

**MYScope**  
MICROSCOPY TRAINING

# Research Data Management

# Train for advanced research

## Welcome

MyScope was developed by Microscopy Australia to provide an online learning environment for those who want to learn about microscopy. The platform provides insights into the fundamental science behind different microscopes, explores what can and cannot be measured by different systems and provides a realistic operating experience on high end microscopes.

We sincerely hope you find our website: [www.myscope.training](http://www.myscope.training) an enjoyable environment. In there you can explore the microscopy space and leave ready to undertake your own exciting experiments.

### Attribution-Share Alike

Unless otherwise noted, content on this site is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.

We cannot revoke these freedoms as long as you follow the license terms.



### **What is research data?**

### **Introduction to research data management**

### **Maintaining quality and integrity**

### **Benefits of managing research data**

### **FAIR data principles**

What is FAIR data?

The FAIR principles explained

Persistent identifiers

Are FAIR data and open data the same?

### **The CARE principles**

What are CARE principles?

What are Indigenous data?

The CARE principles explained

### **Be FAIR and CARE**

FAIR and CARE are complementary

### **Global guidelines**

### **Research data management plans**

### **Credits**

# What is research data?

There is no authoritative definition of research data. It may depend on the discipline and the context, for example, the nature of the project and experiment. As a result, it is advised to check the definition of research data by the relevant institution.

**The Australian Research Data Commons (ARDC)** defines research data as: "Research data, from the point of view of the institution with a responsibility for managing the data includes:

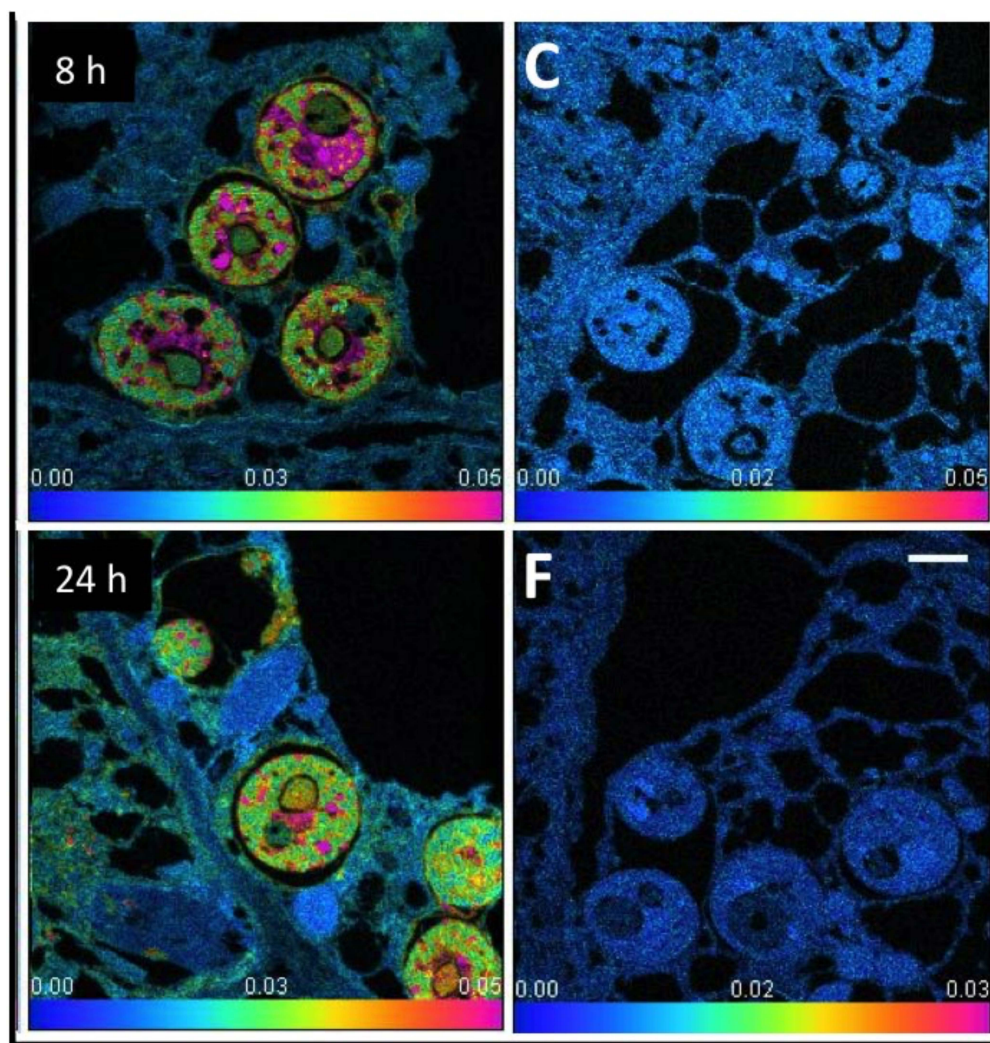
- all data which is created by researchers in the course of their work, and for which the institution has a curatorial responsibility for at least as long as the [Australian Code for the Responsible Conduct of Research] and relevant archives/record keeping acts require, and
- third-party data which may have originated within the institution or come from elsewhere."

