



Waipapa  
Taumata Rau  
**University  
of Auckland**



# Research Data Management

Principles, policies and practices



July 2026



Laura Armstrong & Dr Sarah Hopkins  
Centre for eResearch

# Managing research data

## Learning outcomes for this workshop

At the end of this workshop, you will be able to:

- Describe the importance of **good research data management**
- Identify relevant **policies & processes** for research data management
- Apply **policies, data classification & best practices** to your research across the research data lifecycle



[researchdata@auckland.ac.nz](mailto:researchdata@auckland.ac.nz)

# Overview of this workshop

Duration	Section	Topics covered
10 mins	Introduction	Welcome and outline
20 mins	1. Data management planning	Introduction to RDM Key terms, policies and principles Data management plans
25 mins	2. Data creation and collection	Types and sources of data Approved systems and tools Legal and ethical requirements
5 mins	<b>BREAK</b>	
30 mins	3. Data analysis, storage, and security	Store, backup and organise data Processing and analysis Data security
15 mins	4. Data sharing and archiving	Data access and reuse Archive and retention
10 mins	Wrap up and discussion	



# Data Management Planning

- Introduction to research data management (RDM)
- Key terms, policies and principles in RDM
- Data management plans (DMPs)



# What is research data management?

**Planning and organising the collection, storage, documentation and sharing of research data during and after the project.**

## **Effective RDM:**

- **Aligns with RDM policy** and funder/publisher mandates
- Is well organised, quality-controlled, and well documented. This will enable you (and others) to understand the data in the future and **save you time**
- **Protects** the data against loss or misuse and ensures confidentiality
- Enables **verification** of results, facilitates collaboration and future use, and prevents duplication of effort
- Supports **reproducibility** and ensuring research integrity











**We manage data not just for ourselves, but so it can be shared and reused**



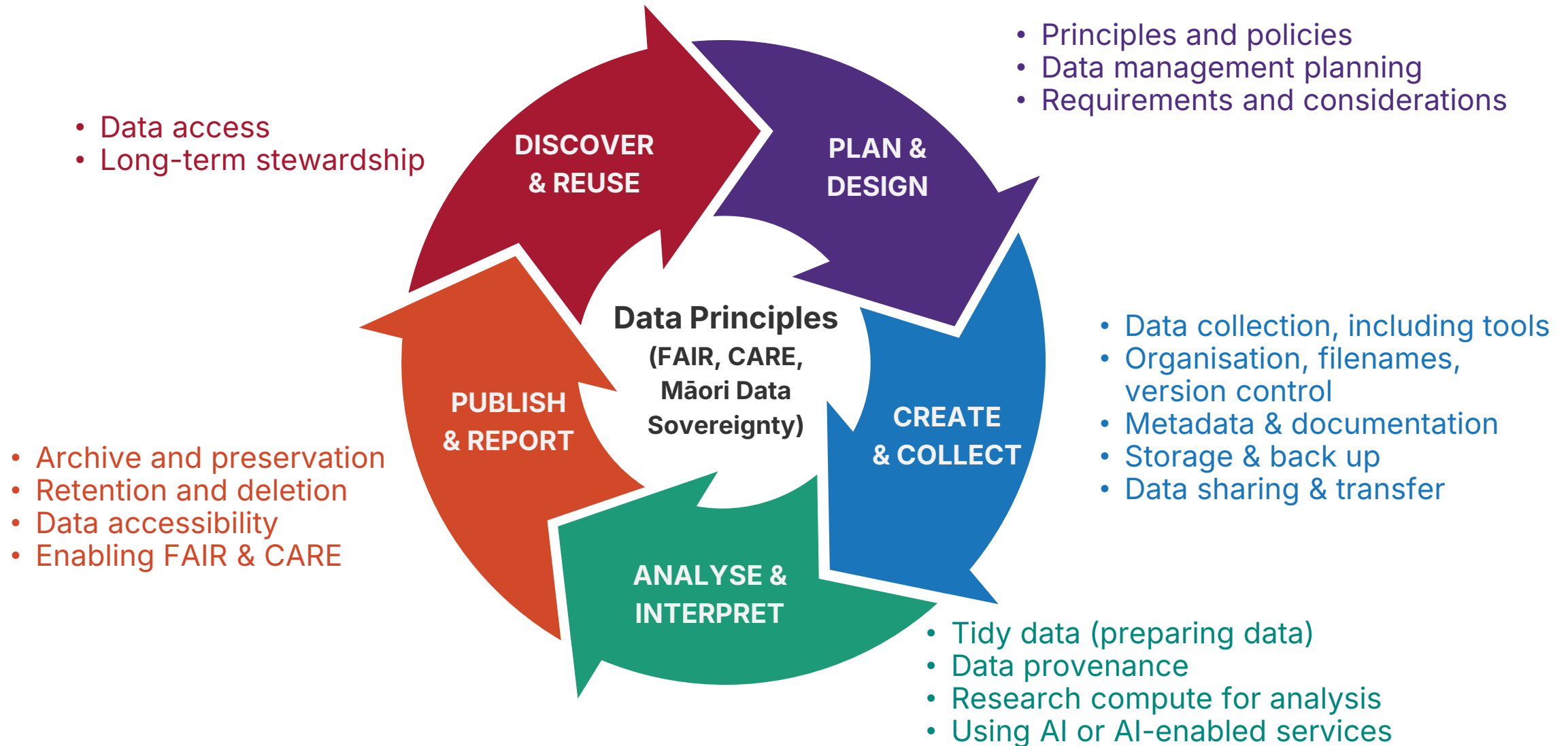
# What is research data?

Research data are the evidence that underpins the answer to a research question and can be used to validate findings regardless of their form (e.g., digital, print, physical).

## Numbers • Text • Images • Audio • Video • Spatial data • Software/code

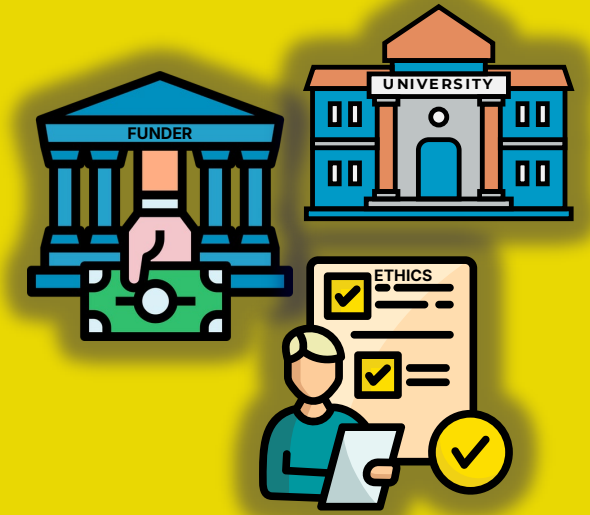
- |   |   |
|---|---|
|  Surveys, questionnaires               |  GIS, maps, spatial data                 |
|  Interview, focus group recordings     |  Laboratory measurements, sensor data    |
|  Transcripts, field notes & documents |  Biological samples, genetic data       |
|  Photographs & images                |  Databases, code & simulation outputs  |
|  Video recordings                    |  Clinical health data, medical imaging |

# Research data lifecycle





# Activities



Zoom poll

What is your biggest fear or concern about managing research data?



# Policies related to research data

## Funder policies

Concerned with obtaining the best outcomes for the research they fund.

**Examples:** • [MBIE Open Research policy](#)

## Government requirements

Legislation or guidelines that impact the handling or management of research data.

**Examples:** • [Trusted Research - Protective Security Requirements guidance](#)

## Publisher policies

Some publishers have policies or mandates related to data availability.

**Examples:** • [Data Availability | PLOS One](#)

## Professional codes of conduct

Concerned with the good conduct of research.

**Examples:** • [Research Charter for Aotearoa New Zealand](#) • [Royal Society Professional Code of Conduct](#)

## Institutional policies related to research data management →

# Research Data Management Policy

Sets out the responsibilities of the University community for the management of research data

Researchers are generally responsible for:

- Planning how data will be managed (DMP)
- Storing data securely
- Documenting data clearly (metadata)
- Sharing data where appropriate
- Retaining data for required periods

Supports FAIR and CARE data principles

 [Link to University of Auckland RDM Policy](#)

... / [Research and innovation](#) / [Research data management](#) / [Research Data Management Policy](#)

## Research Data Management Policy

### Application

This policy applies from the commencement date (1 July 2023) to all **research staff, students, supervisors** and other members of the **University community** that are involved in the management of **research data**.

### Purpose

To articulate the responsibilities of the University community for the management of research data. These responsibilities help to ensure that research data is managed in ways that are consistent with:

- international standards for **FAIR data** and open research that are increasingly required by funders, data providers and publishers
- the University's obligations under Te Tiriti o Waitangi and commitment to becoming a **Māori data sovereignty** organisation,
- the **CARE principles** for the governance of indigenous data, including **Pacific data**, and
- legal, ethical and protective security requirements for research data.

### Policy

#### University responsibilities

The University is responsible for:

1. Defining University roles and responsibilities for the management of research data.
2. Communicating the requirements of this policy and facilitating its adoption through the provision of training and guidance to researchers.
3. Providing the necessary infrastructure and services to enable researchers to meet their responsibilities for research data throughout the research data lifecycle.

