Data Management Planning

Dr Sarah Hopkins,
RDM Programme (DMP team)
Centre for eResearch

June 2025

Overview of this session

- What is a **data management plan (DMP)** and why should you create one?
- What are the common sections and content of a DMP? (Plus, a few **best practices** to consider)
- How can a DMP support your research practices across the **research data lifecycle**?

Activities

Introductions 

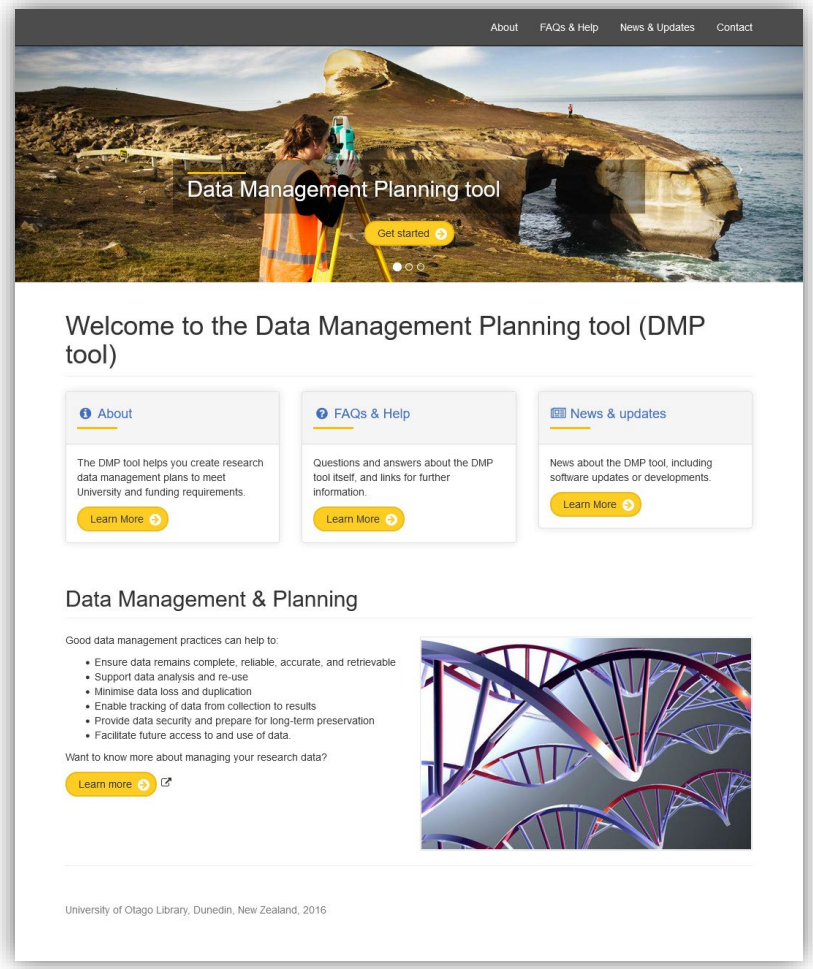
1. What is your **current role or career stage**?

2. What **research discipline** best describes your research area?

What is a data management plan?

A Data Management Plan (DMP) helps researchers to consider and document important decisions about data created or collected for a research project.

- Project specific
- Required for projects involving **sensitive** or **restricted** research data, recommended for all
- Prompts conversations, captures decisions, clarifies roles and responsibilities and helps researchers to align with University policies and processes



Checklist for a Data Management Plan from the **Digital Curation Centre**

Benefits of data management planning

Compliance & Good Research Practice

Protect

Helps organise and store data securely to **reduce the risk of data loss** or unauthorised use of project data.

Connect

Keeps project information in one place where it **can be shared & discussed** with project team members, reducing errors and uncertainty.

