Managing Research Data

Principles and practices

Laura Armstrong & Dr Sarah Hopkins
Centre for eResearch

July 2024

Valuing inclusion

Ensuring all individuals feel respected, accepted, and valued.

- Manaakitanga - show respect, care and support for others
- Whanaungatanga – foster an environment where all in our community have a place
- Kotahitanga – built unity and partnership
- Kaitiakitanga - recognise our responsibilities as kaitiaki (guardians) to protect and respect our environment, traditions, knowledge, culture, languages and other taonga.

https://www.auckland.ac.nz/en/on-campus/life-on-campus/code-of-conduct.html

1. Introductions

…and tell us what question you are hoping to have answered in this workshop.

What are research data?

The evidence that underpins the answer to a research question and can be used to validate findings regardless of its form (e.g., print, digital, or physical).

Data or artefacts, cultural taonga, research evidence, and digital representation of a physical item used in research.

samples, interviews, images, surveys, observations, audio/visual recordings, medical records, maps, instrument data, spreadsheets, bibliographies, manuscript annotations…

Is software/code research data?

Maybe, it depends.

Software/code may be the focus or by-product of the research.

Does the software or code support:

- **Reproduction** - enabling others to find, access & run exactly same software, inputs & computational environment to verify/validate your results.
- **Replication** - using *similar* inputs, tools, environments, to arrive at mostly the same outputs and conclusions to justify the results.

The Turing Way is a handbook to reproducible, ethical and collaborative data science.

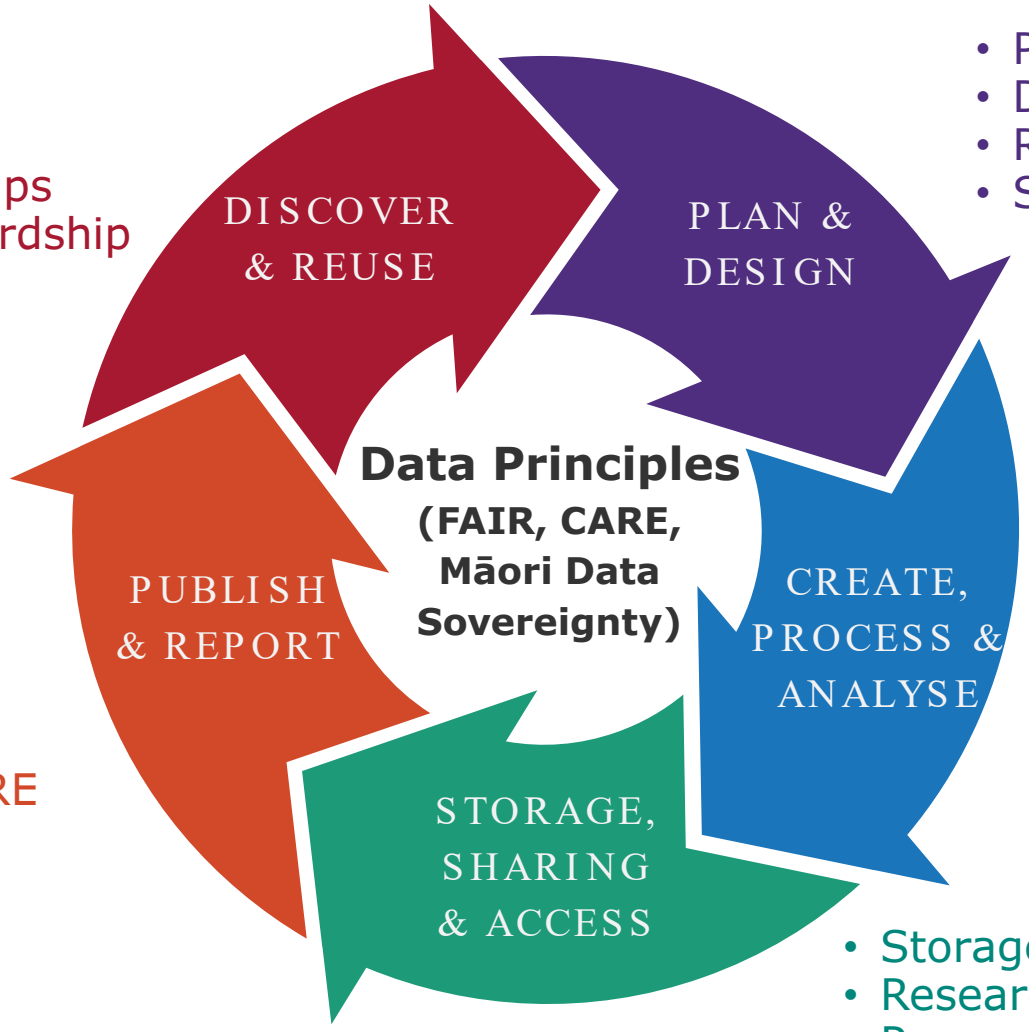
What is research data management?

Process of planning and undertaking the collection, organisation, management, storage, backup, preservation and sharing of data before, during and after the project.

- Understanding
- Integrity
- Collaboration
- Impact
- Other environments

A curated list of RDM resources for researchers and organisations (Mannheim University, Germany)

Research data management



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

- Data collection tools
- Organisation, filenames, version control
- Tidy data (preparing data)
- Computational resources
- Documentation - README

- Storage
- Research data vs. research debris
- Preservation & archive
- Retention/deletion

- Dissemination & copyright
- Data availability
- Enabling FAIR & CARE

- Data access groups
- Long-term stewardship

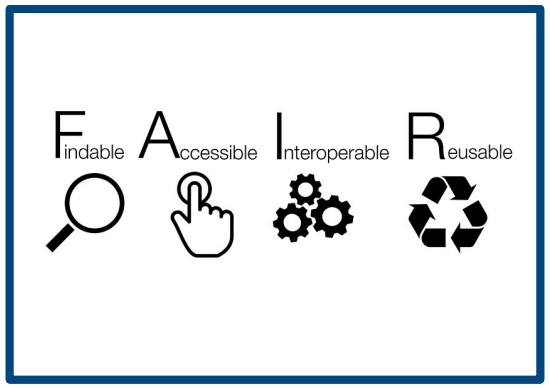
Plan and Design

Plan for data management throughout the research data lifecycle.

- Principles and policies
- Data management planning
- Requirements and considerations
- Managing sensitive data

A culture shift for RDM

What has driven changes over the last decade?



Expectations

- "…maximise knowledge output from funding"
- Reduce **duplication** and increase **reuse** of data
- **FAIR principles** (Findable, Accessible, Interoperable, Reusable)
- **Reproducibility**

Technology

- Increased capacity to generate, store & work with very large datasets,
- Cloud computing
- Lower computing costs
- **Digitisation**

Data sovereignty

"... becoming a **Māori Data Sovereignty** organisation" and **CARE** - data principles

Privacy & security

- **Legal**, **ethical** and **protective security**
- Managing risk

Researchers want to do the right thing but want clarity/direction on best practices, available services & support.

Policies related to research data

National

- Funder and publisher policies, e.g., MBIE Open Research policy
- Trusted Research – Protective Security Requirements guidance
- Research Charter for Aotearoa New Zealand
- Royal Society Professional Code of Conduct

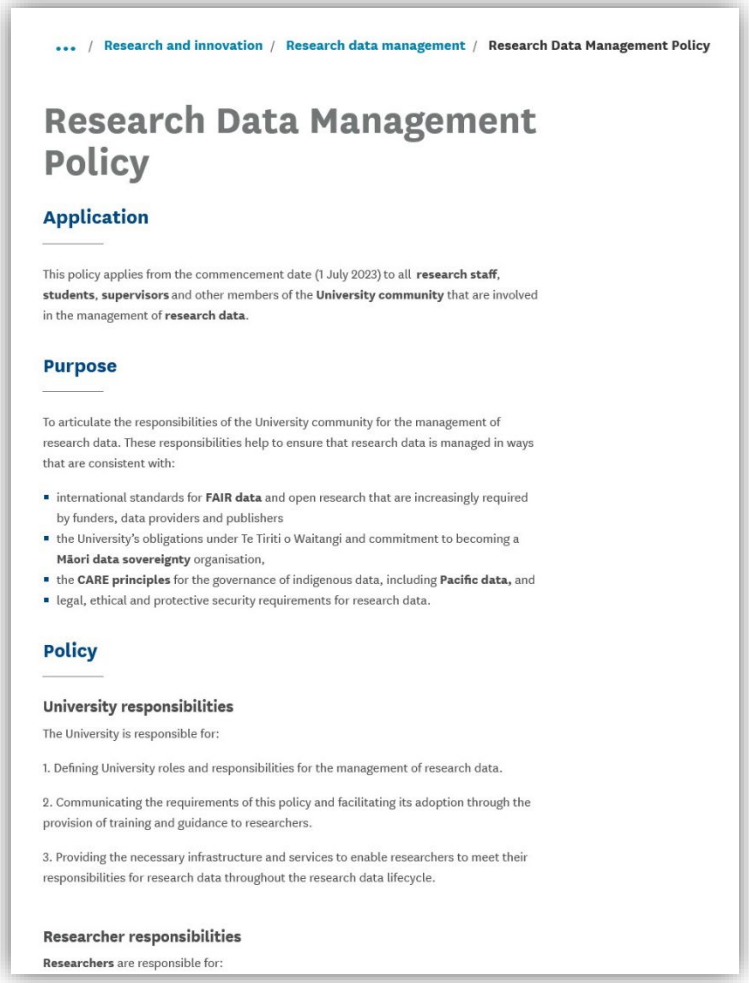
Institutional (University of Auckland, as an example)

- Research Data Management Policy and guidance
- Research Code of Conduct (new Research Integrity Policy under review)
- IP created by staff and students Policy (under review)
- Privacy Policy
- Māori Research Policy (in development)

What does an RDM policy look like?