#### Researcher responsibilities

**Researchers** are responsible for:



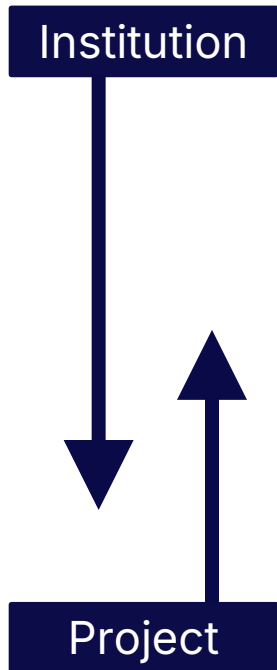
# Indigenous data sovereignty

**Indigenous Peoples have inherent rights and responsibilities to Indigenous data.**

- [CARE](#) principles for Indigenous data sovereignty  
Collective Benefit, Authority to Control, Responsibility, and Ethics
- [Māori Data Sovereignty principles](#)  
Rangatiratanga (Authority), Whakapapa (Relationships), Whanaungatanga (Obligations), Kotahitanga (Collective benefit), Manaakitanga (Reciprocity), Kaitiakitanga (Guardianship)
- [Pacific Data Sovereignty](#)



# Key roles in research data management



Role	Definition
<b>Data custodian</b> Infrastructure & security	Individual(s) or organisation responsible for the care of the data they hold (technical environment and security, e.g., storage, infrastructure, protection)
<b>Data steward</b> Governance & quality	The person(s) responsible for the quality, meaning and governance of data (standards, documentation, and appropriate use)
<b>Data manager</b> Day-to-day handling	The person(s) responsible for the day-to-day handling and organisation of data within a project (operational workflows)

# Data Management Planning

**Data Management Planning** is about preparing for data management across the research data lifecycle.

- Institutional, Ethics Committee or Funder requirement?
  - MBIE, NIH, Wellcome Trust
  - HDEC, University RDM Policy
- Risk-orientated and/or culture change approach
- Project specific
- Prompts conversations, captures decisions, clarifies roles and responsibilities and helps you to align with University policies and processes

About FAQs & Help News & Updates Contact

Data Management Planning tool

Get started

Welcome to the Data Management Planning tool (DMP tool)

About

The DMP tool helps you create research data management plans to meet University and funding requirements.

Learn More

FAQs & Help

Questions and answers about the DMP tool itself, and links for further information.

Learn More

News & updates

News about the DMP tool, including software updates or developments.

Learn More

Data Management & Planning

Good data management practices can help to:

- Ensure data remains complete, reliable, accurate, and retrievable
- Support data analysis and re-use
- Minimise data loss and duplication
- Enable tracking of data from collection to results
- Provide data security and prepare for long-term preservation
- Facilitate future access to and use of data.

Want to know more about managing your research data?

Learn more

University of Otago Library, Dunedin, New Zealand, 2016

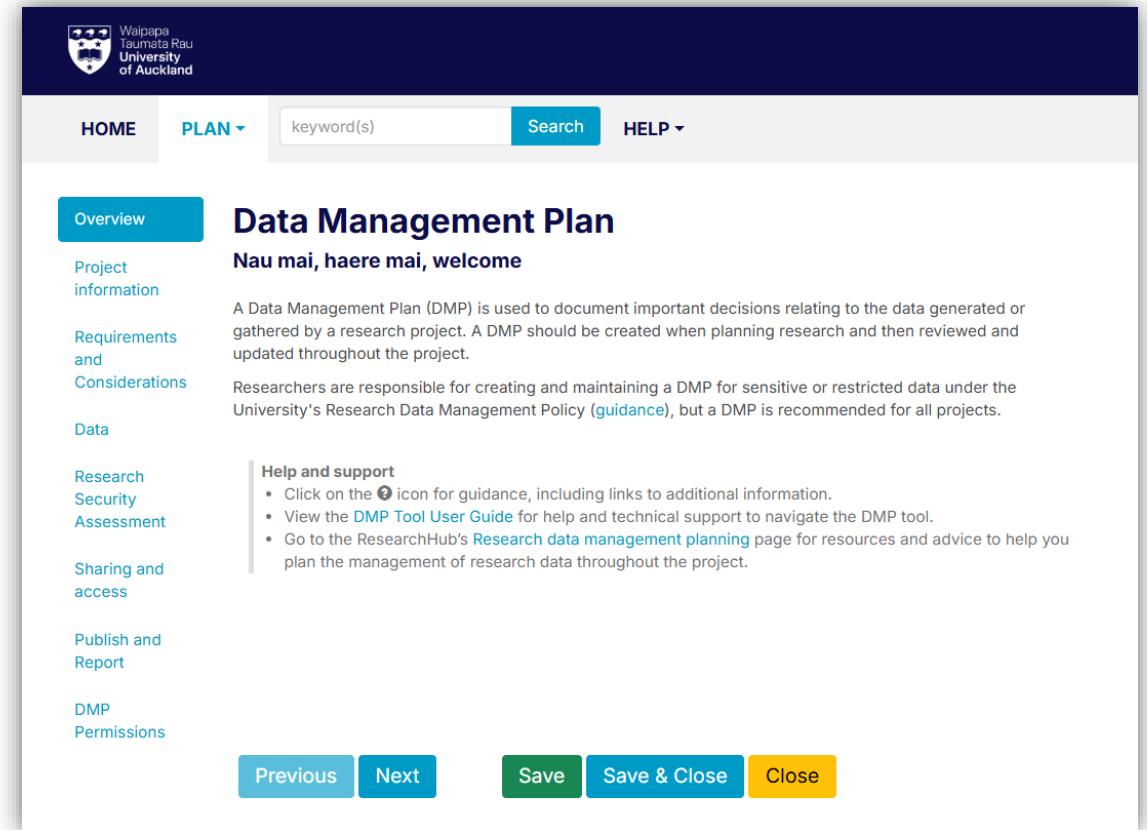


[Checklist for a Data Management Plan from the Digital Curation Centre](#)

# Data Management Planning

- **Project Information:** Purpose, people, roles & responsibilities
- **Requirements:** Data policies & classification
- **Data:** Collection, storage & organisation
- **Considerations:** Ethical, legal, sovereignty, funder, other considerations.
- **Sharing and Access:** Who will have access to the data in the future, retention/deletion
- **Publish and Report:** Enabling FAIR data

Login via Tuakiri & SSO <https://dmp.auckland.ac.nz/>



Waiapapa  
Taumata Rau  
University  
of Auckland

HOME PLAN keyword(s) Search HELP

Overview **Data Management Plan**

Project information **Nau mai, haere mai, welcome**

Requirements and Considerations

Data


Research Security Assessment

Sharing and access

Publish and Report

DMP Permissions

Help and support

- Click on the  icon for guidance, including links to additional information.
- View the [DMP Tool User Guide](#) for help and technical support to navigate the DMP tool.
- Go to the ResearchHub's [Research data management planning](#) page for resources and advice to help you plan the management of research data throughout the project.

Previous Next Save Save & Close Close



# Activities



Zoom poll

What are you currently using to document your research data management plan?

# Data management planning

## Plan your data early

- Use a **Data Management Plan (DMP)** to plan and document your data management decisions
- Identify **ethical, legal, and sovereignty considerations** early
- Clarify **roles and responsibilities** within your project

**Next** → How data is created and organised





# Data Creation and Collection

- Types and sources of data
- Strategies to organise and classify data
- Legal and ethical requirements



# Types and sources of data

Data can come from many sources. New data you collect or existing data you reuse.

Data type	What it is	Examples
 <b>Primary data</b>	Data collected firsthand for the current research project	Surveys, interviews, focus groups, experiments, observations
 <b>Secondary data</b>	Existing data reused for a new research purpose	Census data, government statistics, archived datasets, previous research data



# Considerations for data creation

**Data collection is often when the most identifiable information is present.**

## Considerations:

- What data\* will be created or collected?  
(e.g., type, format, volume, whether pre-existing or new)?  
*\*Raw physical data (inputs) + raw digital data (outputs) + derived digital data + final datasets*
- How will the data be collected or created?
- What standards or methodologies will be used for data collection?
- What quality assurance processes will be adopted?
- Do the chosen formats and software enable sharing and long-term access to the data?



# Secondary data reuse

**If you plan to use existing (secondary) data, you will need to understand and comply with any terms of use under which the data may be used or shared.**

## Considerations:

- Is it **identifiable** or **re-identifiable** data? If yes, ethics approval is required.
- Confirm that **participant consent** included use for secondary analyses. If no, ethics approval and potential re-consent are required.
- If you are working with **routine health data**, there are additional requirements. Check [Health New Zealand | Te Whatu Ora Health Research Support Service](#).
- Check the quality of the data.
- Check data use agreements or licenses to understand how the data can be used.
- If you reuse data, cite it.