Structure and content

DMP tools and templates

Online DMP tools

DMP Tool

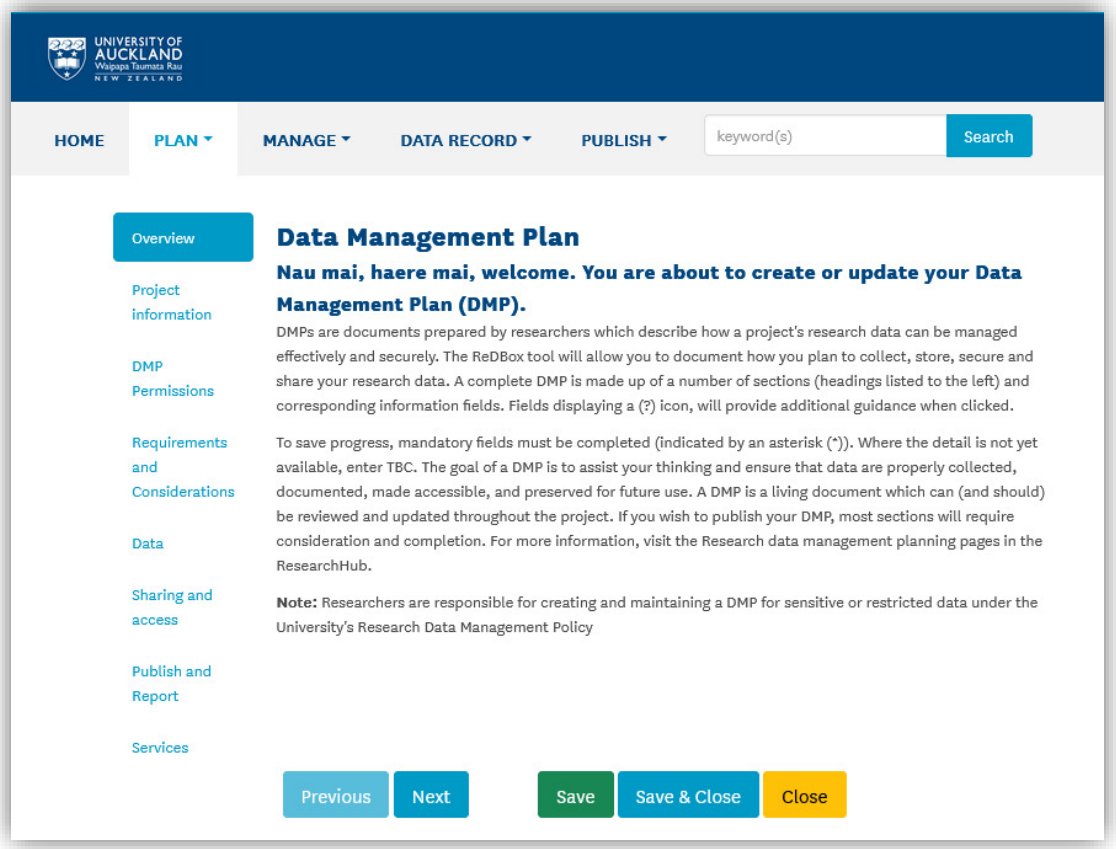
DMPONLINE

ReDBox Research Data Box

DSW DATA STEWARDSHIP WIZARD

DMP templates

Bespoke / configured DMP tools

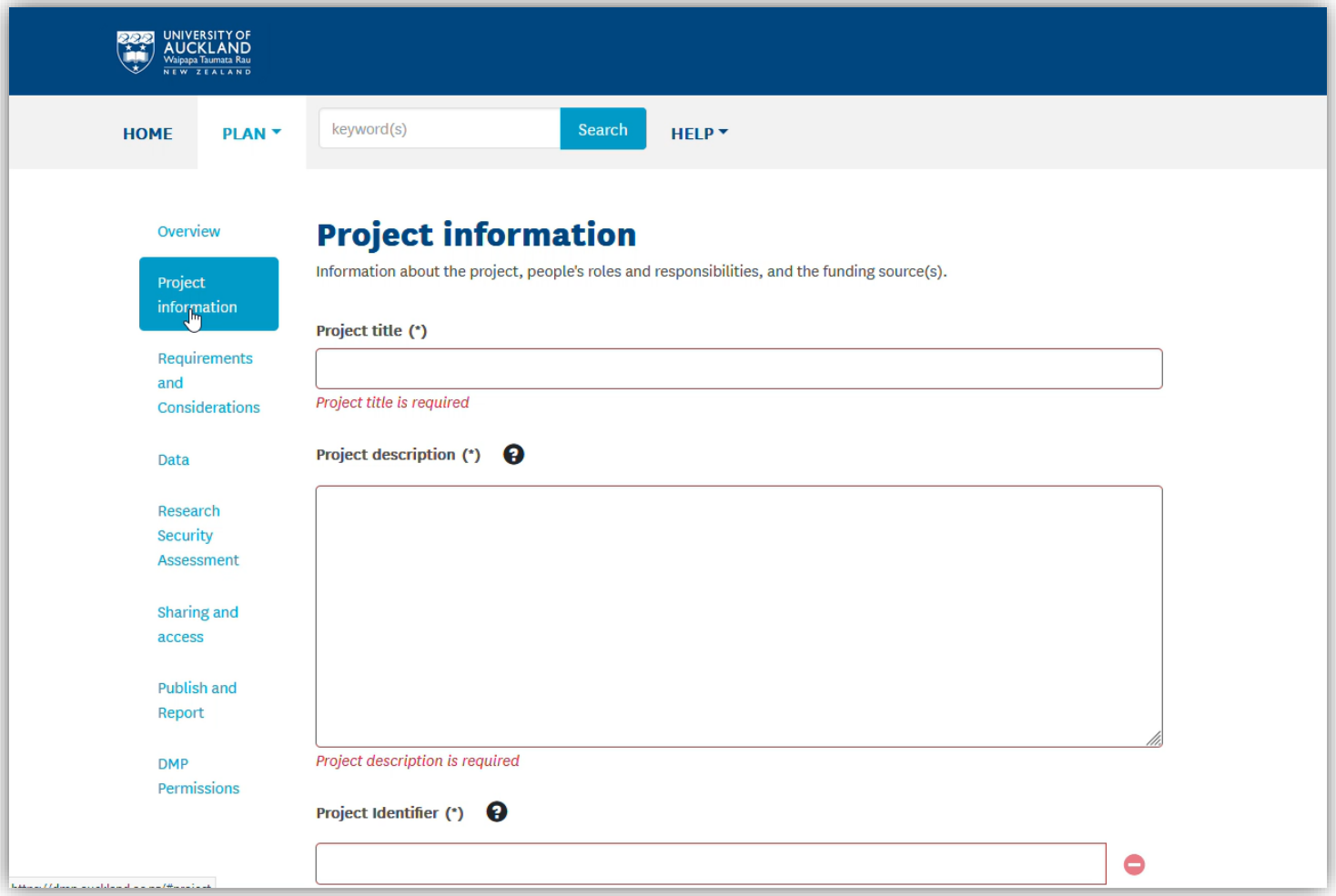


Conversations

Zoom poll…

What does your organisation recommend or support to create DMPs?

Project information



- **Purpose** of the research
- **People involved** and their **roles and responsibilities**
- **Funding**

A good DMP…

- Contains the minimal information required to identify the project.
- Allows for information to be pulled through, where possible, to prevent double entry.