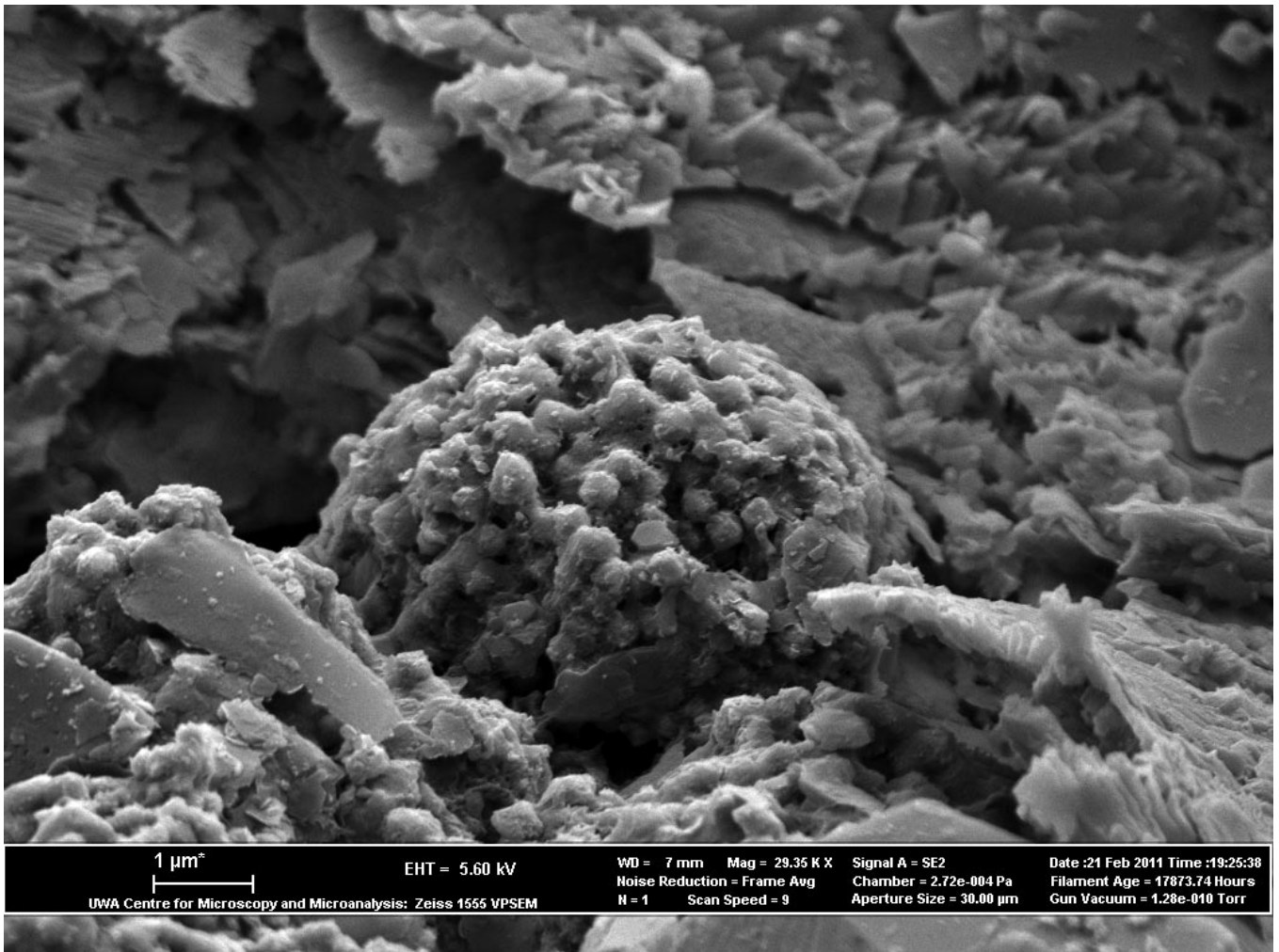
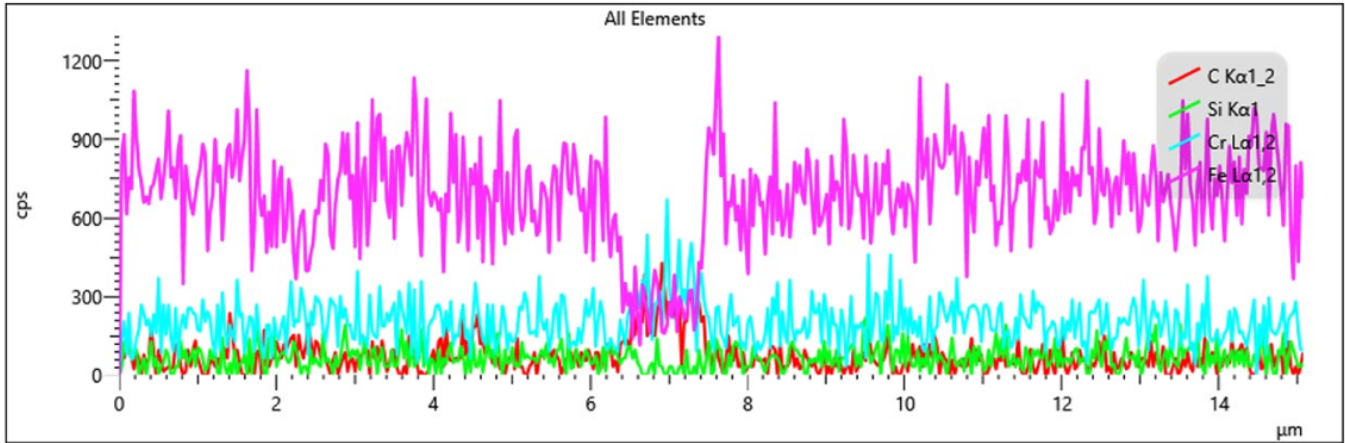
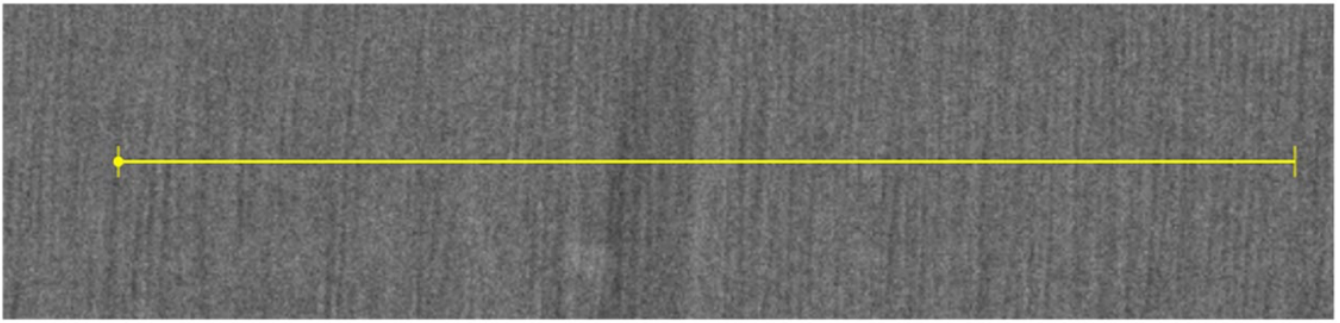
You could put this simply as, research data means information collected or recorded as part of a research project regardless of form or media.