A **research data management policy** sets out researcher and research student responsibilities for the management, preservation and sharing of research data.

- Common elements:
  - Data management planning
  - Support researchers to be “as open as possible, as closed as necessary”
  - Enable long-term stewardship
- Usually paired with guidance on how to implement, e.g., apply governance and storage on a project-by-project basis



... / 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

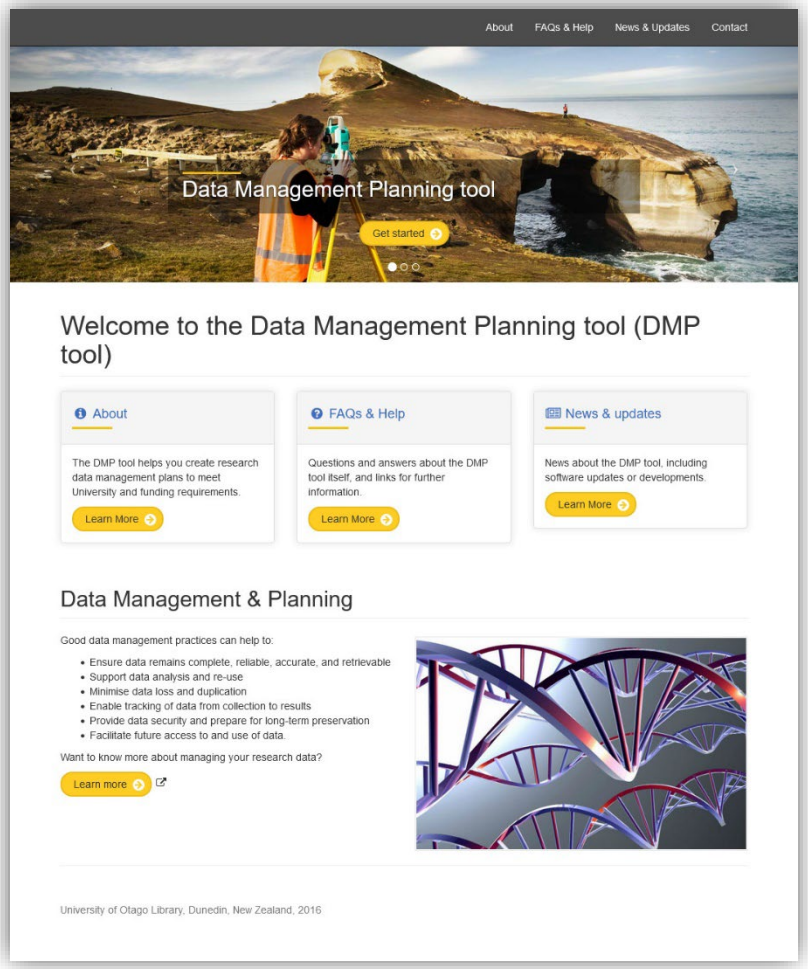
Researchers are responsible for:

2. What policies and guidance on RDM are you required to follow?

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 approach
- Project specific
- Prompts conversations, captures decisions, clarifies roles and responsibilities and helps you to align with University policies and processes

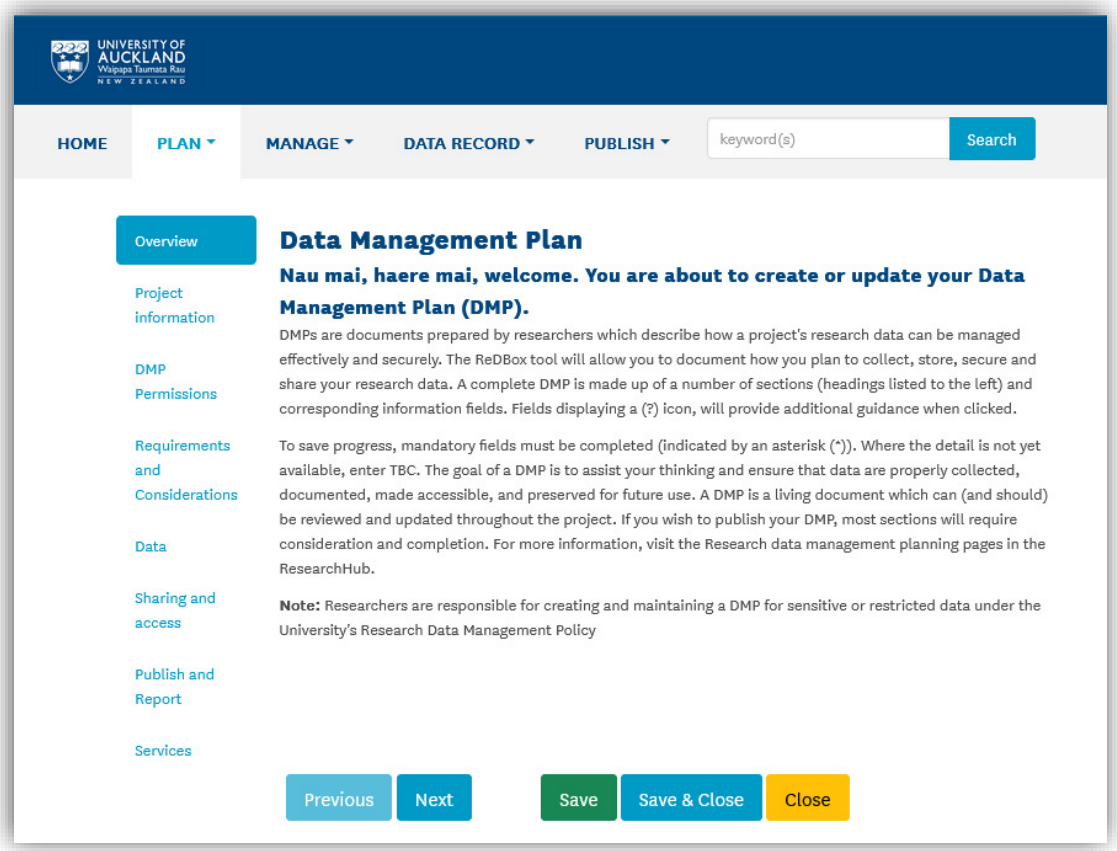


Checklist for a Data Management Plan from the **Digital Curation Centre**

Data Management Planning

Document decisions about:

- **Project information** - purpose, people, roles & responsibilities
- **Requirements and considerations** - ethical, legal, sovereignty, funder, etc.
- **Data** - collection, organisation & sharing, including access restrictions
- **Sharing and access** - storage locations, retention/ deletion, long term governance
- **Publication** - enabling FAIR



ResBaz
RESEARCH BAZAAR

Data Management Planning
Wednesday 10 July, 9am-10am

Data governance

…enables answers to

- Is the data **reliable** (quality) for a given use?
- What is its **value** and associated **risks**?
- Who has **access**?
- Where is the data (**location**), and what happens if it is moved or changed?
- Who and what processes are keeping data **protected**?
- Is access **control** appropriate (security classification, and authority to control)?

…involves

- people - **roles and responsibilities**
- **technologies, systems and tools**
- **processes and controls** to support the consistency, integrity, usability and access
- **policies, procedures and standards** to support clarity and compliance - legal, regulatory, ethics, sovereignty
- **data quality**, including metadata and information security

3. Data management planning

Legal, ethical, sovereignty constraints

Researchers should ensure that **legal**, **ethical**, **data sovereignty**, **protective security and commercial constraints** relating to research data are considered prior to data collection and adhered to throughout the research data lifecycle.

- Constraints will inform how data are collected/gathered, stored, shared and governed throughout the data lifecycle.
- Even if none of the constraints mentioned above apply to specific project, there will likely be values-based considerations.

Legal constraints

Privacy principles covered by the Privacy Act 2020

National Ethics Advisory Committee

Principles for the safe and effective use of data and analytics, 2018

Stats NZ & the Privacy Commission

International: GDPR & HIPAA

Further details >> **https://www.privacy.org.nz/assets/New-order/Resources-/Publications/Guidance-resources/Privacy-Act-2020-information-sheets-full-set.pdf**

De-identifying data

Identifiable

Data that directly or indirectly identifies an individual or business.