Requirements

- **Data policies** applying to the project (include institutional, funder, government, publisher and other policies)
- Identifies **classification** and **sensitivity** of research data

Make sure you know about...

Institutional policies (examples)

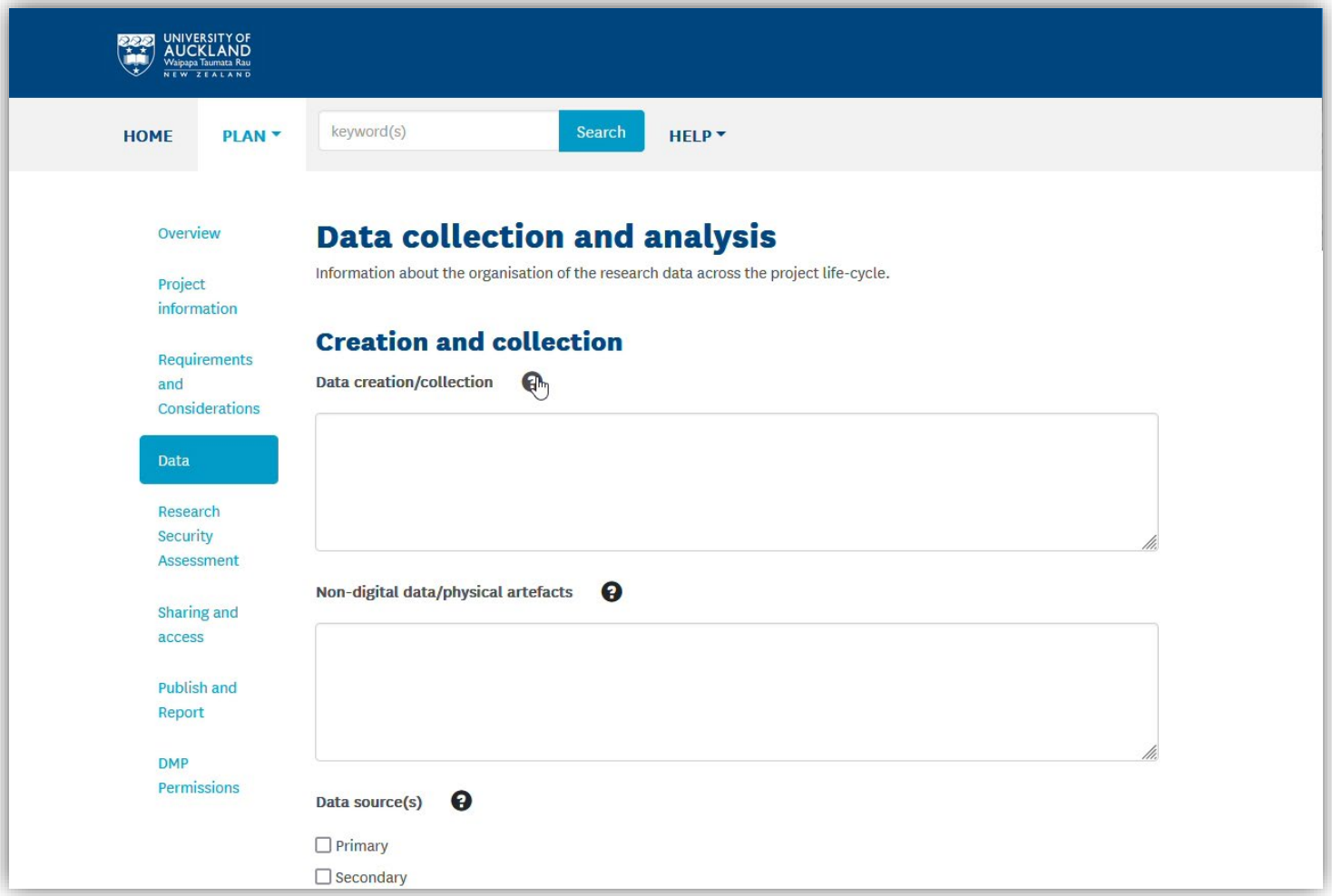
- Research data management policy
- Research integrity policy
- Intellectual property policy
- Privacy policy
- IT security & generative AI guidance

Other policies

- MBIE Open Research policy
- Trusted Research – Protective Security Requirements
- Research Charter for Aotearoa New Zealand

Next section Sharing and Access

Data



- **Describes the data** that will be created or collected and in **what format**.
- Where the data will be **stored** and how it will be **curated** (organisation, formatting and documentation)
- Sections for **data source** considerations**, metadata and documentation, storage, software and equipment.**

Considerations for data collection

- 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/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?

Considerations for 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:

1. Is it identifiable or re-identifiable data? If yes, ethics approval is required.
2. Confirm that participant consent included use for secondary analyses. If no, ethics approval and potential re-consent are required.
3. Check the quality of the data.
4. Check licenses and understand how the data can be used
5. If you reuse data, cite it.

Example | Pilot study

Data creation

DMP Data / Creation and collection / Data creation/collection

Data creation/collection

Identifiable data:

- Name and contact details of participants (entered into REDCap for duration of project)
- Interview audio files, containing potentially identifiable information (.mp3, WAV) ~30-90 mins per session; ~10-100 MB per file depending on quality = 1-5 GB for 20-30 sessions.