# Data collection software & tools

**Data collection is often when the most identifiable information is present, so collection tools should be selected carefully**

- **Use University-supported software or tools, whenever possible.**  
These applications are often security tested and approved
- Always consider where and with whom your data is being sent, stored, or shared during collection and processing.
- Is data backed up and is it secure?



**[An overview of REDCap](#)**

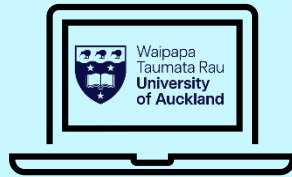
Tuesday 30 June, 1pm-2pm

**[Capturing Longitudinal Data in REDCap](#)**

Wed 1 July, 1pm-2pm

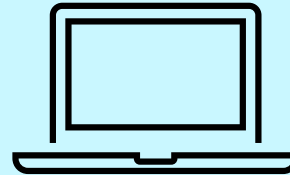
# Audio & video recording

## University-provided



## Personal device

\*up-to-date, screen lock,  
password-protected



## Handheld recording device



## University-managed data capture



Microsoft Teams



- ✓ Sign in to these tools using your University account
- ✓ Record to the cloud
- ✓ For Zoom: turn off the **AI Companion** capabilities (e.g. summaries using generative AI) to prevent data from being shared with third-party genAI models
- ✓ Move audio recordings to secure storage as soon as possible

# Surveys, questionnaires & databases

Qualtrics 

REDCap 

---

Is it easy to use?

Yes

A bit more work upfront

---

Do the surveys look nice?

Yes  
([Qualtrics Demo](#))



Yes  
([REDCap Demo](#))



---

Where is the data stored?

XM data centres, likely  
Australia - **offshore**

University of Auckland data  
centres - **onsite**

---

Is it secure?

Yes

Yes

---

What support is available?

Built-in help  
  
[XM Basecamp](#)

Built-in help  
REDCap Administrator  
Online tutorials  
HHRS (cost recovery)

# Tidy Data principles for tabular data

1. Every column is a variable.

2. Every row is an observation.

3. Every cell is a single value.

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

- 1st row variable names (no spaces, simple).
- 1 sheet/table per file.
- Save in a 'plain text' format (.csv).
- Use a README.txt for background and context.

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

# Use of generative AI in research

- No standalone 'AI Policy' - expectations for how we use AI methods and tools will be woven into existing policies.
- [AI at the University](#)
- Supervisors to approve project- and activity-specific use of generative AI.  
(Implementation of the *Guidelines for the use of GenAI in doctoral research* planned for Q3, 2026)

## University approved AI tools

### Enterprise-wide tools

For secure access, sign in to these tools using your University account. This is your abcd123@aucklanduni.ac.nz address.

- [Microsoft 365 Copilot Chat](#) is approved for public, internal and sensitive data.
- [Google Gemini](#) is approved for public, internal and sensitive data.
- [Google AI Studio](#) is approved for public, internal and sensitive data.
- [NotebookLM](#) is approved for public and internal data.

### Free or user-pays tools

Before using these tools, we recommend going to settings and turning off permissions for Large Language Model (LLM) training.

- [Claude Pro](#) and [Claude Code](#) are approved for public and internal data.
- [ChatGPT](#) is approved for public data.
- [Consensus.ai](#) is approved for public data.
- [Perplexity](#) is approved for use with public and internal data.

### Developer platforms

Costs for developer platforms are based on usage.

- [Watsonx.ai](#) is approved for use with public and internal data.
- [Azure AI](#) is approved for use with public and internal data.




The above tools have been assessed by Digital Services to ensure compliance with our [IT Security Policy](#), [Privacy Policy](#) and [Research Data Management Policy](#).

Further information about [using the right AI tool for your research is available on the ResearchHub](#).



# Research metadata

Metadata describes and gives context for data (allowing discovery and reuse).

Type	What it is	Examples
 <b>Descriptive</b>	Helps people <b>find and understand</b> the dataset	Title, creator/author, description, keywords, dates
 <b>Administrative</b>	Supports <b>access, governance, and compliance</b>	Ownership, access permissions, ethics approvals, licensing, retention/disposal rules
 <b>Technical</b>	Enables <b>correct use and interpretation</b>	File formats, variables, software methods, units of measurement, data structure, instrument settings

# Data security classification

**PUBLIC**

**INTERNAL**

**SENSITIVE**

**RESTRICTED**

- Considers damage or harm to people and the reputation of the University
- Helps you to:
  - Meet requirements (e.g., a DMP is required if working with sensitive or restricted data)
  - Identify systems/controls (e.g., what software or storage services are available if I am working with sensitive data?)

Examples:

- require **human ethics approval**, incl. personal information and de-identified health data
- require animal **ethics approval**
- are **commercially sensitive**
- may be protected as **intellectual property**
- are subject to **export controls**
- are part of **national defence**





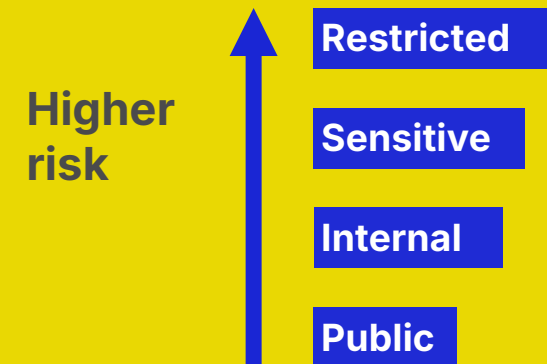
# Activities



**Quiz question:** Select the most likely research data classification for each of the examples below:

1. A publicly available dataset **Public**
2. Interview audio files discussing a controversial or private topic **Sensitive**
3. Research into a new radar/communications system for the military **Restricted**
4. Clinical trial data containing health information with personal identifiers **Sensitive**
5. An anonymous online survey (expedited ethics) **Internal**

## Classification





# Ethical considerations

What are the **ethical considerations** surrounding the research data?

## Privacy and confidentiality

- Does the project involve the collection of personal information (especially health information)? What processes will you use to de-identify data to ensure confidentiality? What is the risk of re-identification?

## Access controls, including storage and transfer of data

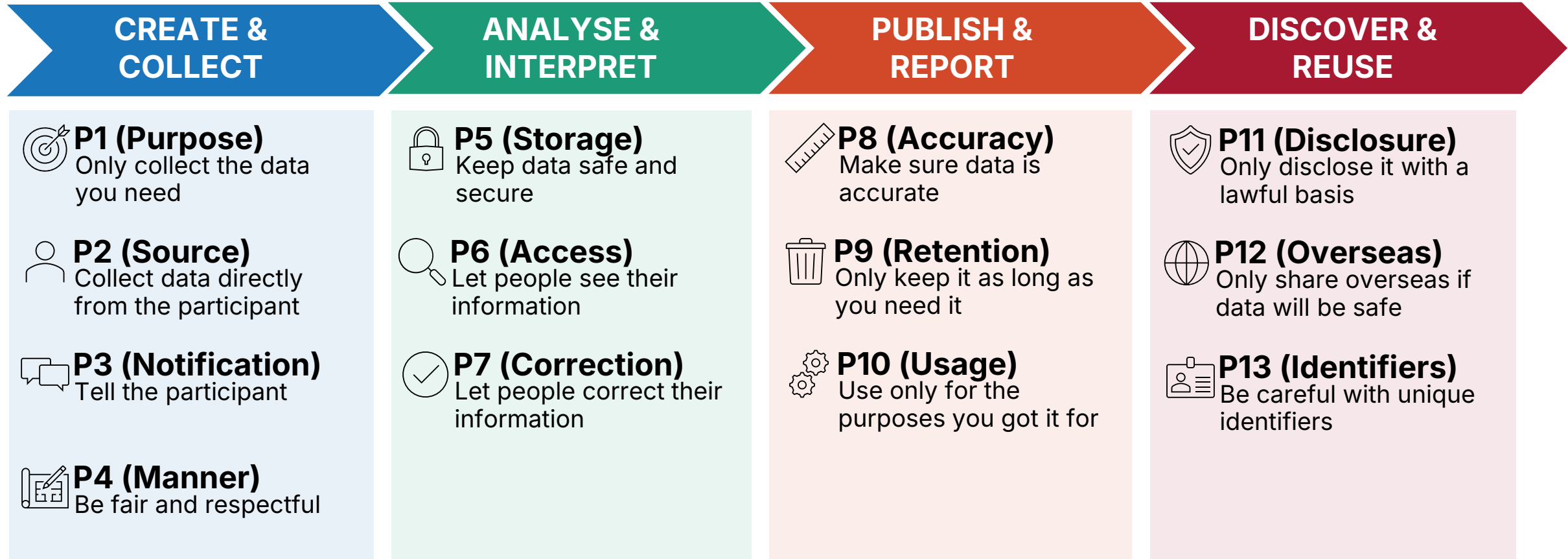
- Where will data be stored? What software or systems will be used to manage or analyse the data? Who will have access to the data, and how will access be controlled?

## Data deletion, archive and/or sharing for reuse

- What are the conditions of data sharing, including applying an appropriate license? Have you gained consent for data preservation and sharing? How and when will the data be archived and/or deleted?