Individual

Name	Hēni
Gender	Female
DOB	31/01/1985
Address	28 My Road Postcode 6012 Wellington

Business

Name	Puzzles
Type	Paper Stationery Manufacturing
Employees	34
Expenditure	$398,000

De-identified

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

Individual

Name	*Unknown*
Gender	Female
DOB	1985
Address	Postcode 6012 Wellington

Business

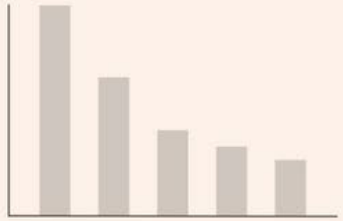
Name	*Unknown*
Type	Manufacturing
Employees	30 - 40
Expenditure	$398,000

Confidentialised

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

Individual

Name	*Unknown*
Gender	Female
Age	30 - 40 years
Address	Wellington



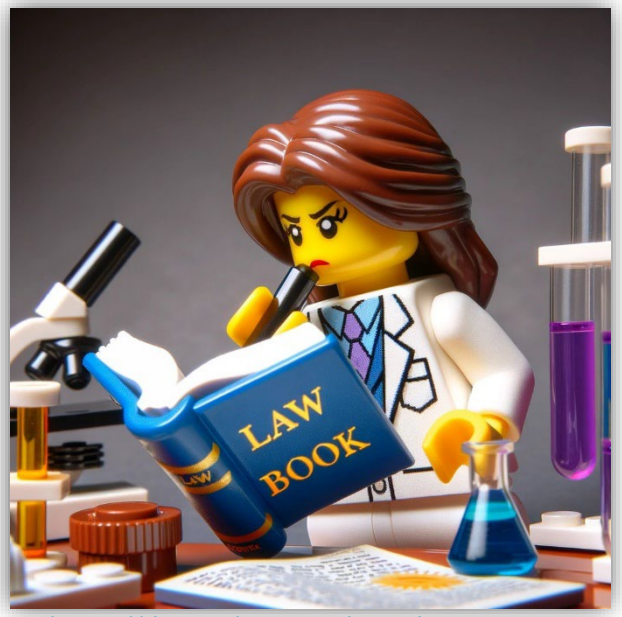
Business

Name	*Unknown*
Type	Manufacturing
Employees	10 - 100
Expenditure	Under $500,000

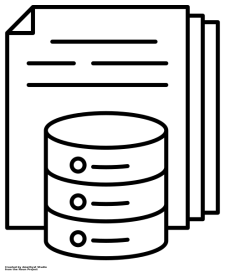
Degrees of identification in data – From data.govt.nz (Creative Commons Attribution 4.0 International)

More legal considerations

- Contracts
- Intellectual Property / commercialisation
- Copyright of incoming data
- Export controls



https://doi.org/10.5281/zenodo.11147887



My fancy data

External data source(s)

New data set

Data sharing agreements

Where research is shared with third parties, intellectual property rights, plans for data preservation and sharing, and legal responsibilities should be agreed in a formal **data sharing agreement**.

- Formal contract sets out:
  - sets out and agrees on the purpose of the data sharing and expected outcomes
  - describes what will to happen to the data at each stage (how the data will be transferred and stored)
  - sets standards and helps all the parties to be clear about their respective roles (access restrictions, stewardship)
- Parties demonstrate their accountability to legal, ethical, data sovereignty, etc. requirements.

More information on data sharing agreements from the UK Information Commissioner's Office

Ethical considerations

Ethics applications, including pre-screening review, **benefit** from being able to demonstrate consideration of:

- What data needs to be collected/gathered, for what purpose, and from whom?
- How will you protect the identify of participants if required?
- How, by whom and when will data collection/gathering occur?
- Where and for how long will data be kept?
- With whom, how and for what purpose can it be shared?
- How will access be restricted and on whose authority will this be controlled?
- Have you gained consent for data preservation and sharing?

Reconsider expectation to *destroy* or store *indefinitely* digital research data.

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

Consider early as these impact the funding application, planning ethics application, consent, storage, metadata, sharing, and publishing of research findings and data throughout the research data lifecycle.

Global Indigenous Data Alliance

ResBaz RESEARCH BAZAAR

Māori Data Sovereignty
Thursday 11 July, 3pm-4pm

Sensitive data

What data might need more **security**, **protection** or **access restriction**?

- Data from or about - human participants, health/clinical providers, environment, indigenous people, culture, politics, industry, defence/national security, animals … ?
- Data may become sensitive - unintended capture or context of use, or attitudes may change over time and place.
- Sensitive data is common.
- Data classification helps to understand constraints & meet requirements.

Consider impact on how data is governed, captured, stored, moved and shared, and future stewardship, etc.

4. Research data classification

Sensitive data

Available resources:

- Sensitive data guide (ARDC)
- Data confidentiality principles & methods (data.govt.nz)
- NEAC National Ethical Standards
- HDEC template for data/tissue management plan
- Te Ira Kāwai, the Auckland Regional Tissue Bank guidance for collection and storage of human tissue for research

Local resources and support?

- Ethics and Integrity
- Māori Responsiveness
- REDCap
- Genomics support
- Health Research / Clinical Trials

5. What do you find most challenging about managing sensitive data?

Collect, process & analyse data

Plan and document data collection & processing so that the end result could be interpreted, replicated from the raw data and reused by others.

- Data collection software & tools
- Organisation, filenames, version control
- Data processing – Tidy Data
- Computational resources
- Documentation – metadata & README

Data collection software & tools

- **Use University-supported software or tools, whenever possible.**
  These applications are often security tested and approved ('Authority to Operate')
- 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?

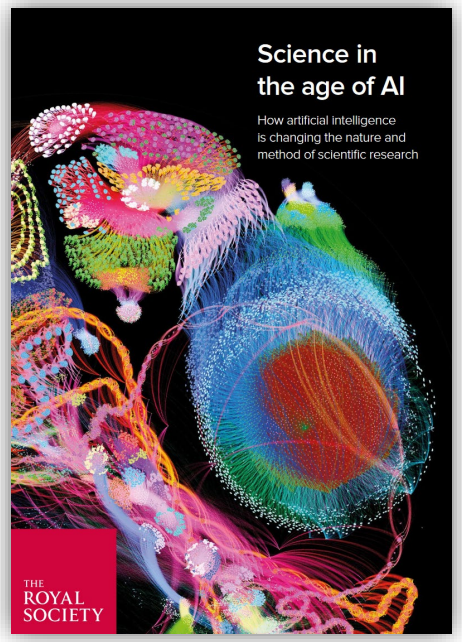
Activity	Approved tools (examples)
Participant surveys	REDCap Qualtrics
Transcription of audio files	MS Word MS Teams Zoom

ResBaz
RESEARCH BAZAAR

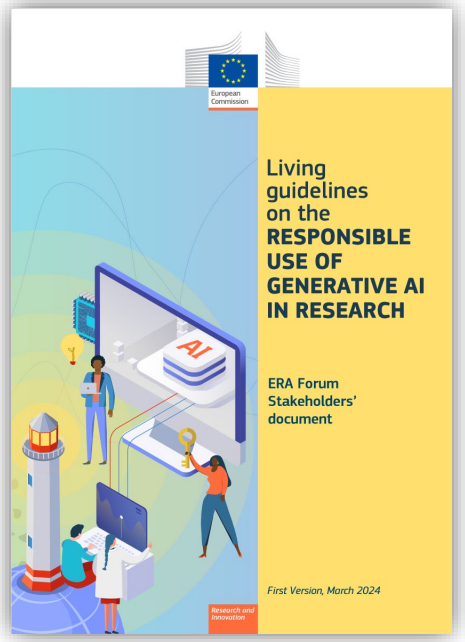
[Digital transcription tools](#)
Thursday 11 July, 10am-11am
[Intro to Qualtrics](#)
Thursday 11 July, 3pm-4pm
[An overview of REDCap](#)
Thursday 11 July, 11am-12pm

Using AI or AI-enabled services

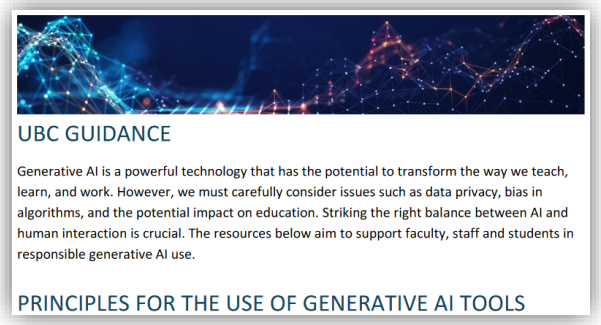
- What are appropriate uses?
- What types of data are appropriate as AI inputs?
- Where does the data go and what happens to it?
- Do I have participant consent?
- How do I cite my use of AI?
- Am I required to follow an institutional policy?



[Royal Society (2024)]



[EU guidelines (March 2024)]



[University of British Columbia Guidance]

6. Use of AI with research data

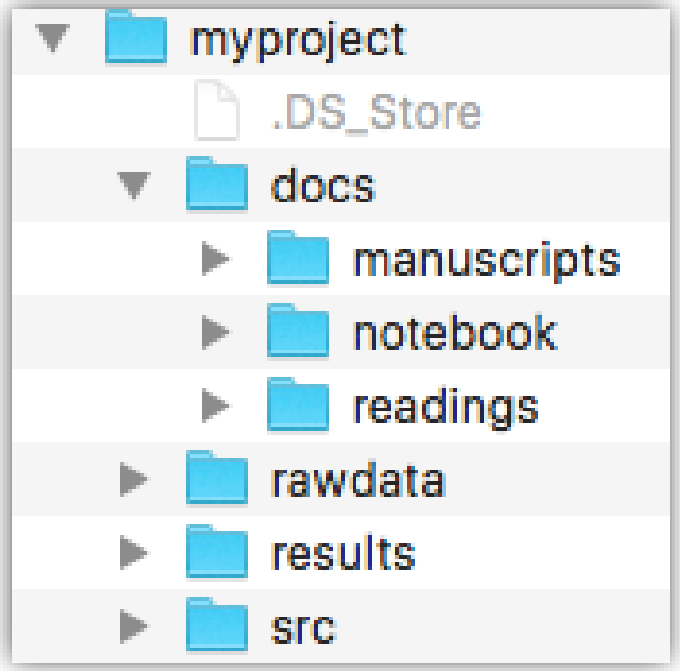
Organisation

- What types of data, what file formats?
- How much? *raw + (analysed * no. analyses) + (backup * redundancies)*
- Will it grow/accumulate?
- Will it change over time?
- How will you organise it?
- Where will you store it?
- How will you document /describe it?