It can include:

- all notebooks
- survey responses
- software and code
- measurements from laboratory or field equipment such as EDS spectra or hygrothermograph charts and 3D data points
- images such as digital photographs, films, scans, or autoradiograms
- audio recordings
- physical samples
- raw, analysed, processed or reduced data.

Some examples of the wide range of microscopy-related research data collected from around Microscopy Australia are below and on the following pages.







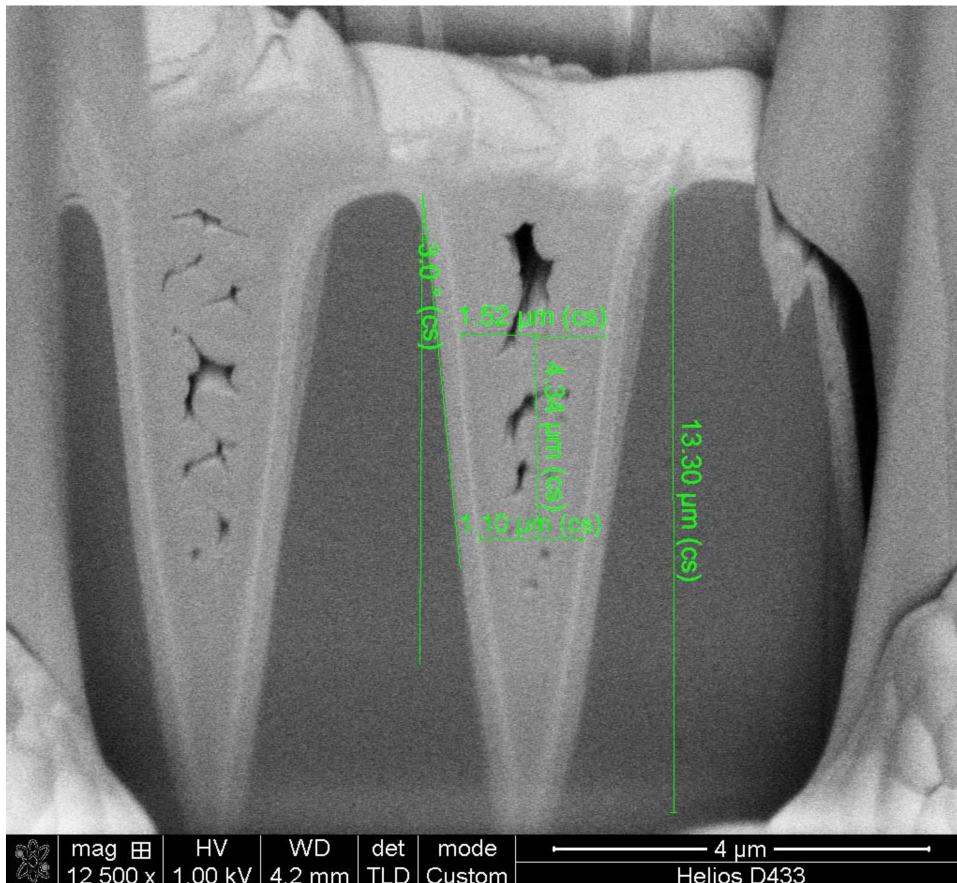
# Introduction to research data management

Research data management has become an integral part of research. Although it is required by many scholarly publishers and research funders, good data management is the best way to maximise impact and value of research through the reuse of research data. Suitable data storage and management protects your data from loss and ensures compliance with any rules and requirements associated with the research project (e.g. ethics and confidentiality).

WHERE IS THE DATA?

UMMMM...

Research data management covers the planning, collection, organisation, storage, preservation, documentation and sharing of research data over the lifetime of a research project, after the end of the project and beyond the publication of research findings. It thus encompasses both the everyday and long-term management of research data.



## Maintaining quality and integrity

---

Research data management is essential for maintaining the quality and the integrity of research data, and ensuring compliance with legal, regulatory, ethical and funding requirements.

WHAT EXACTLY WERE THE  
SETTINGS YOU USED TO  
PROCESS THIS DATA?

ER, I THINK I  
USED IMAGEJ...

**This short video from NYU Health Sciences Library** shows that when research data aren't managed properly, reusing and sharing data is difficult or impossible.



# Benefits of managing research data

---

Research data management has multiple and important benefits:

- It can enhance research impact and value by maximising the visibility of research data and promoting transparency in research;
- It improves accessibility to research data by ensuring that its quality and integrity are maintained during and beyond the lifetime of the research project. It is easier to replicate and validate findings from an experiment when research data are well managed;
- It saves times and optimises the use of storage resources;
- It ensures responsible use and reuse of research data, including by attributing appropriate authorship (citations) and by following confidentiality obligations, legislation, privacy or ethical principles or other guidelines associated with the research data;
- It can safeguard and future-proof research data by establishing appropriate, safe and secure short-term and long-term storage of data (including depositing data to a repository for long-term archiving and access);
- It minimises the risks of research data loss and corruption, as well as breaches in confidentiality or ethical requirements associated with the research data.