De-identified data:

- Online survey data (raw + processed + final dataset)

  Raw data will be collected using REDCap. At the end of data collection, a complete file from REDCap will be downloaded and saved as an 'original' and unchanged version of the data as a .csv file. A copy of the data will be saved and used for statistical analyses.
- Interview transcripts x40 (docx, txt, pdf) ~5000-10,000 words per transcript = ~ 2-5 MB
- Coding and thematic analysis files (NVivo - .nvp or Excel - .xlsx) <100MB total
- App usage data obtained from external app vendor (raw + processed; .csv)
- Participant metadata, including demographic info, interview date (excel; .csv) = <1MB

Digital research data storage

When selecting appropriate research data storage, you will need to take into account:

- Are you collecting personal (identifiable) data?
- Who needs access?
- How much storage space do you need?
- How will the data be protected against loss?
- How will you ensure the data is secure?
- What does your funder and/or organisation require?

Be aware of the 3-2-1 rule for backups

At least **3** copies

Using at least **2** different storage media

At least **1** copy offsite

Project specific

People/access

Legal, ethics etc. conditions

Files, incl. README & DMP

Security classification

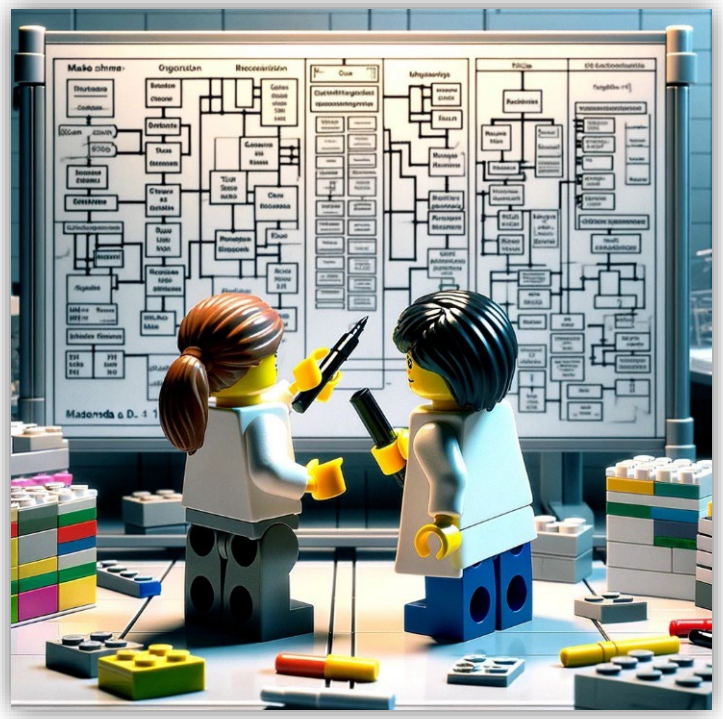
Governance

Retention period

What metadata do you need?

Metadata is a list of information you expect will be needed for the data to be read and interpreted in the future.

- Ensures data can be shared, discovered and reused
- Facilitates reproducibility and scientific integrity