Further information: Organising data (UK Data Service)

Project-based Organisation

- 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



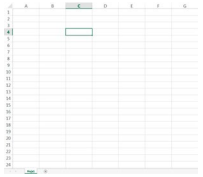
Further information: Project structure by Danielle Navarro

File Naming

- 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]

Version Control

Originally for software development, now used by the data science community.

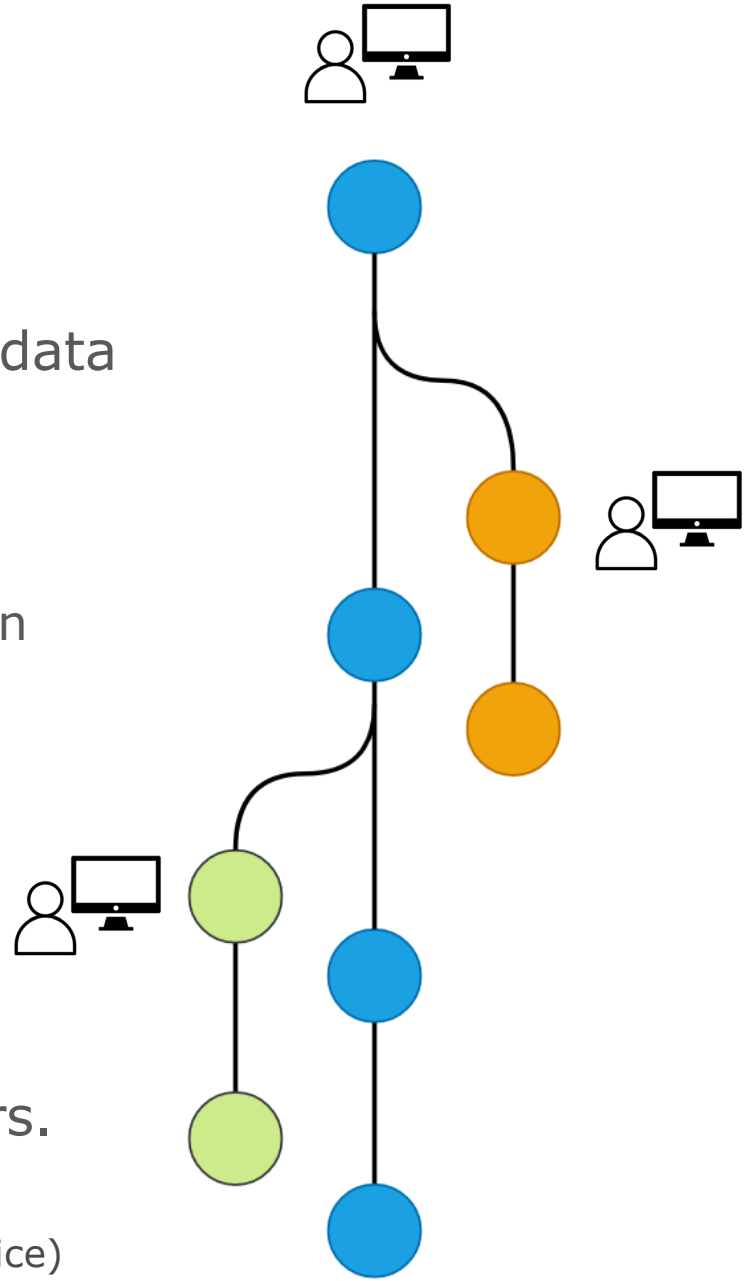
Git = Version Control System

- Manage changes to plain text files (code documents) in an ordered way.
- Commit changes to a repository.
- Branching model.

GitHub = Remote repository

- Files are pushed up to a remote/cloud repository.

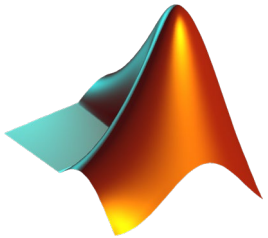
A modern research workflow used by many researchers.



Further reading: Version control strategy and best practice (UK Data Service)

Data Provenance

- Always keep a copy of the raw data untouched
- Have a separate copy which is your tidy dataset
- Keep a record of your ‘recipe’ (exact steps taken) to get from raw to tidy data
- Keep contextual information in a README



Tidy Data

1. Every column is a variable.

2. Every row is an observation.

3. Every cell is a single value.

- 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

Identify resources for RDM

Researchers are responsible for **identifying resources** required for the management of research data, including storage, compute, unique or special infrastructure or governance requirements.

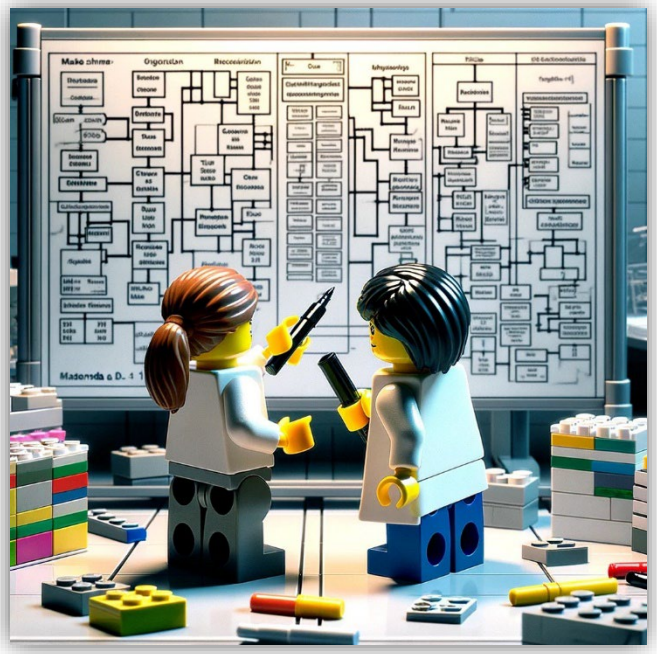
- Large or non-standard storage?
- Additional computer power? (access to GPUs, virtual machines, high performance computing [HPC], machine learning)
- Special infrastructure requirement? (software, hardware, technical expertise)
- Governance? (access committee, advisory groups)

Documentation and metadata

Ensuring research data is accompanied by appropriate **documentation** and **metadata** will help you and your collaborators to understand what you did and why. This supports the **reproducibility** of findings and is **good research practice.**

Documentation:

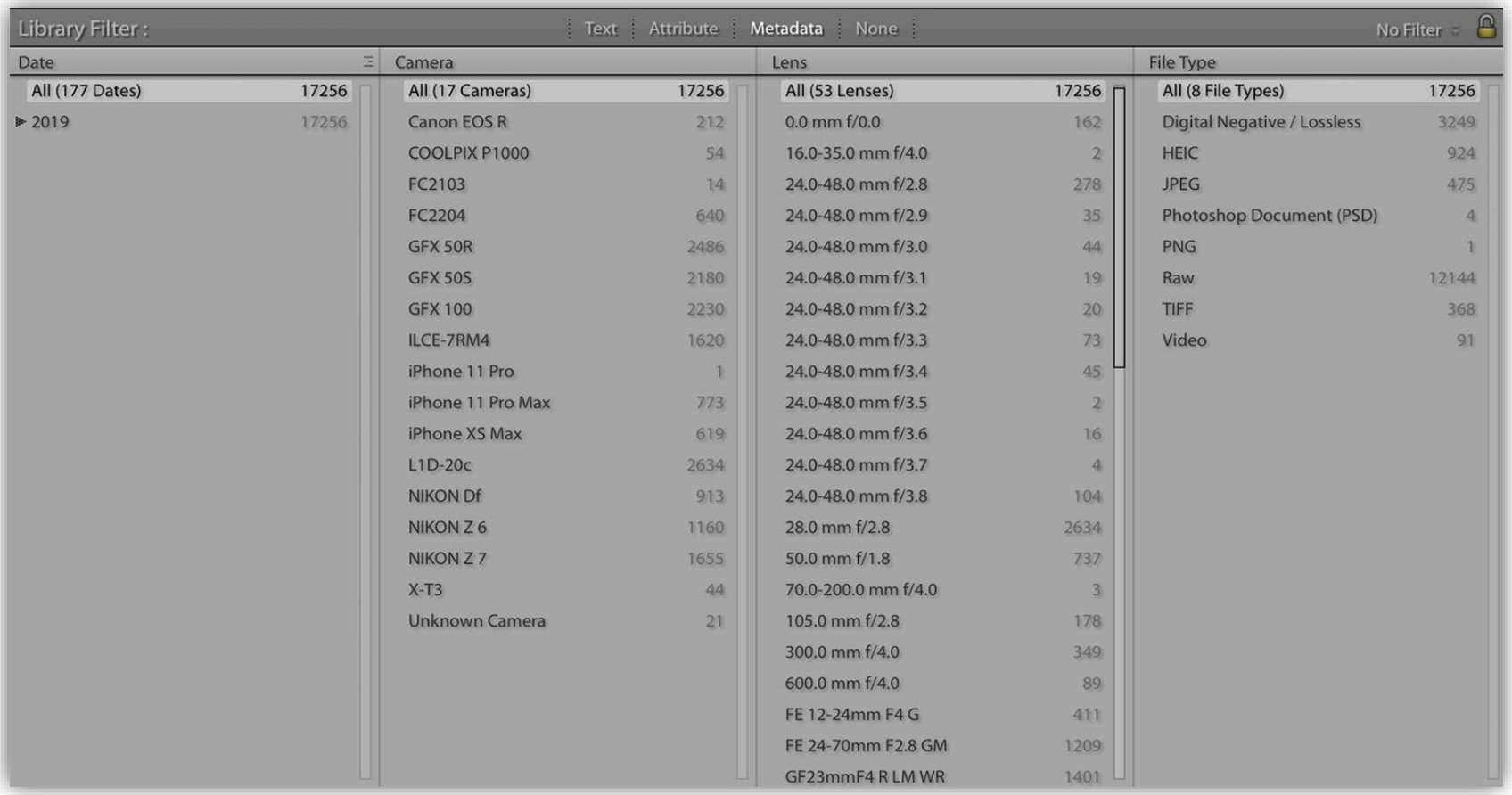
- README
- Data Dictionary
- Codebook
- Metadata