# Privacy Act 2020

Principles governing the collection, handling and use of personal information



Further details >> [Information sheet about the Privacy Act 2020 from the Privacy Commission](#)  
>> [Information sheet about the Health Information Privacy Code](#)



# De-identifying data



Working With Personally Identifiable Research Data  
Monday 29 June, 1pm – 2pm

## Identifiable

Data that directly or indirectly identifies an individual or business.

## De-identified

Data which has had information removed from it to reduce risk of spontaneous recognition.

## Confidentialised

Data which has had statistical methods applied to it to protect against disclosing unauthorised information.

Identifiable		De-identified		Confidentialised	
Individual	Business	Individual	Business	Individual	Business
<b>Name</b> Hēni <b>Gender</b> Female <b>DOB</b> 31/01/1985 <b>Address</b> 28 My Road Postcode 6012 Wellington	<b>Name</b> Puzzles <b>Type</b> Paper Stationery Manufacturing  <b>Employees</b> 34 <b>Expenditure</b> \$398,000	<b>Name</b> <i>Unknown</i> <b>Gender</b> Female <b>DOB</b> 1985 <b>Address</b> Postcode 6012 Wellington	<b>Name</b> <i>Unknown</i> <b>Type</b> Manufacturing  <b>Employees</b> 30 - 40 <b>Expenditure</b> \$398,000	<b>Name</b> <i>Unknown</i> <b>Gender</b> Female <b>Age</b> 30 - 40 years <b>Address</b> Wellington	<b>Name</b> <i>Unknown</i> <b>Type</b> Manufacturing  <b>Employees</b> 10 - 100 <b>Expenditure</b> Under \$500,000

# Data creation and collection

## Create structured, usable data

- Use **approved tools**
- Use **metadata** to describe the data – what it is, how it was created and how it should be used
- **Data security classification** is a decision-making tool to help you to assess data sensitivity and consider what level of protection it requires

**Next** → How to analyse and secure your data



# Data Analysis, Storage & Security

- Storage and backup
- Organisation, filenames, version control
- Processing and analysis
- Data security

# Digital research data storage

Every stage of the research data lifecycle involves consideration of data storage

## Key considerations:

- Where is the primary copy? \*Where will research data be stored?
- Where does data collection and analysis happen?
- What needs to be backed up or protected?
- What does your funder, ethics and/or organisation require?

Researchers at the University of Auckland are responsible for:

9. Ensuring that digital forms of research data are stored on an appropriate University-managed research storage service or other trusted storage service approved by the University Chief Information Security Officer.



# Research data storage

Remove from...



CREATE & COLLECT

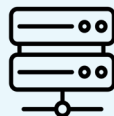
ANALYSE & INTERPRET

AFTER THE PROJECT

Research data  
"Primary copy"

University-managed research data storage

Research Drive



Dropbox for Researchers

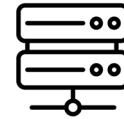
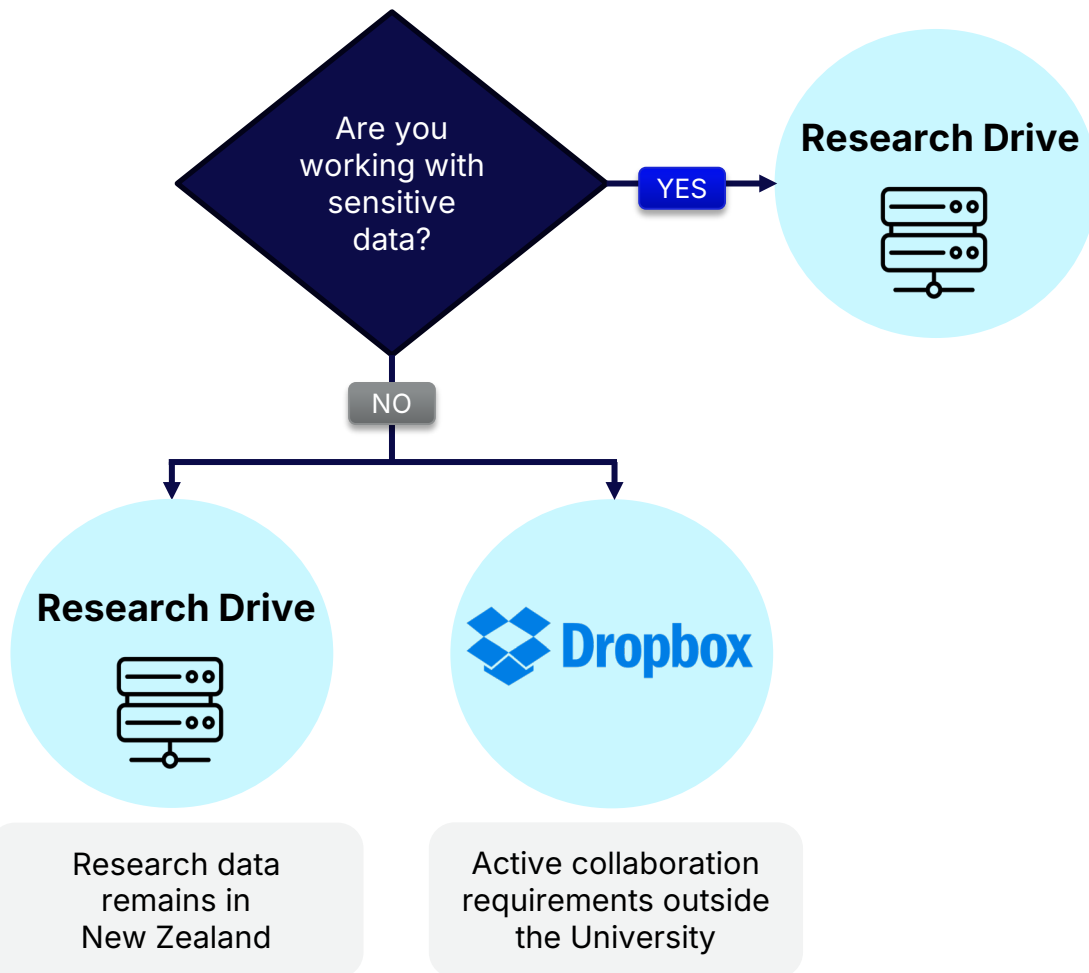


Locations  
where data  
collection &  
analysis happen  
(Temporary)

University-managed  
data capture



# Choosing research data storage



## Research Drive

- Located in New Zealand
- Requires University credentials (includes external collaborators)
- Access controls: read-write or read-only access *across the entire drive*



## Dropbox for researchers

- Not located in New Zealand
- Sub-folders can be shared with specific people (including anyone outside of the University)

**OneDrive** is *not appropriate* for research data storage (at University of Auckland) because of the increased risk for data loss, limited storage space and insufficient data provenance (data history) for research integrity.

# Backing up your research data

## Use University-managed storage services

- ✓ Research Drive & Dropbox for Research are backed up automatically.
- ✓ Files can be recovered from deletion up to 180 days.

## General principles

- What data/files need to be backed up? How often? Where? By whom?
- At least 2 people should have access to the data.

## Be aware of the 3-2-1 rule



At least  
3 copies



Using at least 2 different  
storage media

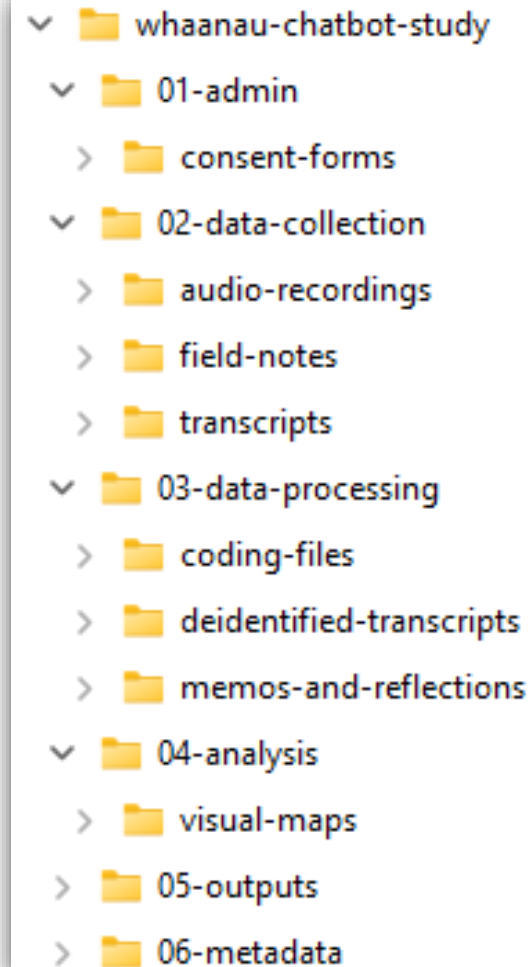


At least  
1 copy offsite

# Organising research data

## Folder structures

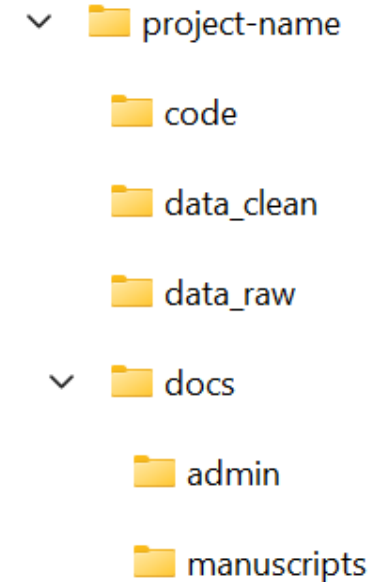
- CLEAR, CONCISE, CONSISTENT
- Folder hierarchy
  - short, descriptive folder names
  - avoid overlapping categories
  - limit size and depth of folders
- Consistent strategy prevents confusion
- Things are easy to find and to sort
- Document your strategy
- Set up and use databases if necessary



```
graph TD; A[waaanau-chatbot-study] --> B[01-admin]; A --> C[02-data-collection]; A --> D[03-data-processing]; A --> E[04-analysis]; A --> F[05-outputs]; A --> G[06-metadata]; B --> B1[consent-forms]; C --> C1[audio-recordings]; C --> C2[field-notes]; C --> C3[transcripts]; D --> D1[coding-files]; D --> D2[deidentified-transcripts]; D --> D3[memos-and-reflections]; E --> E1[visual-maps];
```