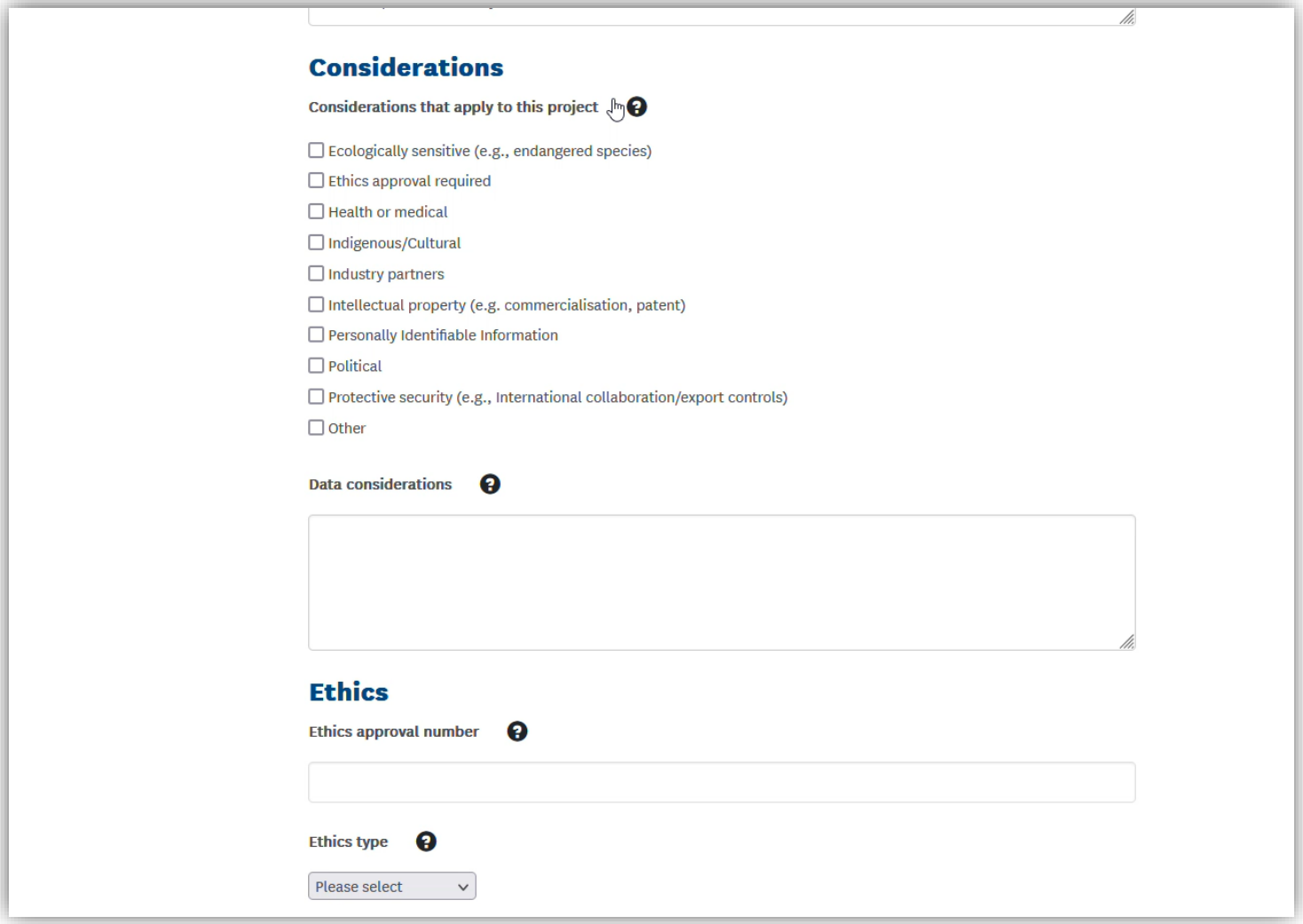


Metadata across the research data lifecycle:

- **Who** created the data?
- **What** does the data file contain?
- **When** were the data created?
- **Where** were the data created?
- **Why** were the data created?
- **How** were the data created?

Considerations

DMP **Requirements and Considerations** / **Considerations** / ...



Considerations

Considerations that apply to this project

- [ ] Ecologically sensitive (e.g., endangered species)
- [ ] Ethics approval required
- [ ] Health or medical
- [ ] Indigenous/Cultural
- [ ] Industry partners
- [ ] Intellectual property (e.g. commercialisation, patent)
- [ ] Personally Identifiable Information
- [ ] Political
- [ ] Protective security (e.g., International collaboration/export controls)
- [ ] Other

Data considerations

Ethics

Ethics approval number

Ethics type

Please select

- Identifies **legal, ethical, data sovereignty**, **IP** and other considerations for the research data.
- **Ethics**
  - Consent, protection of privacy, access controls
- **Data sovereignty**
  - Plans to consult and enable rights of stakeholders and indigenous peoples involved in the research
- **Contracts and copyright.**

Ethical considerations

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

Informed consent

- Has consent been obtained for data preservation, sharing and possible reuse?

Privacy considerations

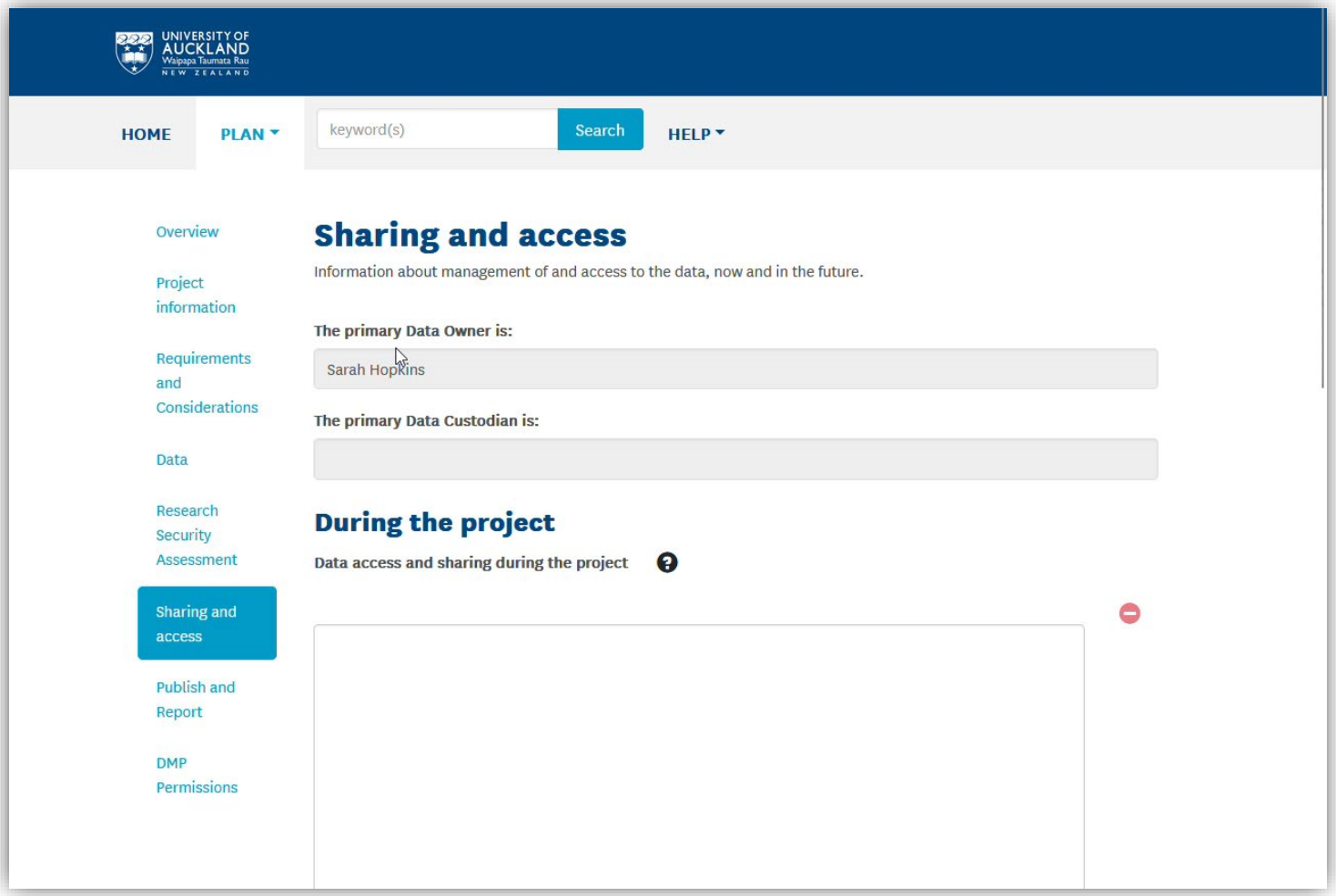
- Are personally identifiable information being collected? What processes will you use to de-identify data to ensure confidentiality? What is the risk of re-identification?