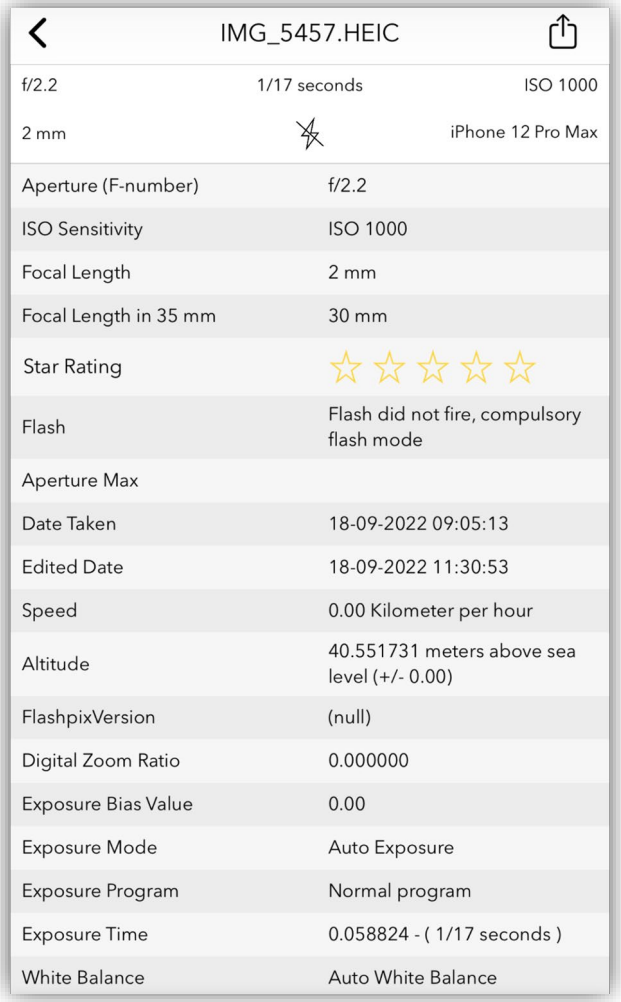


Metadata enables:

- Collaboration
- Governance, including data sovereignty
- Impact
- Application of data principles

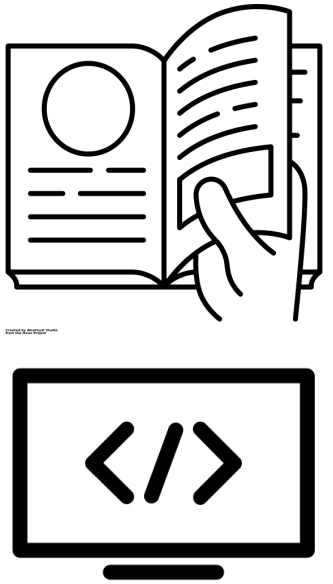
Data about the data - metadata



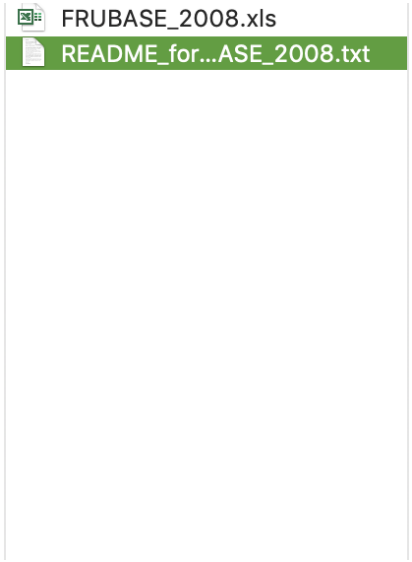


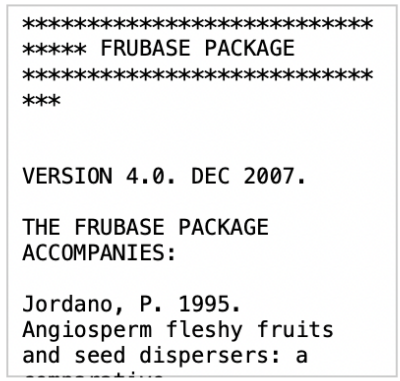
Metadata enables

- Collaboration at different levels
- Discovery (Findable)- human and machine readable
- Access and governance (CARE, MDSov)
- Understanding (Interoperation and Reuse) README.txt, electronic lab notebooks, data dictionary, code book
- Preservation (Reuse)



Documentation/README





README_for_FRUBASE_2008.txt

Plain Text Document - 16 KB

```
THE FRUBASE PACKAGE ACCOMPANIES:

Jordano, P. 1995. Angiosperm fleshy fruits and seed dispersers: a comparative
analysis of adaptation and constraints in plant-animal interactions. American
Naturalist 145: 163-191.

It contains a copy of the main data file exactly as used for this paper, as
well as other accompanying files (see below).

Taxonomic arrangement follows:
Cronquist, A. (1981). An integrated system of classification of flowering
plants. Columbia University Press.

Nomenclature follows Stevens, P. F. (2001 onwards). Angiosperm Phylogeny
Website. Version 8, June 2007. http://www.mobot.org/MOBOT/research/APweb/.
This scheme follows: A.P.G. [= Angiosperm Phylogeny Group] II. 2003. An update
of the Angiosperm Phylogeny Group classification for the orders and families
of flowering plants: APG II. Bot. J. Linnean Soc. 141: 399-436.

Plant names and names of higher taxonomic categories have been checked with:
Mabberley, D.J. 1987. The plant-book. A portable dictionary of the higher
plants. Cambridge University Press, Cambridge, UK.

Please, contact me if you have suggestions, find errors, inconsistencies, or
any other bug in the file. As well, please let me know about your uses of this
data and send manuscripts and reprints when available. I'll be happy to help
you in any case, as far as I can.

I am periodically updating this data base since I started writing my PhD
thesis more than 20 years ago. Thus, I'd like to receive suggestions for new
data sources and provide updated versions to those interested.

Please, use these data files for peaceful purposes, enjoy doing science with
them as I have enjoyed writing the paper quoted above, and learn as much as
you can with them. They are the result of splendid work by many people working
with plant-frugivore interactions and are embedded in papers reporting very
interesting results, descriptions and discussions on these interactions;
please read them.
********** CONTENTS **********
All files are plain ASCII text files, with the exception of SUMMARY and
FRUBASE.xls.
Those with data have TABs as their field delimiters so they can be readily
imported in any statistical package or spreadsheet program. The FRUBASE.txt is
readily imported by any spreadsheet application. Please, contact me if you need
the files formatted in other ways (e.g., my original SAS datasets, or EXCEL
worksheets).

1. README.txt - This file. Including a description of the variables and a
   listing of the literature sources with the numeric codes.
2. Summary.doc - A summary file (originally intended to appear as an Appendix
   in my 1995 paper) summarizing mean values for the main families and genera
   in the data base. This is a Microsoft WORD (version 6.0) file, which can be
   read directly either by the Mac or Windows versions of the program.
3. REFS.txt - A long list with the source reference used for each species in
   the data file. The file is TAB delimited and has a header line with variable
   names: FAMILY, GENUS, SPECIES, NEWREF, and REFERENCE (authors and year).
4. FRUBASE.txt - The data file itself. Missing data are indicated by dots (.).
   The file is TAB delimited and has a header line with variable names as in
   the list below. The file is sorted by FAMILY, GENUS, and SPECIES names, in
   ascending order.
5. FRUBASE.xls - The data file itself, now in Excel format for spraedsheets.
   See (4).
**********

Variable names and descriptions in FRUBASE

      CL        Class
      SCL       SubClass
      ORD       Order
      FAM       Family
      GEN       Genus
      SP        Species
      REF       Reference number - This is my maintenance code for updates.
      NEWREF    New Reference number - These are the refs numbers in the
                                       files REFS and SUMMARY.
      FAMLAB    Family Label -  An 8-character label for family.
      GENLAB    Genus Label -   An 8-character label for genus.
      SPLAB     Species Label - An 8-character label for species.
      COD       Species code -  A 5-character code for the species.
      DISPCAT   Disperser type category - BIRDS, MIXED, MAMMALS.
      DISP      Disperser type - Finer categorization. Not yet completed.
                                 Needs revision.
```