▼ waaanau-chatbot-study

- ▼ 01-admin
  - > consent-forms
- ▼ 02-data-collection
  - > audio-recordings
  - > field-notes
  - > transcripts
- ▼ 03-data-processing
  - > coding-files
  - > deidentified-transcripts
  - > memos-and-reflections
- ▼ 04-analysis
  - > visual-maps
- > 05-outputs
- > 06-metadata



```
graph TD; A[project-name] --> B[code]; A --> C[data_clean]; A --> D[data_raw]; A --> E[docs]; E --> F[admin]; E --> G[manuscripts];
```

▼ project-name

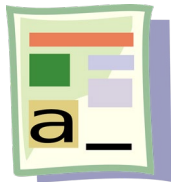
- code
- data\_clean
- data\_raw
- ▼ docs
  - admin
  - manuscripts



Further reading: [Project structure by Danielle Navarro](#)

# File naming conventions

- Create a template and document it
- Short, descriptive and use only important fields
- Avoid spaces or special characters and ambiguity

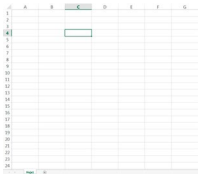


20170310-tmr-literature-review.docx

[date]-[creator]-[subject].[ext]

arthnz-rat-rbw-food-weights.xlsx

[project]-[animal model]-[creator]-[data type].[ext]



UCollege\_AndersonM\_Ped-Resp-Infection-Genomic-Determinants\_Biosketch\_20160125.pdf

[CTSA]\_[InvestigatorLastNameFirstInitial]\_[ProtocolShortTitle]\_[Document]\_[YYYYMMDD].[ext]

teko-van-kuyk\_pineapple-41-white-purple-black\_35x50\_2017

[artist-name]\_[artworks-name]\_[length-X-height(depth in case of sculpture)]\_[date or year]



fr3s-140623-129C-2653-w.jpg

[studysite,depth of water]-[yyymmdd]-[tile#,treatment]-[photo#]-[photo coverage].[ext]



# Research data documentation

## Data-level documentation

- Provides information on individual data objects, such as a variable in a data file or interview transcript.
- Should be brief and concise, include units of measurement, and use established classifications and standards, where appropriate.

## Study-level documentation

- Provides high-level information on the research context and design, the data collection methods, data preparations and manipulations, etc.

### Examples:

- Electronic lab notebook and protocol
- Data Dictionary
- Codebook
- Software syntax and output files
- Metadata record or database schema
- Links to reports and publications (DOIs)
- Provenance information (third-party data)

# A README example

Title: Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002

Authors:

- TORU HIRAWAKE (INVESTIGATOR, TECHNICAL CONTACT)
- DAVE CONNELL (DIF AUTHOR)

Owner: Commonwealth of Australia

Citation: Hirawake, T. (2005) Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002, Ver. 1, <em>Australian Antarctic Data Centre</em> - <a href="http://dx.doi.org/doi:10.4225/15/5a384270f2b61">doi:10.4225/15/5a384270f2b61</a>, Accessed: 2020-07-02

Revision Date: 2017-12-18

Metadata URL: [https://data.aad.gov.au/metadata/records/chlorophyll\\_65-02](https://data.aad.gov.au/metadata/records/chlorophyll_65-02)

Dataset API URL: [https://data.aad.gov.au/s3/api/bucket/datasets/science/chlorophyll\\_65-02?export=json](https://data.aad.gov.au/s3/api/bucket/datasets/science/chlorophyll_65-02?export=json)

Abstract: The variation in the phytoplankton biomass over a decadal time scale, and its relationship with the Antarctic Circumpolar Wave (ACW) and climate change, has been poorly interpreted because of the limited satellite chlorophylla (chl a) data compared with the physical parameters from satellite. We analysed a long-term chl a dataset along the Japanese Antarctic Research Expedition (JARE) cruise tracks since 1965 to investigate inter-annual variation of phytoplankton biomass. In the Southern Ocean, increasing trends of chl a and the spreading of higher chl a area to the north with 3-7 year cycles were found. Although relationships between the decadal change in chl a and climate change such as variation of sea ice extent and the El Nino are still obscure, large variation of primary production in proportion to the chl a is implied.

The chl a concentration of sea surface water has been measured routinely on board the icebreakers Fuji and Shirase during almost every cruise of the JARE.

The download file contains chlorophyll a data collected from ship tracks on JARE voyages between 1965 and 2002.

The field in this dataset are:

Date (local time)  
Year  
Latitude  
Longitude  
Corrected Chlorophyll a

See the attached paper for more details.

The publications on the data collected during the 1965-1976 and 1988-1993 cruises are listed in Fukuchi [1980] and Suzuki and Fukuchi [1997], respectively. For data on the 1977-1985 and 1994-1997 cruises, see [Kanda and Fukuchi, 1979; Fukuchi and Tamura, 1982; Tanimura, 1981; Watanabe and Nakajima, 1983; Ino and Fukuchi, 1984; Sasaki, 1984; Hamada et al., 1985; Fukuda et al., 1986; Hattori and Fukuchi, 1988; Midorikawa et al., 2000]. Data post 1998-2002 cruises is in Hirawake and Fukuchi [2004]. Data from the 1986-1987 will be published in the JARE data report of digital media, including all cruise data.

Auxiliary Material for paper 2004GL021394 Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002. Toru Hirawake, Tsuneo Odate and Mitsuo Fukuchi (National Institute of Polar Research, Tokyo) Geophys. Res. Lett., Vol (Num), doi:10.1029/2004GL021394

All of the chl a data have been reported in the publications of the National Institute of Polar Research (NIPR).

Use Constraints: This data set conforms to the CCBY Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

Please follow instructions listed in the citation reference provided at [http://data.aad.gov.au/aadc/metadata/citation.cfm?entry\\_id=chlorophyll\\_65-02](http://data.aad.gov.au/aadc/metadata/citation.cfm?entry_id=chlorophyll_65-02) when using these data.

ID: 1360

Metadata ID: chlorophyll\_65-02

UUID: EBD86EDA-DE08-4C0F-AC94-AF171DBCE46A



# Compute resources for data analysis

How you analyse data directly depends on the computing environment you use.

## Key considerations:

- Data size & growth
- Processing intensity
- Security & sensitivity
- Collaboration needs
- Reproducibility & documentation
- Cost & access constraints

## No single best compute option.

Match your compute environment to the data – considering size, complexity, sensitivity and your analysis goal.





# Compute options

Data size, complexity, and format influence compute needs for analysis

Compute	Data / Analysis requirements	Examples
<b>Local machine</b>	Small, simple workflows	<i>Desktops, laptops</i>
<b>Virtual machines</b>	Moderate, reproducible workflows	<i>Managed VMs, Nectar</i>
<b>HPC</b>	Large-scale, structured processing	<i>REANNZ</i>
ML / AI platforms	Image processing, prediction	
GPUs	Deep learning, AI models	<i>GPU workstations</i>

# Transcription tools

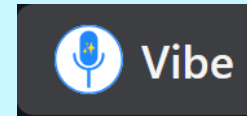
Transcribe:



Record and transcribe:



Transcribe Desktop  
on Nectar virtual  
desktop



Local machine



## Considerations for de-identification of transcripts

- Create a de-identification log of all replacements or removals made
- Store the log separately from the de-identified data files
- Identify replacements in text in a meaningful way – such as indicating replaced text with square brackets [] in transcribed interviews, or using XML markup tags like <anon>...</anon>
- For qualitative data like transcribed interviews or survey textual answers, use pseudonyms or generic descriptors rather than blanking out information

# What is reproducibility?

Reproducibility means that research findings can be independently verified by repeating the analysis using the original data, code, and documented methods.

If another researcher can follow your documented process and get the same results, your research is **reproducible**.

Term	Same data?	Same methods?	Purpose
Reproducibility	✓	✓	Verify analysis
Replication	✗	✓	Test generalisability
Repeatability	✓ (same team)	✓	Internal consistency



# Reproducible data workflows

Analysis is where **data provenance** is most easily lost – and most important.