**WHERE ARE THE REST  
OF THE IMAGES FROM  
THAT EXPERIMENT?**

**WE DUMPED THEM BECAUSE  
THE SERVER WAS FULL.  
ANYWAY, WE'VE ALREADY  
PUBLISHED THE PAPER.**



# FAIR data principles - What is FAIR data?

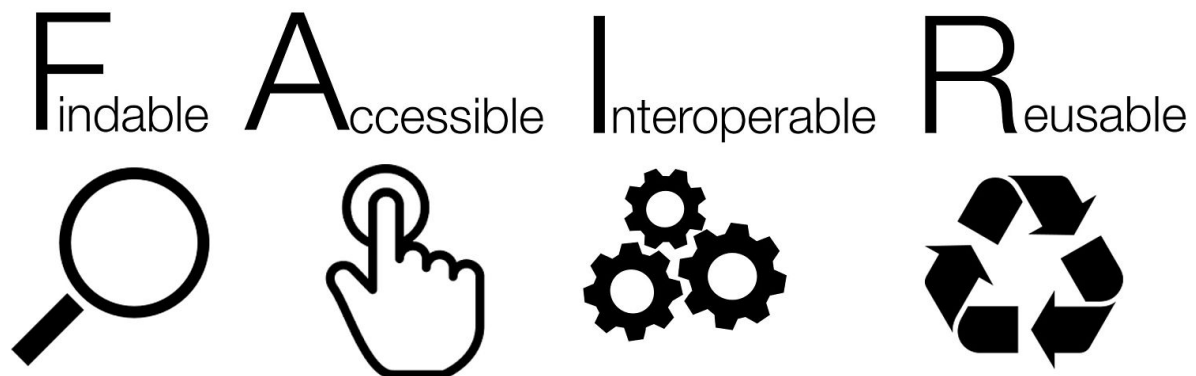
---

Good data management has become increasingly recognised as integral to research. However, it is not a goal in itself. Instead, it favours the discovery, sharing and reuse of research data after the publication of the studies they underpin. The FAIR Guiding Principles for scientific data management and stewardship promote practices that enable research data to be Findable, Accessible, Interoperable and Reusable for both machines and humans. They enhance the exposure of research data thereby increasing the impact of research.

Machine-actionability is especially central to FAIR data. It means that computers can find, access, interoperate and reuse data with limited or no human intervention. In this era of big data, when the volumes, level of complexity and velocity of data keep increasing, machine-actionable data will play an important role.

The FAIR principles build on earlier efforts that have aimed to facilitate the reuse and sharing of data. A major change associated with the FAIR principles is that raw research data have become usable, analysable, citable scholarly outputs. This has had consequences for research organisations, publishers, funders and data repositories because for research data to be FAIR, metadata (information about data) and supporting infrastructure (for data storage and sharing) must also be compatible with the FAIR principles. So the FAIR principles require standardised approaches to how research data can be:

- described and identified using metadata (information about data) and persistent identifiers (PIDs) following community-agreed conventions;
- stored in a way that maximises discoverability, reuse and sharing of data while keeping a balance between sustainability of storage and permanence of data.



# The FAIR principles explained

---

## Findable data

Reusing data implies that, first, it is possible to find them. It should be easy for both humans and computers to find data and the associated metadata. Data and metadata are assigned a globally unique and persistent identifier that is machine-readable. Metadata and PIDs are essential for automatic and reliable discovery of datasets and services. Metadata must clearly and explicitly include the identifier of the data it describes. Using community-endorsed, general or discipline-specific, standards for data and metadata (such as data models and metadata schemas) is recommended.

## Accessible data

Once data are found, it must be clear how to access them. This can include information on whether authentication or authorisation is required prior to accessing the data. Data and metadata should be retrievable by their identifier using a standardised and open communications protocol that is open, free and universally implementable. It must allow for authentication and authorisation procedures for restrictions if necessary. Importantly, metadata should remain available and accessible even when the data are no longer available.

## Interoperable data

It should be possible to integrate the data with other data, applications and workflows. This means that the format of the data should be open and interpretable by various tools for analysis, storage and processing (which proprietary software applications generally don't allow). They must use a formal, accessible, shared and broadly applicable language for knowledge representation. Interoperability applies both at the data and metadata level. Common formats and standards for data and metadata, as well as controlled vocabularies (that follow FAIR principles), are essential to ensure data and metadata interoperability.

## Reusable data

The ultimate goal of FAIR is to optimise data reuse. To achieve this, data and metadata should be well-described, have a clear and accessible license for data usage, include detailed information on provenance, and meet domain-relevant community standards. This will ensure repeatability of experiments and reproducibility of results.

