary portal

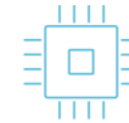
<https://vocabs.ardc.edu.au>

Australia's national research cloud

Advancing Australian research since 2012



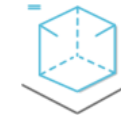
8,000
Virtual Machines



20,000
Physical CPU Cores



60,000+
Virtual CPU Cores



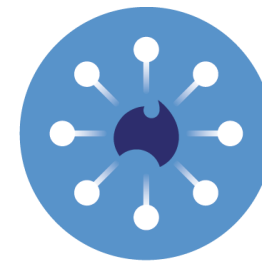
1.8TB
Object Storage



5PB
File Storage

Research Platforms

Domain-oriented online environments that draw together research data, models, analysis tools and workflows to support collaborative research across institutional and discipline boundaries.



Virtual Geophysics
Laboratory (VGL)



<https://ardc.edu.au/services/research-platforms/>

ARDC Identifier Services

- Services to create and manage Persistent Identifiers (PIDs) for research data, samples, files, document, projects and other digital objects
- Contributing to national and international standards and practice for PIDs, inc samples and instruments



Digital Object Identifier services



Handle service



International Geo Sample
Number



Research Activity Identifier

<https://ardc.edu.au/services/identifier/>

Persistent Identifiers (PIDs)

Linking the elements of the research community

PIDs for people (researchers) include ISNIs and ORCID iDs



PIDs for places (research organizations) include GRID and ROR



PIDs for things (research outputs/inputs like grants, reviews, preprints, projects, etc.) include Crossref and DataCite DOIs, IGSNs, RAiDs, and more



Data Infrastructure Co-investment Programs

- Platforms
- National Data Assets
- Translational Research Data Challenges
- Data Retention

Add your project name

Institutional Underpinnings

Catalyse system-wide good institutional practice for research data.

Participating 25 Universities will collaboratively develop, test and implement a national framework for institutional research data management.

RISE framework
Research Infrastructure Self Assessment



<https://ardc.edu.au/news/universities-unite-for-consistency-in-research-data-management/>

Advice and Expertise



Australian Research Data Commons

Vocabularies

Metadata

Virtualisation and
Containerisation

Data Architecture

FAIR

Sensitive Data

International
Linkages

AI and ML

Research Software

Data
Governance

Research Data
Management

eResearch Skills
Development

Communities

Platforms for information exchange, problem solving, peer support and development of best practice.

Including:

- Sensitive Data Community
- Machine Learning Community of Practice for Australia (ML4AU)
- Identifiers for Instruments in Australia
- Australian Research Data Management Community
- Data Management Plans Interest Group
- Australian Research Containers Orchestration Service (ARCOS)
- National Training Registry Working Group (Digital Research Skills Australasia portal)
- Research Data Roundtables (in each state)
- Tech Talks

<https://ardc.edu.au/resources/communities-of-practice/>

Skills Development

Facilitate eResearch and digital skills development

- eResearch and Data Skills Summits
- The Carpentries consortium
- DReSA: <https://dresa.org.au>
Digital Research Skills Australasia
Training events, materials, providers
and trainers



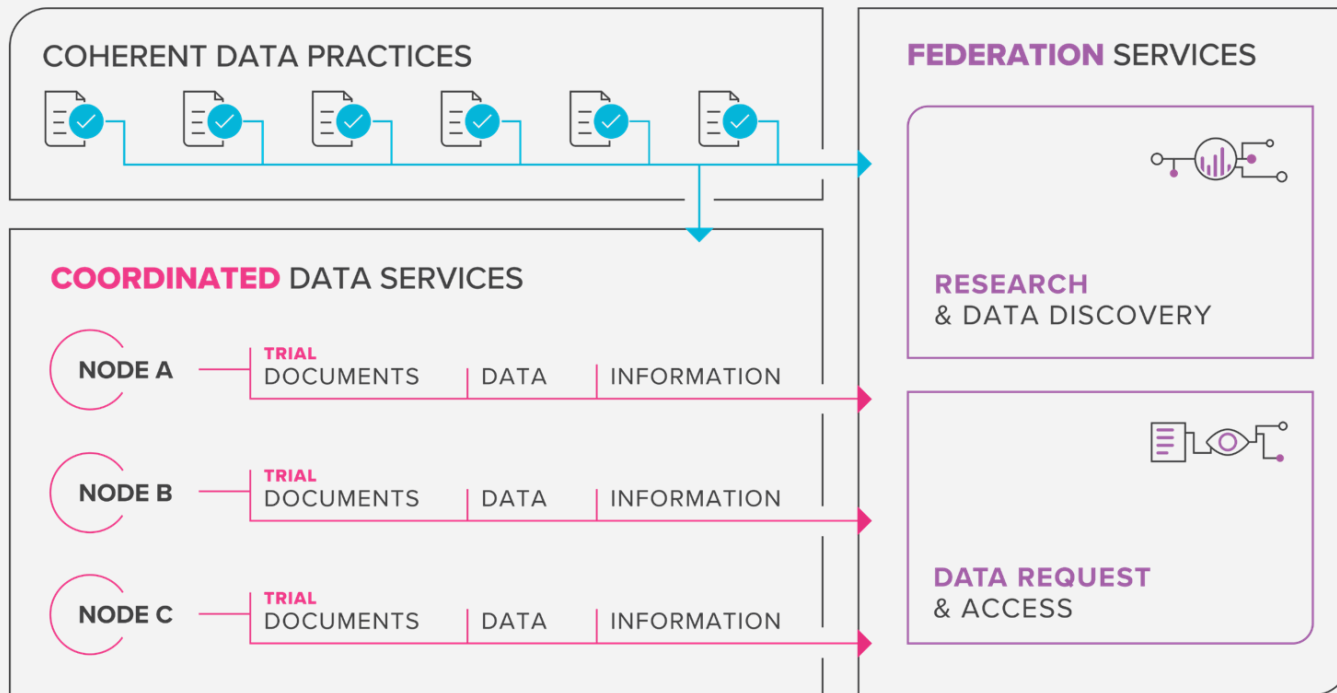


HeSANDA - Health Studies Australian National Data Asset

A strategic partnership with the Australian health research community to build a distributed national data asset from the outputs of health studies to support health data sharing & secondary use.

- Co-designed by the health research community
- Initial focus on **investigator-initiated clinical trials** as a proof-of-concept
- External Advisory Committee of peak national bodies
- \$3M co-investment over 2 years

Consultation Phase outputs



- Infrastructure model:
 - Coherent practices
 - Coordinated services
 - Federated systems
- Research community & stakeholder guiding principles



Subscribe to the
ARDC CONNECT
newsletter

THANK YOU



ardc.edu.au



contact@ardc.edu.au



+61 3 9902 0585



[@ARDC_AU](https://twitter.com/ARDC_AU)



[Australian-Research-Data-Commons](https://www.linkedin.com/company/Australian-Research-Data-Commons)

Repository of Open Access DataSets



CONNECT

 /flindersuniversity

 @flinders

 @flindersuniversity

 school/flinders-university/

 @flindersuniversity