Access controls, including storage and transfer of data

- Where and for how long will data be kept (including when collected)? With whom, how and for what purpose can it be shared? How will access be restricted, and on whose authority will this be controlled? What are the conditions of data sharing, including applying for an appropriate license?

Next section Publish and Report

Sharing and Access



- Who will own and have **access to the data**.
- How will the data will **preserved and shared** for reproducibility and/or reuse.
- **Retention and disposal** procedures and provisions.

Archive and retention

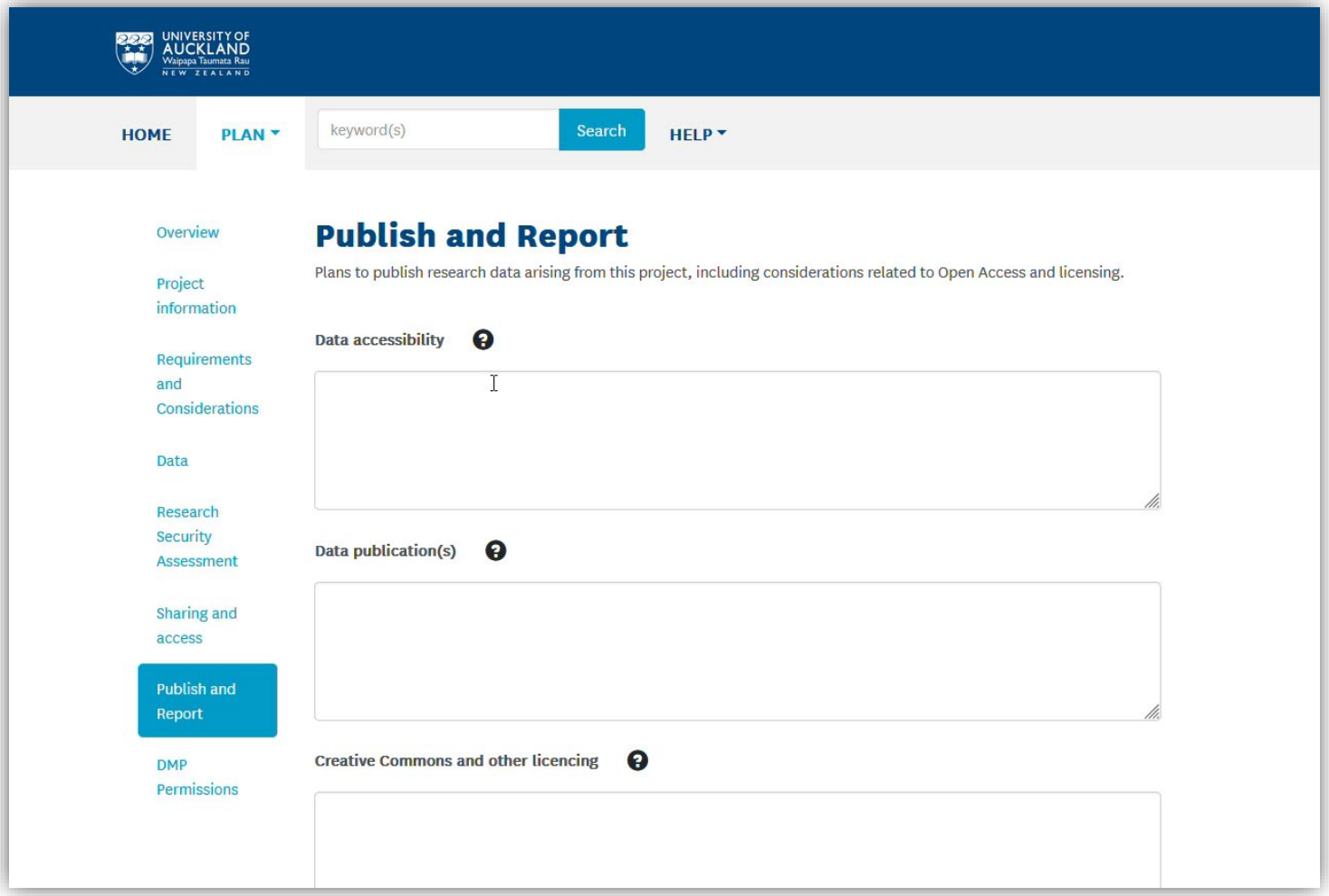
When you have finished working with your research data, you should…

- 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)

Research data retention

- What is the minimum retention period? (for transparency and reproducibility, compliance with regulations and ethical requirements, verification of research findings)
- During this period, the data should be **archived** on secure organizational storage and non-digital data held locally.

Publish and Report



- Describes how the data and/or metadata will be made **discoverable and shared**.
- Provides name of **data repository**, data catalogue or registry where data and/or metadata will or could be shared.
- States the **license** under which data may be publicly shared.