# Persistent identifiers

---

Persistent identifiers (PIDs) are long-lasting, globally unique, digital references to objects (for example, publications, datasets, funding grants), people or organisations. The figure below shows examples of established PIDs that can be used. Additional PIDs are being developed so their levels of awareness and adoption can vary across communities

PIDs are a critical piece of infrastructure that underpins the FAIR principles:

- They link entities in the research ecosystem (for example, researchers, organisations, articles, datasets, samples, software, funders, grants, instruments) so they can favour innovation and improve efficiency;
- They enable tracking of research impact by enabling research provenance and attribution;
- They support research integrity and reproducibility of results by identifying research inputs (for example, samples, instruments, software) and research outputs (for example, datasets, articles).

Overall, the integration of PIDs into information systems used in research organisations (for example, laboratory information management systems, image repositories, research management systems, finances, human resources) facilitates information exchange between those systems across and between organisations. It also eliminates the need to re-enter information manually multiple times into multiple systems, for example about a grant, a publication or a person (such as their publications), which can lead to significant cost savings for research organisations.

## Researchers



ORCID ID  
(Open Researcher and  
Contributor ID)  
*orcid.org*

## Objects (physical, digital, abstract)



DOI  
(Digital Object Identifier)  
*doi.org*

## Research organisations



ROR ID  
(Research Organization Registry)  
*ror.org*

## Research projects



RAiD  
(Research Activity Identifier)  
*raid.org*

## Key resources in biomedical research



RRID  
(Research Resource Identifier)  
*rrids.org*

## Physical samples



IGSN ID  
(International Generic Sample  
Number Organisation)  
*igsn.org*

*Examples of PIDs used in research*

## Are FAIR data and open data the same?

---

No. FAIR data doesn't necessarily mean "open" or "free" data. Data can be restricted and still be FAIR. The important aspect of "accessible data" is that the conditions to access the data are explained in a way that both a human and machine can understand. FAIR data can be as open as possible and as closed as necessary. Both private and protected data can be FAIR. For example, accessible data can mean that users register an account to access a repository.

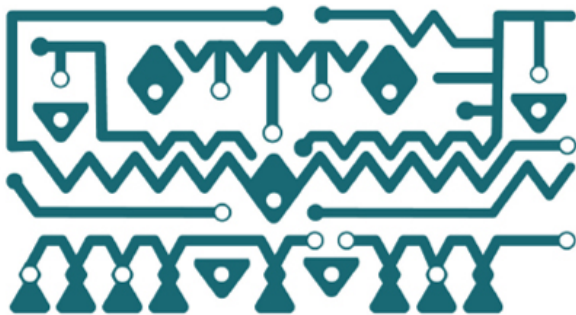
## The CARE principles - What are CARE principles?

---

The “CARE Principles for Indigenous Data Governance” were defined following the FAIR principles. They focus on Collective benefit, Authority to control, Responsibility and Ethics in the use and sharing of Indigenous data and they highlight Indigenous Peoples’ rights and interests in data governance [3]. FAIR and CARE principles are complementary.

Despite the current movement towards FAIR data, open data and open science to promote accessibility and sharing of research data, Indigenous Peoples’ rights and interests over Indigenous data have often been overlooked or not explicitly considered. For example, the emphasis on greater data sharing needs to take into account the fact that Indigenous Peoples are also asserting greater control over the application and use of Indigenous data and Indigenous Knowledge for collective benefit.

The CARE Principles for Indigenous Data Governance (Collective benefit, Authority to control, Responsibility and Ethics) were developed by the Global Indigenous Data Alliance (GIDA) (Figure below). They provide guidance to data producers, users, managers and publishers on the inclusion of Indigenous Peoples in data governance and management for a better discovery, access, use, reuse and attribution of Indigenous data. The CARE principles build upon earlier work by the Te Mana Raraunga Māori Data Sovereignty Network in New Zealand, the US Indigenous Data Sovereignty Network, the Maiam nayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Collective in Australia and numerous other Indigenous Peoples, nations, and communities.



## CARE Principles for Indigenous Data Governance

Source: <https://www.gida-global.org/care>

## What are Indigenous data?

---

They are data, information and knowledge, in any format, that impact or concern Indigenous Peoples, Nations and Communities at the collective or individual levels. They include data about:

- Individuals: for example, administrative, legal, health, social, commercial, corporate, services;
- Collectives: for example, traditional and cultural information, languages knowledge systems, ancestral and clan knowledges;
- Resources and environments: for example, land, water, geology, titles, air, soil, sacred ecosystems, territories, plants, animals.