**To track how data changes (cleaning/processing) and versions:**

- Keep a read-only copy of raw data
- Work on a separate, editable version for cleaning and analysis
- Document the steps used to get from raw data to tidy data
- Create a README or data dictionary describing variables, units and context

**Scripted cleaning and analyses (e.g., R, Python, Stata, SQL)** – transparent, shareable, reviewable, scalable. Supports **reproducibility**, protects **data integrity**, and enables **collaboration and reuse**.



# Research data storage

Remove from...



CREATE & COLLECT

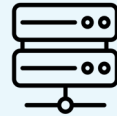
ANALYSE & INTERPRET

AFTER THE PROJECT

Research data  
"Primary copy"

University-managed research data storage

Research Drive



Dropbox for Researchers



Locations  
where data  
collection &  
analysis happen  
(Temporary)

University-approved  
data capture



Compute-adjacent  
storage (VMs, HPC)



Secure Research  
Environment





# Activities

## Documenting research data analysis



In the Zoom chat...

What is one analysis step in your research that would be hard for someone else to reproduce?

Identify one piece of documentation that would fix that.

**If you can't explain it later, you'll struggle to defend it.**



# Principles of data security

**Good security balances protection and usability**

Principle	What is it?	Example
<b>Confidentiality</b>	Only authorised people can access data	Restricting access to identifiable data
<b>Integrity</b>	Data is accurate and complete, not altered improperly	Preventing accidental overwriting of datasets
<b>Availability</b>	Data is accessible when needed	Backups ensure data isn't lost



[CIA Triad](#) – described by National Institute of Standards and Technology (NIST)



# Data security risk assessment

## Identifying what could go wrong and how likely/serious it is

### Common risks in research

- Unauthorised access to sensitive data
- Lost or stolen devices (laptops, USBs)
- Weak passwords or shared logins
- Accidental sharing (emailing wrong file)
- Poor data handling or storage practices

### Simple approach

1. Identify risks
2. Assess likelihood and impact
3. Apply appropriate controls

# Security controls

## Encryption

- Protects data if stolen or intercepted
- **At rest** (stored files, drives)
- **In transit** (email, uploads, downloads)

Using FileSender to share data/files

## Access Controls

- Limit data access to **those who need it**
- Role-based access

Restrict access to identifiable data

## Authentication

- Verifies user identity
- Strong passwords + multi-factor authentication (MFA)
- Role-based access

Institutional SSO + MFA

## Audit & Monitoring

- Track who accessed or changed data
- Helps detect and investigate issues

System logs of file access

## Good personal practices

- Lock devices when unattended
- Don't share passwords
- Store data in approved systems (not personal drives)
- Be cautious with email attachments/links

# Secure research environment (SRE)

Project-specific virtual platform for the secure storage and analysis of sensitive data.



## Security

- Controlled access for authorised people
- Project data is encrypted
- Auditable data import, export and access



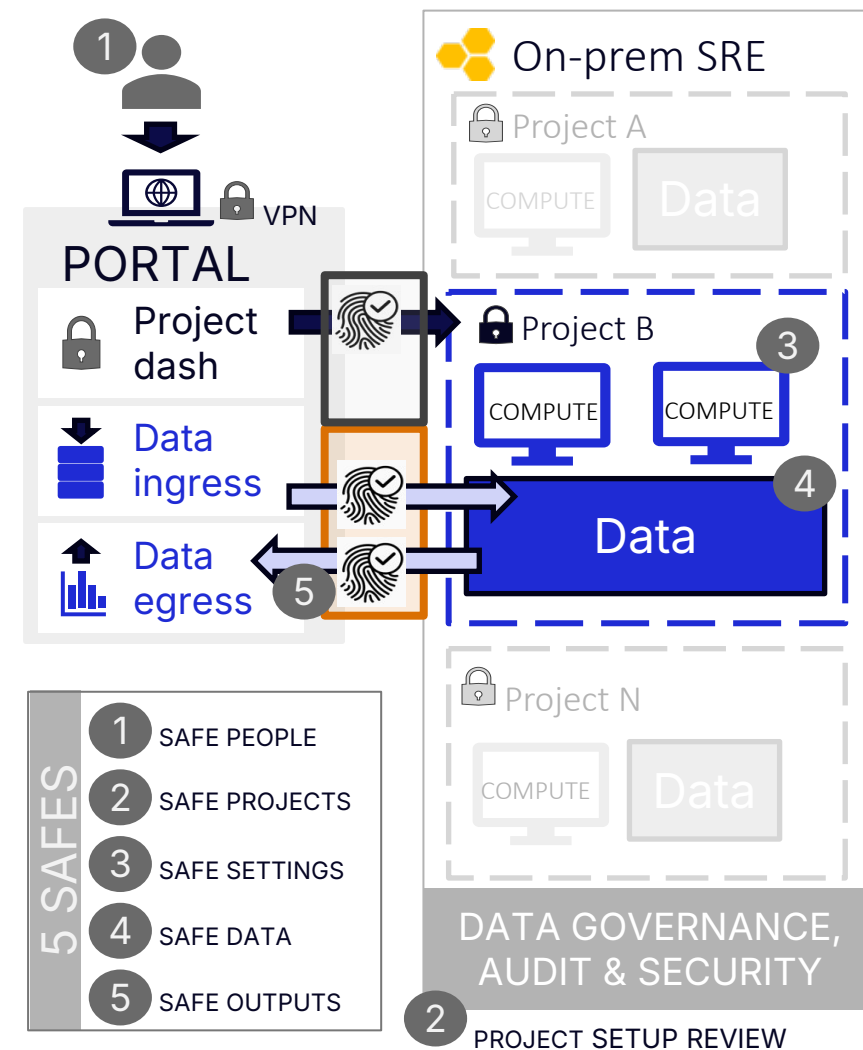
## Collaboration

- Shared workspace for a project with data and analytical tools in one place



## Compliance

- University approved for sensitive data
- Aligned with University security and governance standards





# What to do in case of a data breach

**A data breach = unauthorised access, loss, or disclosure**

## Immediate steps

1. Contain the breach (e.g., disconnect device / revoke access)
2. Report immediately and follow institutional process
  - [Privacy Breach Management Procedures](#) (University of Auckland)
  - [Report an information security incident](#) (University of Auckland)
3. Document what happened (e.g. what data, who was affected, when)

Support teams will help with next steps – assess risk and impact, notify affected parties (if required) ,review and improve security practices.



# Data analysis, storage and security

## Make analysis reproducible and protect data

- Keep raw data intact and document all changes
- Choose appropriate compute and storage environments
- Apply data security practices to protect sensitive data

**Next** → How to share and preserve your data



[Five Safes in Action: A Tour of a Secure Research Environment](#)

Thurs 2 July, 4pm – 5pm



# Data Sharing and Archiving

- Rationale for data sharing
- Access conditions for sensitive data (Five Safes)
- Archive and retention

# FAIR and CARE data principles



Global Indigenous Data Alliance <https://www.gida-global.org/care>

## FAIR data principles

- ✓ Make data easier to discover
- ✓ Promotes (human) actions now to enable wider sharing and reuse later (by humans and machines)
- ✓ Supports reproducibility and open research

## CARE principles for indigenous data

- ✓ Complement FAIR by considering people and purpose
- ✓ Promotes the governance of indigenous data and advancement of indigenous data sovereignty, innovation and self-determination

F

**Findable**

Metadata (descriptive information), DOI and process for access are external facing, human and machine readable E.g., internet search results, bibliographic databases.

A

**Accessible**

(others know how to access)

I

**Interoperable**

Metadata is with data and disciplinarily specific. Combining and using data are enabled by format and file type(s).  
E.g., Data Management Plan, Protocol, README.txt

R

**Reusable**

# Australian Antarctic Data Centre

Data management and spatial data services

[Menu](#)[Login](#)[Support](#)

[Australian Antarctic Data Centre](#) / [Discover and Manage Data](#) / [Records](#) / chlorophyll\_65-02

## Metadata details

[Request DOI](#)[Edit record](#)

### chlorophyll\_65-02

[View the full metadata record](#)

#### Citation

Hirawake, T. (2005) Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002, Ver. 1, *Australian Antarctic Data Centre* - doi:10.4225/15/5a384270f2b61, Accessed: 2025-04-07

#### Title

Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002

#### Data Centre

Australian Antarctic Data Centre, Australia

#### DOI

doi:10.4225/15/5a384270f2b61

#### Created Date

2005-08-22

#### Revision Date

2017-12-18

#### Parent record

None

Unique persistent  
identifier (DOI)  
= Findable

## Datasets and documents

### chlorophyll\_65-02

Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002

[Download dataset](#) [View dataset contents](#)

Public

Submitted 22 Aug 2005

### chlorophyll\_65-02

Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002

[Download dataset](#)

Released - AAD Only

Submitted 22 Aug 2005

## Related links

[Download point for the data - Excel spreadsheet](#)

[Download point for the data - papers - AAD Staff Only](#)

[Citation reference for this metadata record and dataset](#)

## Access

These data are publicly available for download from the provided URL. A copy of some of the referenced publications is available for download by AAD staff only.