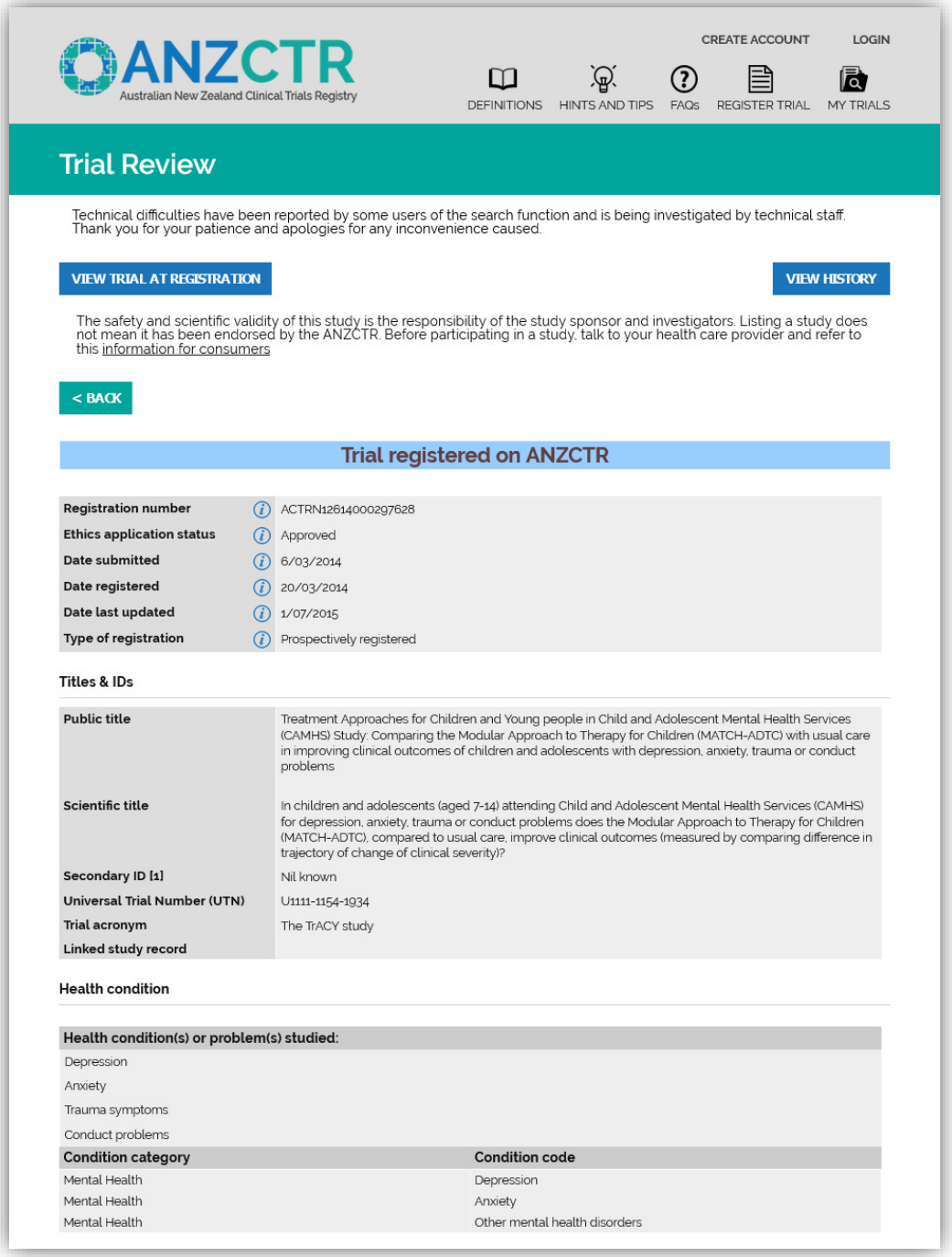
README guide

Metadata standards

Standard, structured and formalised fields of metadata that enable people and machines to share and comprehend (reproduce) datasets. Eg.Clinical trials, EML, CellML

- Clinical trials registry (*pictured, right*)

- Research Data Alliance (UK)
- FAIR sharing
- Digital Curation Centre (UK)



ANZCTR
Australian New Zealand Clinical Trials Registry

CREATE ACCOUNT LOGIN

DEFINITIONS HINTS AND TIPS FAQs REGISTER TRIAL MY TRIALS

Trial Review

Technical difficulties have been reported by some users of the search function and is being investigated by technical staff. Thank you for your patience and apologies for any inconvenience caused.

VIEW TRIAL AT REGISTRATION VIEW HISTORY

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been endorsed by the ANZCTR. Before participating in a study, talk to your health care provider and refer to this information for consumers

< BACK

Trial registered on ANZCTR

Registration number	ACTRN12614000297628
Ethics application status	Approved
Date submitted	6/03/2014
Date registered	20/03/2014
Date last updated	1/07/2015
Type of registration	Prospectively registered

Titles & IDs

Public title	Treatment Approaches for Children and Young people in Child and Adolescent Mental Health Services (CAMHS) Study: Comparing the Modular Approach to Therapy for Children (MATCH-ADTC) with usual care in improving clinical outcomes of children and adolescents with depression, anxiety, trauma or conduct problems
Scientific title	In children and adolescents (aged 7-14) attending Child and Adolescent Mental Health Services (CAMHS) for depression, anxiety, trauma or conduct problems does the Modular Approach to Therapy for Children (MATCH-ADTC), compared to usual care, improve clinical outcomes (measured by comparing difference in trajectory of change of clinical severity)?
Secondary ID [1]	Nil known
Universal Trial Number (UTN)	U1111-1154-1934
Trial acronym	The TrACY study
Linked study record	

Health condition

Health condition(s) or problem(s) studied:	
Depression	
Anxiety	
Trauma symptoms	
Conduct problems	
Condition category	**Condition code**
Mental Health	Depression
Mental Health	Anxiety
Mental Health	Other mental health disorders

Storage, Sharing and Access

Store data so that they are protected against corruption and loss. Research data should be prepared for preservation or archive to substantiate research findings.

- Data storage
- Research data vs. research debris
- Preservation and archive
- Retention and deletion

Digital research data storage

Researchers should ensure that digital forms of research data are stored so that they are **protected against corruption and loss**, **secured** appropriately and **findable** by those who need to exert long-term governance or stewardship.

Storage considerations and approach

- What data are being collected/gathered?
- Are you working with confidential or sensitive data?
- Where are you required to, or have agreed to, store data?
- Who needs access, and where are they when they access?
- Does the project need granular access restrictions?
- What processing or analysis will be needed?
- What is available?

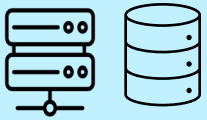
Backing up your research data

What is your backup routine?

- What data/files need to be backed up? How often? Where? By whom?
- Regularly check if data is not corrupted.
- A least 2 people should have access to the data.

Follow the 3-2-1 rule

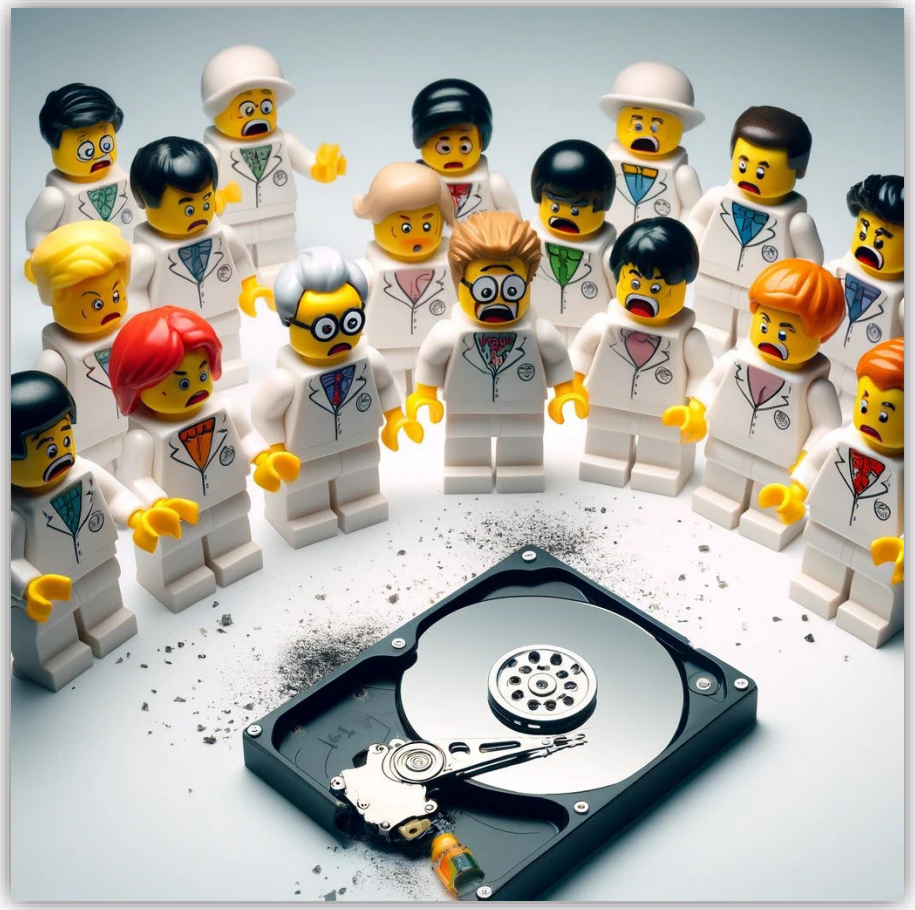
At least **3** copies



Using at least **2** different storage media

At least **1** copy offsite

NB. Network drives are backed up on tape every night in multiple physical locations.