# The CARE principles explained

---

## **Collective benefit**

Data ecosystems should be designed and function in ways that enable Indigenous Peoples to derive benefit and equitable outcomes from Indigenous data (such as inclusive development, innovation, improved governance and citizen engagement) through Indigenous nation and community use and reuse of data, use of data for policy decisions and evaluation of services, and creation and use of data that reflect community values.

## **Authority to control**

Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data.

## **Responsibility**

Those working with Indigenous data have a responsibility to build relationships based on respect, reciprocity, trust and mutual understanding, as defined by the Indigenous Peoples to whom those data relate. Information must be shared about how data are used to support Indigenous Peoples' self-determination and collective benefit through openly available and meaningful evidence. This includes enhancing data literacy and supporting the development of an Indigenous digital infrastructure, able to generate data grounded in the languages, worldviews and lived experiences of Indigenous Peoples.

## **Ethics**

Indigenous Peoples' rights and well-being should be the primary concern at all stages of the data life cycle and across the data ecosystem. Representation and justice, as well as consideration of potential future use (or harm) should be incorporated. This includes acknowledging the provenance and limitations or obligations for secondary use, especially in issues of consent.



## Be FAIR and CARE - FAIR and CARE are complementary

The CARE principles are complementary to the FAIR principles. While the FAIR principles focus on the characteristics of data, the CARE principles highlight and consider the people and purpose for which data exist and are used, reflecting the important role that data can play to advance Indigenous innovation and self-determination [6].

It is important to be aware that Indigenous data may be buried in larger datasets, or that samples or additional data used in experiments or projects may be associated with Indigenous data. They may be hard to find, mislabelled and controlled by others in a manner inconsistent with the FAIR and CARE principles. As a result, data could be subject to both CARE and FAIR. It is in general easier to apply the CARE principles to data that are already FAIR.



Source: <https://www.gida-global.org/care>

# Global guidelines

An increasing number of funding bodies and scholarly publishers have required or encouraged appropriate management of research data in Australia and other parts of the world. In Australia, the [Australian Code for the Responsible Conduct of Research, 2018](#) sets out basic requirements in research data management. Compliance with it is mandated by the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) to be eligible to their funding.

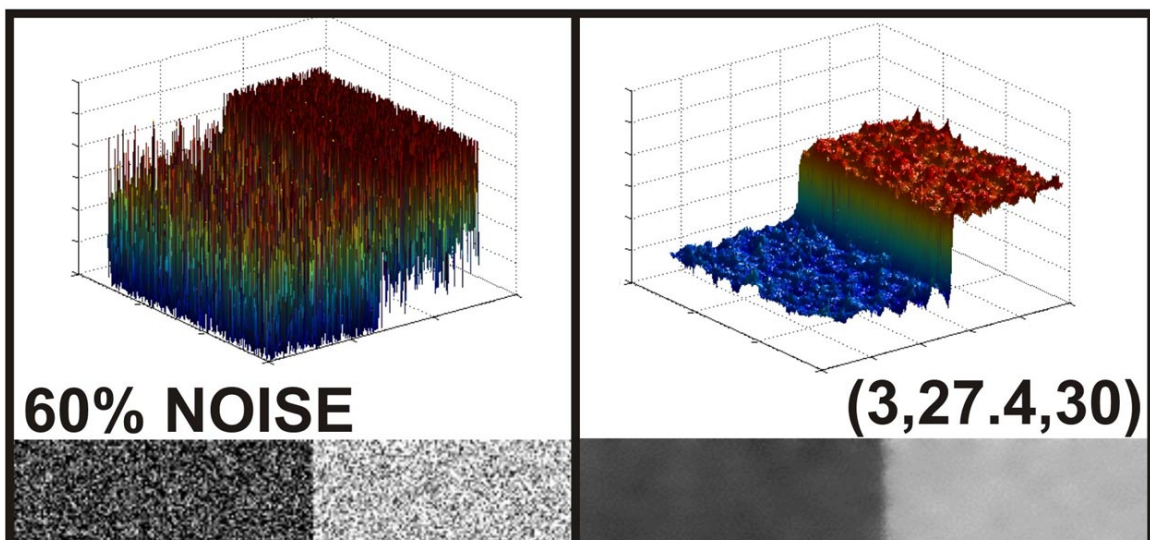
**WHAT DO THE NUMBERS  
IN COLUMN 'D' REFER TO?**

**MY EX-STUDENT MIGHT KNOW  
THAT BUT I'M NOT SURE WHERE  
SHE IS. I THINK SHE WENT  
TRAVELLING.**

Globally, the Organisation for Economic Co-operation and Development (OECD) encourages good practices in data management to ensure long-term sustainability of research data from public funding ([OECD Principles and Guidelines for Access to Research Data from Public Funding](#)).