## Temporal Coverages

- **Start date:** 1965-11-23 - **Stop date:** 2002-12-08

## Spatial Coverages



Latitude	Longitude
<b>Northernmost:</b> 24.567	<b>Westernmost:</b> 100.147
<b>Southernmost:</b> -54.985	<b>Easternmost:</b> 137.95

Describe access conditions = Accessible

## Science Keywords

- EARTH SCIENCE > CLIMATE INDICATORS > ATMOSPHERIC/OCEAN INDICATORS > TELECONNECTIONS > ANTARCTIC OSCILLATION
- EARTH SCIENCE > CLIMATE INDICATORS > ATMOSPHERIC/OCEAN INDICATORS > TELECONNECTIONS > EL NINO SOUTHERN OSCILLATION (ENSO)
- EARTH SCIENCE > BIOSPHERE > ECOSYSTEMS > AQUATIC ECOSYSTEMS > PLANKTON
- EARTH SCIENCE > OCEANS > OCEAN CHEMISTRY > PIGMENTS > CHLOROPHYLL
- EARTH SCIENCE > BIOSPHERE > ECOLOGICAL DYNAMICS > ECOSYSTEM FUNCTIONS > BIOMASS DYNAMICS

## Additional Keywords

- CHLOROPHYLL A
- JARE
- PHYTOPLANKTON
- SOUTHERN OCEAN

## Locations

- OCEAN > INDIAN OCEAN
- OCEAN > SOUTHERN OCEAN
- OCEAN > PACIFIC OCEAN
- GEOGRAPHIC REGION > POLAR

## Use Constraints

This data set conforms to the CC BY Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

Please follow instructions listed in the citation reference provided at [http://data.aad.gov.au/aadc/metadata/citation.cfm?entry\\_id=chlorophyll\\_65-02](http://data.aad.gov.au/aadc/metadata/citation.cfm?entry_id=chlorophyll_65-02) when using these data.

### Project

### ISO Topic

### Dataset Language

- BIOTA
- CLIMATOLOGY/METEOROLOGY /ATMOSPHERE
- OCEANS

- ENGLISH

### Originating Centre

### Dataset Progress

### IDN Node

- JARE

- COMPLETE

- AMD/AU
- CEOS
- AMD

## Publications

- Fukuchi, M. (1980) Phytoplankton chlorophyll stocks in the Antarctic Ocean, *J. Oceanogr. Soc. Jpn.*, 36, 73-84
- Fukuchi, M., and S. Tamura (1982) Chlorophyll a distribution in the Indian sector of the Antarctic Ocean in 1978-1979, *Antarct. Rec.*, 74, 143-162
- Fukuda, Y., M. Ohno, K. Iwanami, and H. Touju (1986) Chlorophyll a content in the surface and subsurface waters along the course of the Shirase to Antarctica in 1984-1985, *Antarct. Rec.*, 30, 103-112
- Hamada, E., A. Taniguchi, M. Okazaki, and Y. Naito (1985) Report on the phytoplankton pigments measured during the JARE-25 Cruise to Syowa Station, Antarctica, November 1983 to April 1984, *JARE Data Rep.*, 89, Natl. Inst. Polar Res., Tokyo, 103
- Hattori, H., and M. Fukuchi (1988) Report on the phytoplankton pigments concentrations, zooplankton and benthos sampling during the JARE-27 cruise, November 1985 - April 1986, *JARE Data Rep.*, 28, Natl. Inst. Polar Res., Tokyo, 135
- Hirawake, T., and M. Fukuchi (2004) Chlorophyll a concentration of phytoplankton during the cruises of 40-44th Japanese Antarctic Research Expedition in 1998-2003, *JARE Data Rep.*, 31, Natl. Inst. Polar Res., Tokyo, 279
- Ino, Y., and M. Fukuchi (1984) Report on chlorophyll a distribution along the course of the Fuji in 1981-1982, *Antarct. Rec.*, 81, 38-44
- Kanda, H., and M. Fukuchi (1979) Surface chlorophyll a concentration along the course of the Fuji to and from Antarctica in 1977-1978, *Antarct. Rec.*, 66, 37-49
- Midorikawa, T., K. Nomura, Y. Miyamoto, T. Odate, A. Ishikawa, N. Washiyama, T. Hirawake, M. Namiki (2000) Report on phytoplankton pigments measured during the JARE-36-39 cruises to Syowa Station, Antarctica in 1994-1998, *JARE Data Rep.*, 249, 36, Natl. Inst. Polar Res., Tokyo
- Sasaki, H. (1984) Distribution of nano- and microplankton in the Indian sector of the Southern Ocean, *Mem. Natl. Inst. Polar Res. Spec. Issue*, 32, 38-50
- Suzuki, T., and M. Fukuchi (1997) Chlorophyll a concentration measured with a continuous water monitoring system during the cruise to Syowa Station, Antarctica, JARE-27 (1985/86) to JARE-35 (1993/94), 60, Natl. Inst. Polar Res., Tokyo
- Tanimura, A. (1981) Distribution of the surface chlorophyll a along the course of the Fuji to and from Antarctica in 1979-1980, *Antarct. Rec.*, 72, 35-48
- Watanabe, K., and Y. Nakajima (1983) Surface distribution of chlorophyll a along the course of the Fuji (1980/81) in the Southern Ocean, *Antarct. Rec.*, 77, 33-43



# chlorophyll\_65-02

Metadata Entry ID: chlorophyll\_65-02

Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002

[View Metadata Record](#)[Download Dataset](#)

## Contents

### Resource

[LICENSE](#)[README](#)[chlorophyll\\_65-02.csv](#)[chlorophyll\\_65-02.xml](#)

Apply a license  
= Reusable

Metadata is with data  
= Interoperable

Use open file formats,  
= Interoperable

4 records

## Citation

Hirawake, T. (2005) Long-term variation of surface phytoplankton chlorophyll a in the Southern Ocean during 1965-2002, Ver. 1, *Australian Antarctic Data Centre* - doi:10.4225/15/5a384270f2b61, Accessed: 2025-04-08

## Use Constraints

This data set conforms to the CCBY Attribution License (<http://creativecommons.org/licenses/by/4.0/>). Please follow instructions listed in the citation reference provided at [http://data.aad.gov.au/aadc/metadata/citation.cfm?entry\\_id=chlorophyll\\_65-02](http://data.aad.gov.au/aadc/metadata/citation.cfm?entry_id=chlorophyll_65-02) when using these data.



C

**Collective benefit**

A

**Authority to control** their data

F

A

I

R

Findable

Accessible

R

Interoperable

Reusable

**Responsibility** to engage respectfully with those communities



Everyone

E

Indigenous Peoples' **ethics** should inform the use of data across time



Specific people and purpose

# Five Safes framework

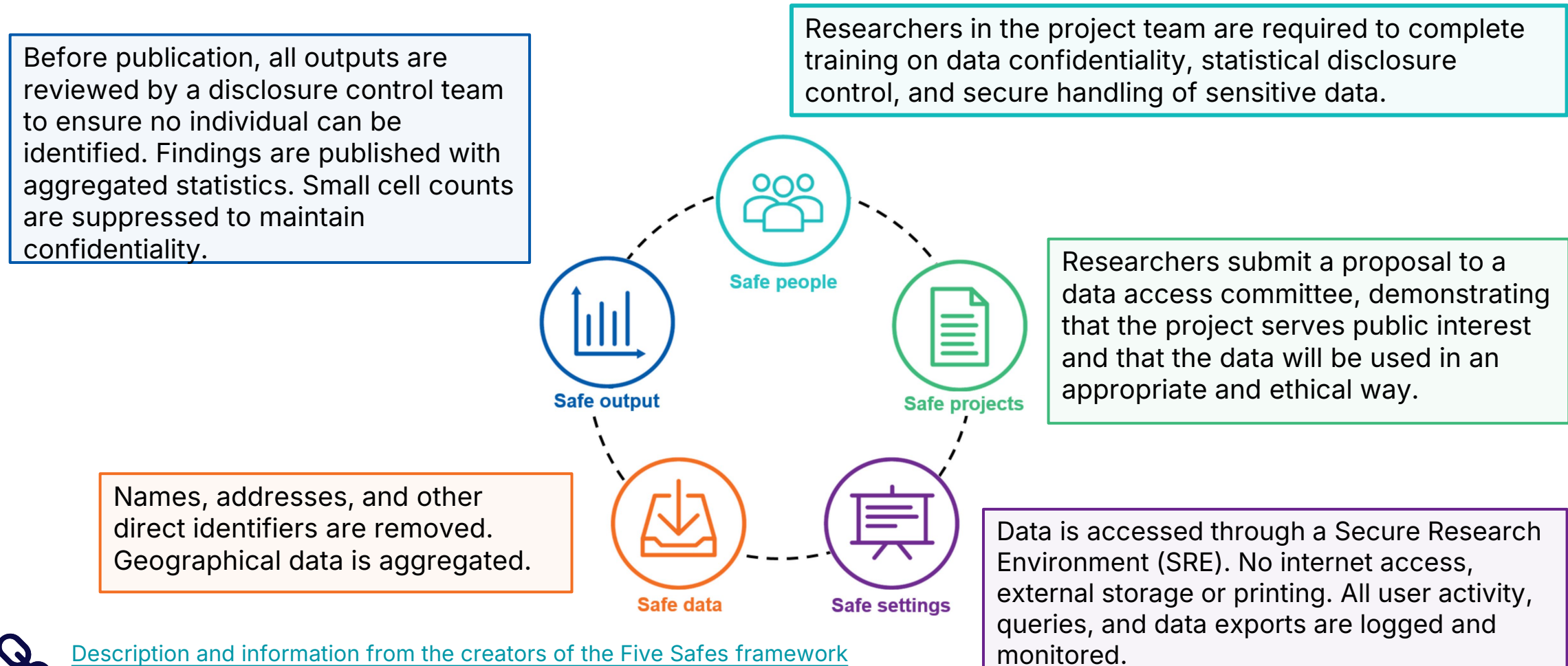
Decision-making framework to help researchers decide how to share data safely

- **Safe people:** Who can access the data? Do they have the right training and knowledge?
- **Safe project:** Is the use of the data appropriate?
- **Safe settings:** Where will the data be accessed?
- **Safe data:** How is risk reduced within the data itself?
- **Safe output:** What can be published or released?