7. Data storage

Research data storage options

Personal devices are usually not appropriate



Data capture

Self-service or managed

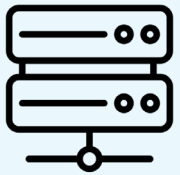
REDCap

zoom

qualtrics XM

Storage

Network storage



Dropbox (cloud storage)

Dropbox

MS OneDrive (personal), Teams, SharePoint (online)

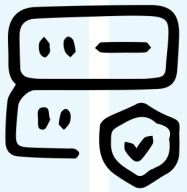
Microsoft Teams

SharePoint Online

OneDrive

Analysis & visualisation with research compute - virtual machines and HPC

Secure Research Environment

Sustainable use of storage

Data storage is **expensive** and has **environmental** costs to consider.

Tips to take control of your storage space:

1. Regularly review your files
2. Understand and identify **research data** vs. **research debris**
3. Enable future use e.g., open formats
4. Use different storage tiers/products – slower=cheaper

Preservation and Archive

Preserving or **archiving** data ensures that it lives beyond the end of a specific research project.

- What data must be retained (& destroyed?) and/or preserved?
- Archive vs. preservation (passive vs. active)
- Publishing as archive.
- What are you trying to achieve or enable?
- Where have you left your data? How long for?
- Who is responsible, who is the data steward?
- Will the future be able to open and make sense of it?

https://library.si.edu/research/best-practices-storing-archiving-and-preserving-data

Return, retention, deletion, & destruction

Researchers should ensure that research data are **returned**, **retained**, **deleted** and/or **destroyed** in accordance with legal, ethical, data sovereignty and commercial constraints.

- Is there a requirement to return research data?
- What is the minimum retention period? During this period, the data will be:
  - **Published**? Assigned DOI, licensed for re-use, access controls – if required.
  - **Archived**? Confidential and non-digital data held locally.
- Deletion of archived digital research data ✓
- Destruction of digital research data files stored on University-managed storage is not achievable in most cases. ✘

8. Retention

Publish and Report

Wherever possible, data should be shared (published) so that it can be consulted and re-used by other researchers.

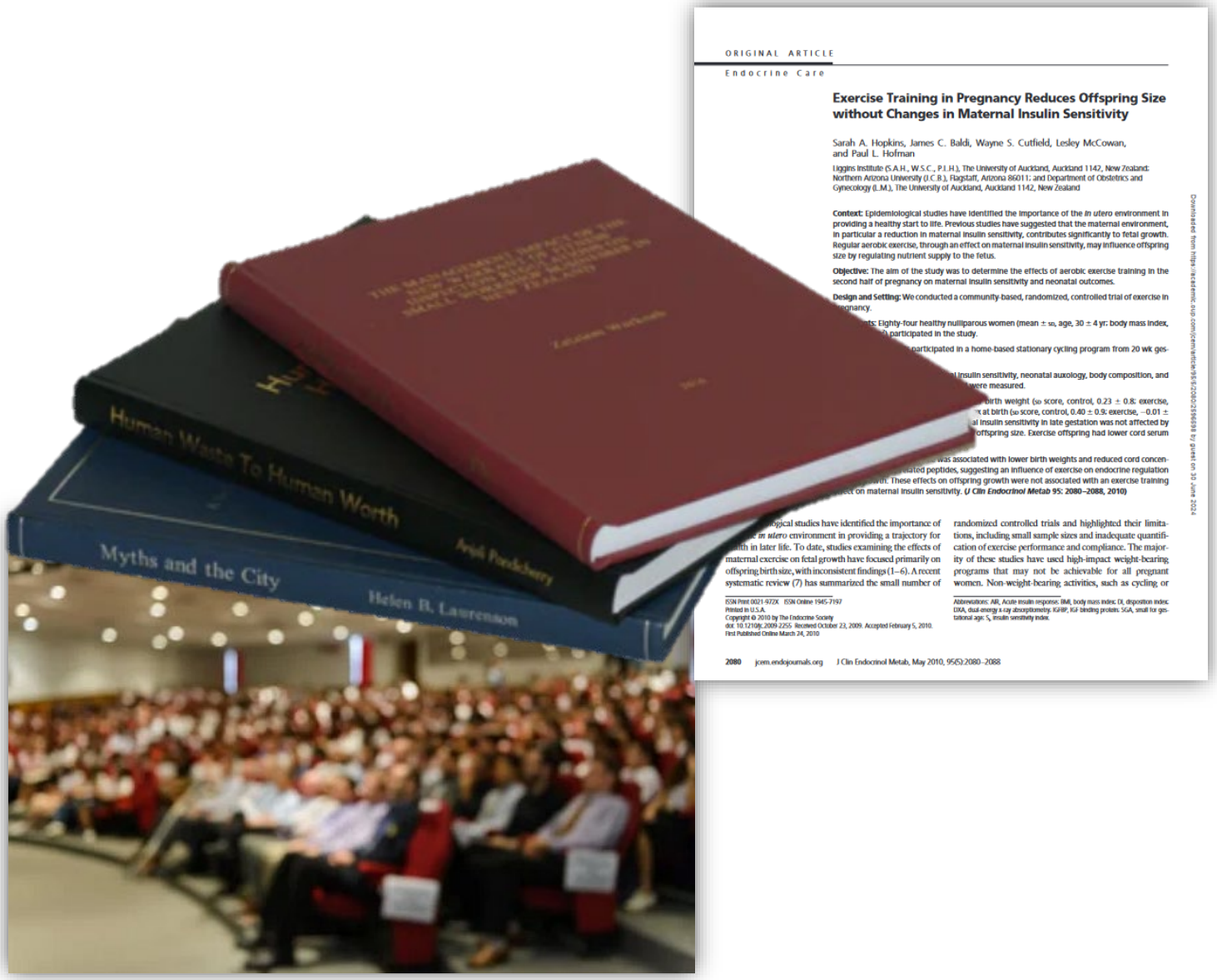
- Dissemination & copyright (rightsholders)
- Publish data & metadata
  - FAIR & CARE
  - Licenses
- Data availability

Dissemination

Research findings are usually shared with multiple audiences:

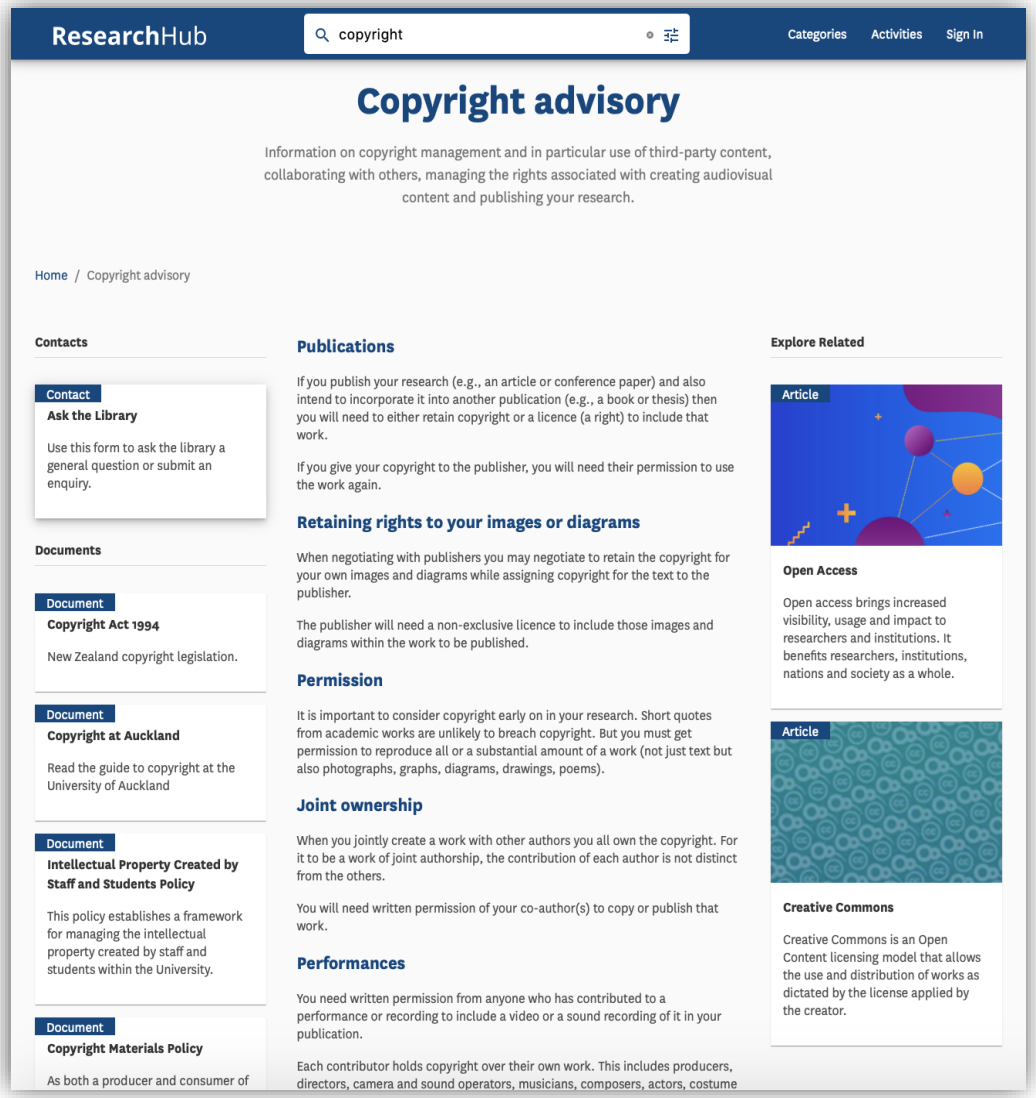
- Thesis submission
- Manuscript preparation
- Community dissemination

- Who makes decisions about what is disseminated?
- Is the sharing of **research data** a condition of publication?