Am I enabling FAIR data principles?

F	A	I	R
Findable	**Accessible** (others know how to access)	**Interoperable**	**Reusable**

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

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

Australian Antarctic Data Centre

Data management and spatial data services

Menu | Search | Search | 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.

Describe access conditions = **Accessible**

Temporal Coverages

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

Spatial Coverages

Latitude		Longitude	
Northernmost:	24.567	**Westernmost:**	100.147
Southernmost:	-54.985	**Easternmost:**	137.95

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 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 when using these data.

Project

ISO Topic

- BIOTA
- CLIMATOLOGY/METEOROLOGY /ATMOSPHERE
- OCEANS

Dataset Language

- ENGLISH

Orignating Centre

- JARE

Dataset Progress

- COMPLETE

IDN Node

- 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, ARE 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

View Metadata Record | Download Dataset

Metadata Entry ID: chlorophyll_65-02

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

Contents

Resource	ize
LICENSE	
README	KB
chlorophyll_65-02.csv	KB
chlorophyll_65-02.xml	KB

4 records

Apply a license
= **Reusable**

Metadata is with data
= **Interoperable**

Use open file formats,
= **Interoperable**

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 when using these data.

Balancing FAIR and CARE
(or other constraint to publishing data)

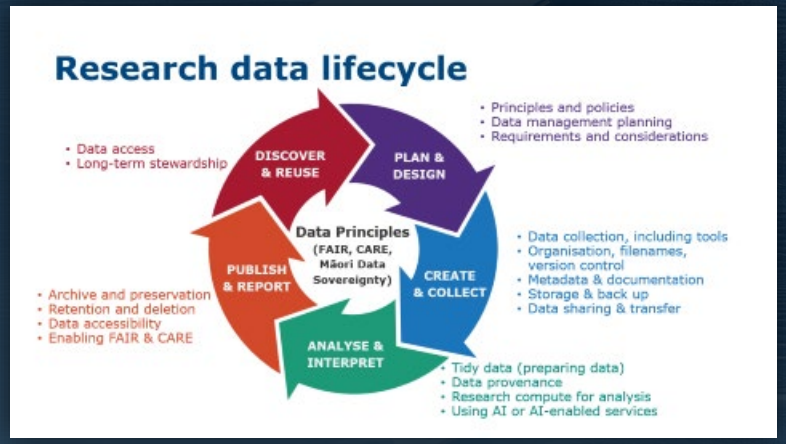
Practical steps:

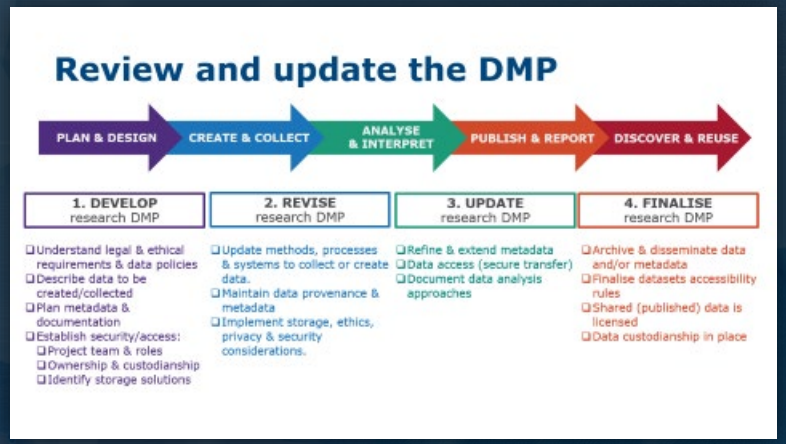
1. Publish a **descriptive or metadata-only record**
2. Create a **mediated access process**
3. Use a **data sharing agreement**
4. Produce a **data availability statement** linking data DOI to research outputs

Carroll, S.R., Herczog, E., Hudson, M. *et al.* Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Sci Data* **8**, 108 (2021). https://doi.org/10.1038/s41597-021-00892-0