# Evaluating the secondary reuse of data

## Using the Five Safes Framework to evaluate access to sensitive or confidential data



[Description and information from the creators of the Five Safes framework](#)  
[How Stats NZ uses Five Safes to manage access to integrated data in NZ](#)

# Ngā Tikanga Paihere

Decision-making tool to help with safe, ethical, and culturally-appropriate data practices

- Draws on tikanga (Te Ao Māori – Māori world concepts)
- Encourages community participation in data decisions
- Used alongside Five Safes



[Description of Ngā Tikanga Paihere from Stats NZ](#)

[How Stats NZ uses Five Safes and Ngā Tikanga Paihere to manage access to integrated data in NZ](#)

# How is access to sensitive data controlled?

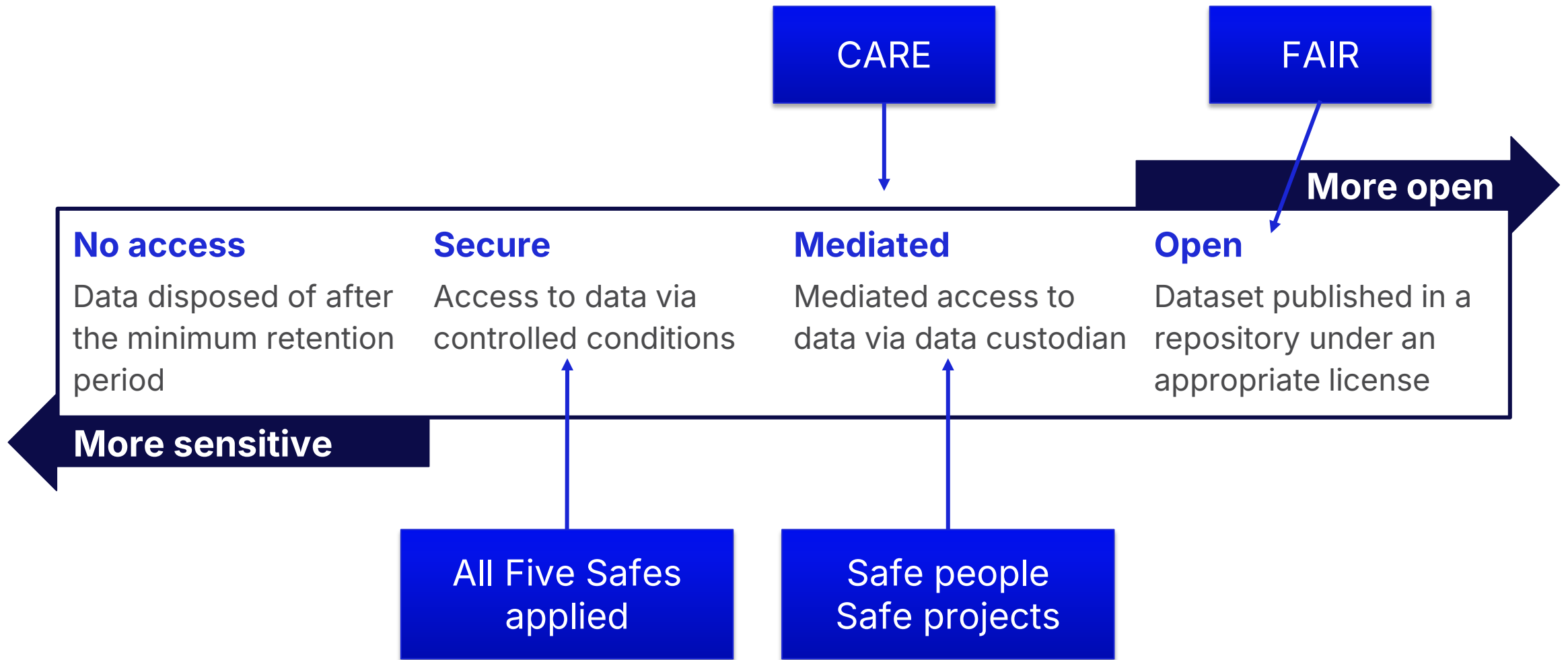
When data can't be openly shared, access is usually governed by agreements.

	Data Use Agreement (DUA)	Data Sharing Agreement (DSA)
<b>Purpose</b>	Controls how you <i>use</i> data	Controls how data is <i>shared between organisations</i>
<b>Examples</b>	Accessing an existing dataset (e.g. Health NZ data)	Active collaboration or transfer between organisations

**Five Safes helps us decide what conditions are needed**

	Include in agreements...
<b>Safe people</b>	Who is allowed access
<b>Safe projects</b>	What the data can be used for
<b>Safe settings</b>	Where data can be accessed
<b>Safe outputs</b>	Publication restrictions



# From open to restricted access





# Prepare to archive

When you have finished working with your research data, you should...

- Evaluate your digital research files to identify data  from debris 
- Ensure that data is stored safely, in a suitable file format, and accompanied by adequate and self-explanatory documentation (e.g., README)
- Digitise **non-digital research data** whenever possible  
(Guidance <https://www.openaire.eu/non-digital-data-guide>)

**Researchers** are responsible for:

**15. Ensuring that non-digital forms of research data created after the commencement date of this policy are promptly digitised where possible and managed in accordance with this policy. The original non-digital research data are to be securely destroyed following digitisation unless there is a requirement for retention, in which case researchers will ensure appropriate storage arrangements.**



# Retention requirements



DMP

Sharing and access / After the project / Data retention and deletion

- What is the minimum retention period?
- During this period, the data should be **archived** on University-managed storage and non-digital data held locally.

## Research data retention requirements at the University of Auckland

- Minimum **6** years from completion of the research activity
- Health data – **10** years
  - Clinical trial data or routinely collected health data used for research
  - Until children turn 26 years
- Patent – **20** years from date of filing
- Ethics approved – check
- Community or heritage value

**Research resources**

Research data / Research data policies ... / [Research data retention](#)

# Research data storage

Remove from...



CREATE & COLLECT

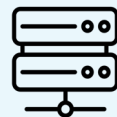
ANALYSE & INTERPRET

AFTER THE PROJECT

Research data  
"Primary copy"

University-managed research data storage

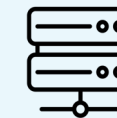
Research Drive



Dropbox for Researchers



Long-term  
storage  
(retention)



Locations  
where data  
collection &  
analysis happen  
(Temporary)

University-approved  
data capture



Compute-adjacent  
storage (VMs, HPC)



Secure Research  
Environment



Data  
repository



# When researchers leave...

**Ensure appropriate arrangements are made if researchers require AND have rights to continued access to research data after leaving the project or moving to another research organisation/institution.**

## **Handover of research data management, including:**

- Updating the Data Management Plan (DMP)
- Off-boarding meetings to discuss arrangements for ensuring ongoing access
- Update all agreements and ethics approvals

## **May also include:**

- Data Transfer Agreement to transfer the research data to another institution (where permitted by ethics approvals and agreements),
- Establishing external collaborator status for the departing staff member or student where ongoing access to research data held at the University is required and permitted.

## **What happens when a PhD candidate hands in their thesis?**

- Students generally retain 'ownership' of data created for postgraduate thesis submission.
- Ethics restrictions regarding where, and for how long, sensitive data is retained still apply.

# Data sharing and archiving

## Share data responsibly

- Aim to share data “as open as possible, as closed as necessary”
- Use frameworks like **Five Safes** to determine how sensitive data can be shared safely
- Use agreements to define and control data access where needed
- Retain data in line with **policy requirements**, and plan for long-term preservation



**What haven't we discussed?  
(Add your questions into the chat  
or raise your hand)**





Waipapa  
Taumata Rau  
**University  
of Auckland**

Questions? Get in touch...



[researchdata@auckland.ac.nz](mailto:researchdata@auckland.ac.nz)

*Research data are a treasure. Managing data is about caring for data to reflect this.*

Thank you