In the European Union, the European Commission encourages or requires data management planning be included in projects participating in Horizon 2020, the funding programme for research and innovation run by the EU.

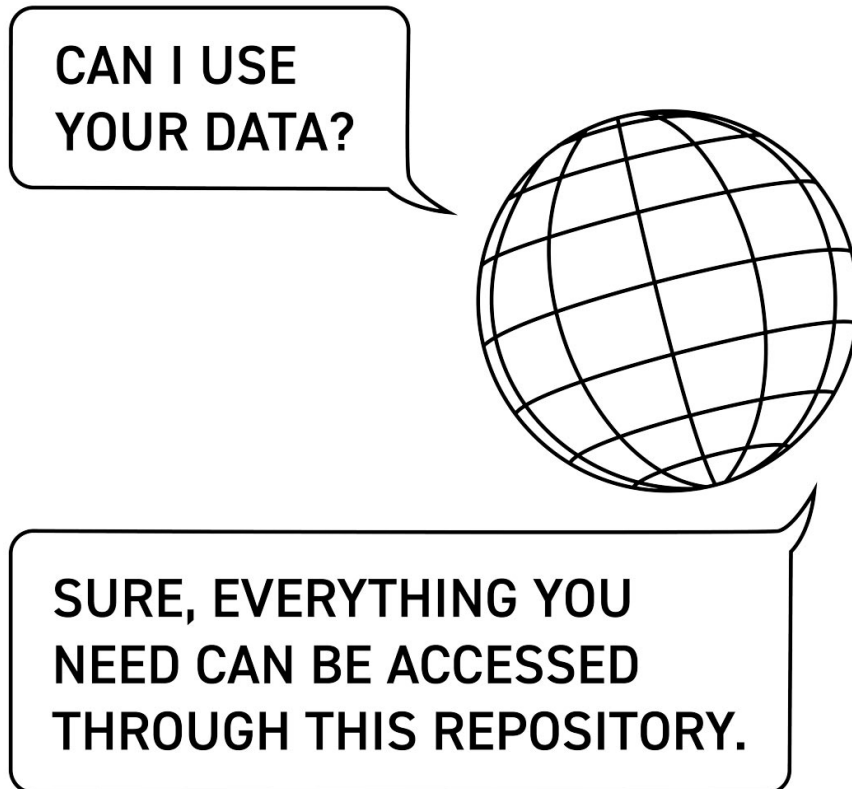
In the USA, many funding agencies such as the National Science Foundation (NSF) and the National Institutes of Health (NIH) require appropriate data management planning with every funding request.



# Research data management plans

---

Effective and successful research data management requires planning. A research data management plan records the formal plan for the storage, retention and disposal of, and access to, research data that are created or generated over the course of a research project.



There are a range of issues to consider when planning, including:

- the nature of the research data and the description of the file format(s) (e.g. will new data be generated or collected? Will some existing data be reused and/or transformed? Will the data be in proprietary/non-proprietary formats, converted to other formats and archived in shareable formats?). Whenever possible, open, shareable formats are preferred;
- the description and documentation of the data, in particular the metadata and persistent identifiers (PIDs) collected, the metadata/PIDs standard(s) and format(s) used and the tools and software (including versions) used to capture, read and use the data and the metadata/PIDs;
- the policies and procedures for the organisation, storage, archiving (including backups), deletion and accessibility of research data throughout and after the research project, including whether the data will be deposited to a public repository;
- the institutional, legal, regulatory, ethical, funding and contractual policies, arrangements and requirements applicable to the research data, including privacy, confidentiality and security issues;
- the roles and responsibilities for data management amongst the members of the research team, in particular in relation to data ownership, stewardship and/or custodianship and upon the departure of a team member from the research project or the institution.

Institutions often provide online tools, templates and various other resources to create research data management plans. They often offer training sessions for students and researchers to help them manage their data.

# Credits

---

Microscopy Australia acknowledges the huge input of time and expertise by the many staff members and associates who have contributed to the development of MyScope over the years.

For the RDM module we thank: David Poger.