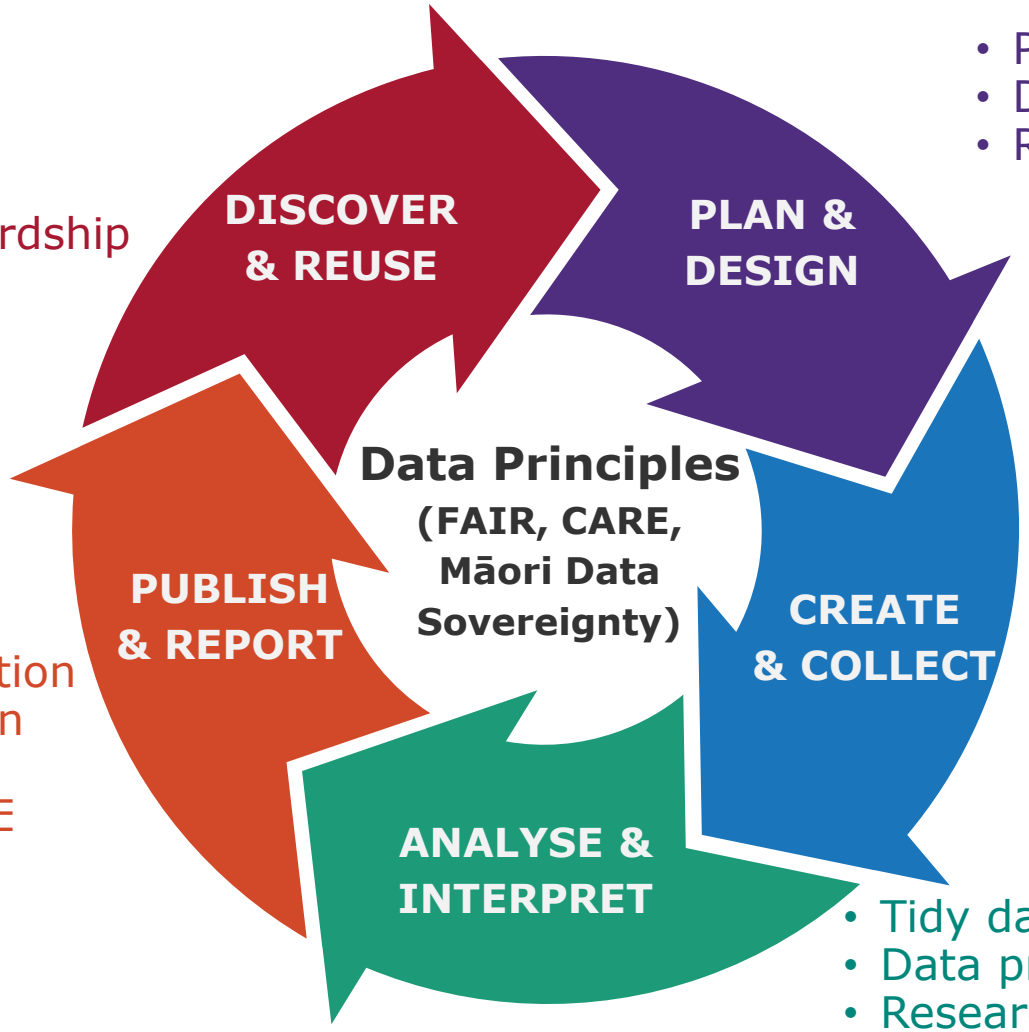
DMP across the research data lifecycle

"DMPing" is an ongoing active process across the RDM lifecycle.
Here are four times having a good quality DMP can save you time and energy.





Research data lifecycle



- Principles and policies
- Data management planning
- Requirements and considerations

- Data collection, including tools
- Organisation, filenames, version control
- Metadata & documentation
- Storage & back up
- Data sharing & transfer

- Tidy data (preparing data)
- Data provenance
- Research compute for analysis
- Using AI or AI-enabled services

- Archive and preservation
- Retention and deletion
- Data accessibility
- Enabling FAIR & CARE

- Data access
- Long-term stewardship

Review and update the DMP

1. DEVELOP research DMP

- ❑ Understand legal & ethical requirements & data policies
- ❑ Describe data to be created/collected
- ❑ Plan metadata & documentation
- ❑ Establish security/access:
  - ❑ Project team & roles
  - ❑ Ownership & custodianship
  - ❑ Identify storage solutions

2. REVISE research DMP

- ❑ Update methods, processes & systems to collect or create data.
- ❑ Maintain data provenance & metadata
- ❑ Implement storage, ethics, privacy & security considerations.

3. UPDATE research DMP

- ❑ Refine & extend metadata
- ❑ Data access (secure transfer)
- ❑ Document data analysis approaches

4. FINALISE research DMP

- ❑ Archive & disseminate data and/or metadata
- ❑ Finalise datasets accessibility rules
- ❑ Shared (published) data is licensed
- ❑ Data custodianship in place

Good DMPs support funding success

Increases competitive advantage for research funding.

Document RDM practices to help you through **ethics approval**.

Demonstrate your awareness and understanding of RDM best practices to funders and to obtain ethics approval

DMPs connect project team members

Plan and document project roles and responsibilities with the project team so that everyone is on the same page.

DMPs connect project team members

Keeps project information in one place so that you can onboard new project team members efficiently and accurately

DMPs support discovery and reuse

Return research data to participants, where appropriate and agreed

Enable **secondary reuse** of data (**FAIR** data principles, increase impact & visibility)

Comply with institutional and other **data retention requirements**

Finalise your DMP so that research data can be retained, returned, shared and/or undergo secure disposal

Pātai / kōrero about DMPs

Some questions to consider and discuss

Conversations

Zoom poll...

1. Have you completed a **data management plan (DMP)** before?

2. In what situations did you find a research DMP **most helpful**?
   - At the start of the project to plan research data management
   - To obtain ethics approval
   - To discuss plans and responsibilities with the project team
   - During the study when I needed to refer to decisions made, metadata or data organisation or other details
   - At the end of the study to plan for post-study retention requirements and/or sharing for reuse

Conversations

The best strategy to support me to create and maintain a research DMP is to:

- Have lots of training available
- Train and support local data stewards (or advisors) to individually support researchers
- Provide lots of self-directed resources
- Make it easy – integrate systems and workflows better so that they share information (e.g., title, abstract, project team members, funding)

Zoom poll…

Rank the strategies from most important > least important

In the Zoom chat…

Share any comments/thoughts or raise your hand to speak

Conversations

DMPs should be mandated (compulsory) for all research projects…

Zoom poll…

Choose whether you agree or disagree with this statement

In the Zoom chat…

Share any comments/thoughts or raise your hand to speak

Conversations

DMPs should be a compulsory provisional/first year goal for all doctoral students…

Zoom poll…

Choose whether you agree or disagree with this statement

In the Zoom chat…

Share any comments/thoughts or raise your hand to speak

Waipapa Taumata Rau
University of Auckland

Questions? Get in touch…


s.hopkins@auckland.ac.nz


researchdata@auckland.ac.nz

Thank you