Who holds the copyright?

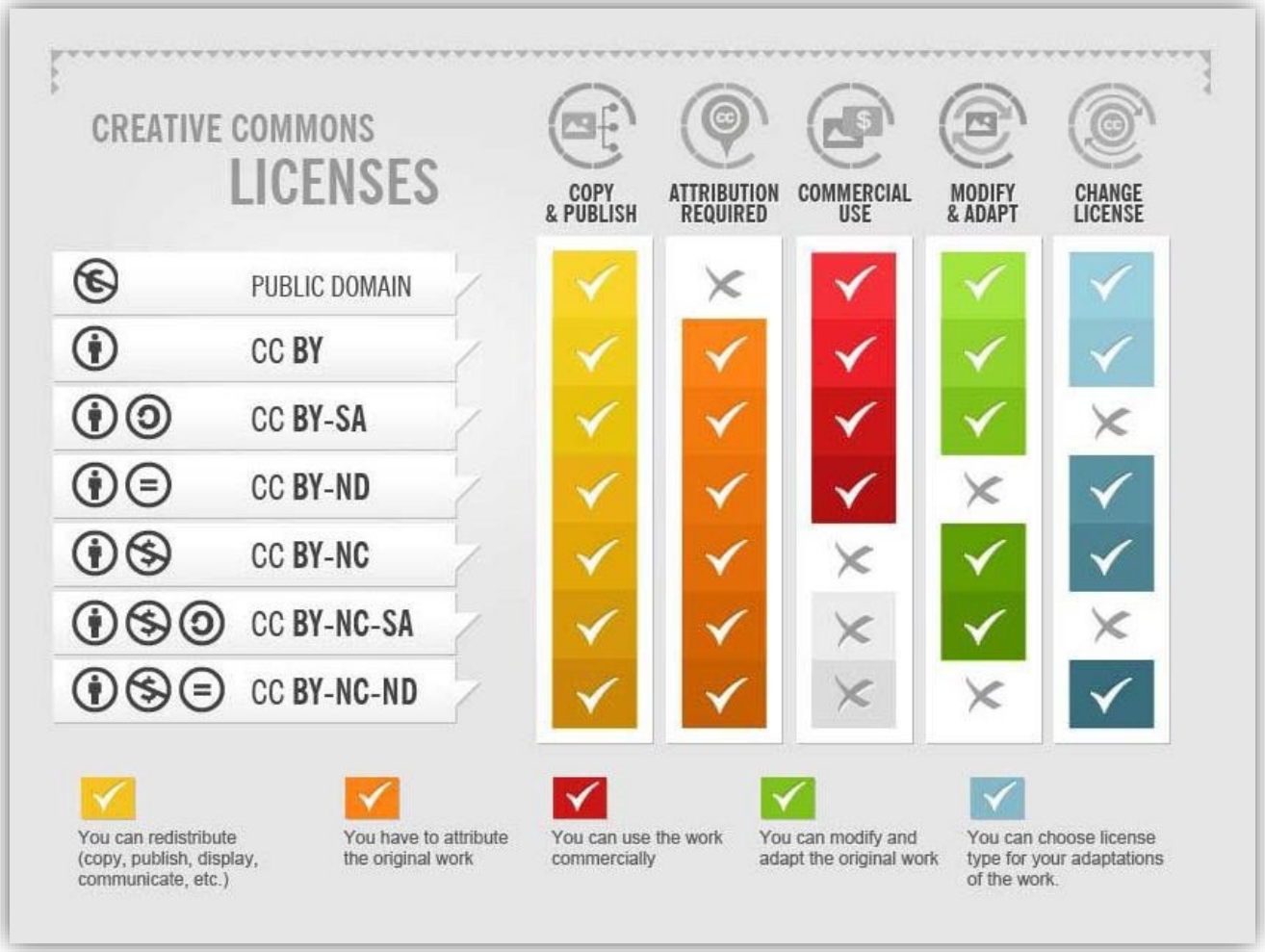
- Copyright holder can license a work
- Choose open licences
- Check rights and permissions if re-using data
- Start ownership discussions early
- Explicit copyright display
- “As open as possible, as closed as necessary”



https://research-hub.auckland.ac.nz/article/copyright-advisory

Creative Commons Licenses

- Make it easy to allow reuse of your works by others
- Creative commons licences to choose from
- Most open licences allow adaptation, remix and sharing of materials



Publish data

Researchers are strongly encouraged to **publish digital forms of research data** in a suitable research data repository unless the data cannot be published due to legal, ethical, data sovereignty or commercial constraints.

Include, where possible, a **data availability statement** in all accepted manuscripts and final accepted theses describing how and on what terms any supporting research data may be accessed.

- How do you know whether you can, or should, publish data?

F

Findable

A

Accessible
(others know how to access)

I

Interoperable

R

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

FAIR principles

Findable
Accessible
Interoperable
Reusable

Exemplar

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	
Northernmost:		**Westernmost:**	
24.567		100.147	
Southernmost:		**Easternmost:**	
-54.985		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

Platforms

Instruments

Researchers

Use Constraints

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

Metadata Revision History

2010-07-27 - record updated by Dave Connell to change URL Content Type. 2017-12-18 - record updated by Dave Connell - basic udpates.

Dataset

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	Type	Last Modified	File Size
LICENSE	File	2022-08-19	250 B
README	File	2022-08-19	3.31 KB
chlorophyll_65-02.csv	File	2022-08-19	51.67 KB
chlorophyll_65-02.xml	File	2022-08-19	16.25 KB

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: 2024-04-19

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.

C

Collective benefit

A

Authority to control their data

F A R I R

Findable Accessible Interoperable Reusable

R

Responsibility to engage respectfully with those communities

Everyone

E

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

Specific people and purpose

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

9. Data or metadata-only publishing

Data availability statements

Data Availability Statement: All data generated or analyzed during this study are included in this paper and its Supporting Information files, except the sound .wav files, which are available through Figshare (https://auckland.figshare.com/articles/media/Sound_wav_files_use/20103734; DOI: 10.17608/k6.auckland.20103734).

PLOS ONE

RESEARCH ARTICLE

Direct liquid transmission of sound has little impact on fermentation performance in *Saccharomyces cerevisiae*

Rachel Benitez[1☯], Alastair Harris[1☯], Evie Mansfield[1], Pat Silcock[2], Graham Eyres[2], Silas G. Villas-Bôas[3], Andrew Jeffs[1], Austen R. D. Ganley[1*]

1 School of Biological Sciences, University of Auckland, Auckland CBD, New Zealand, New Zealand, 2 Department of Food Science, University of Otago, Dunedin, New Zealand, 3 Luxembourg Institute of Science and Technology, Z.A.E. Robert Steichen, Luxembourg, Luxembourg

☯ These authors contributed equally to this work.
* a.ganley@auckland.ac.nz

Check for updates

OPEN ACCESS

Citation: Benitez R, Harris A, Mansfield E, Silcock P, Eyres G, Villas-Bôas SG, et al. (2023) Direct liquid transmission of sound has little impact on fermentation performance in *Saccharomyces cerevisiae*. PLoS ONE 18(2): e0281762. https://doi.org/10.1371/journal.pone.0281762

Editor: Shashi Kant Bhatia, Konkuk University, REPUBLIC OF KOREA

Received: October 5, 2022

Accepted: January 31, 2023

Published: February 17, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0281762

Data Availability Statement: All data generated or analyzed during this study are included in this paper and its Supporting Information files, except the sound .wav files, which are available through

Abstract

Sound is a physical stimulus that has the potential to affect various growth parameters of microorganisms. However, the effects of audible sound on microbes reported in the literature are inconsistent. Most published studies involve transmitting sound from external speakers through air toward liquid cultures of the microorganisms. However, the density differential between air and liquid culture could greatly alter the sound characteristics to which the microorganisms are exposed. In this study we apply white noise sound in a highly controlled experimental system that we previously established for transmitting sound underwater directly into liquid cultures to examine the effects of two key sound parameters, frequency and intensity, on the fermentation performance of a commercial *Saccharomyces cerevisiae* ale yeast growing in a maltose minimal medium. We performed these experiments in an anechoic chamber to minimise extraneous sound, and find little consistent effect of either sound frequency or intensity on the growth rate, maltose consumption, or ethanol production of this yeast strain. These results, while in contrast to those reported in most published studies, are consistent with our previous study showing that direct underwater exposure to white noise sound has little impact on *S. cerevisiae* volatile production and sugar utilization in beer medium. Thus, our results suggest the possibility that reported microorganism responses to sound may be an artefact associated with applying sound to cultures externally via transmission through air.

Introduction

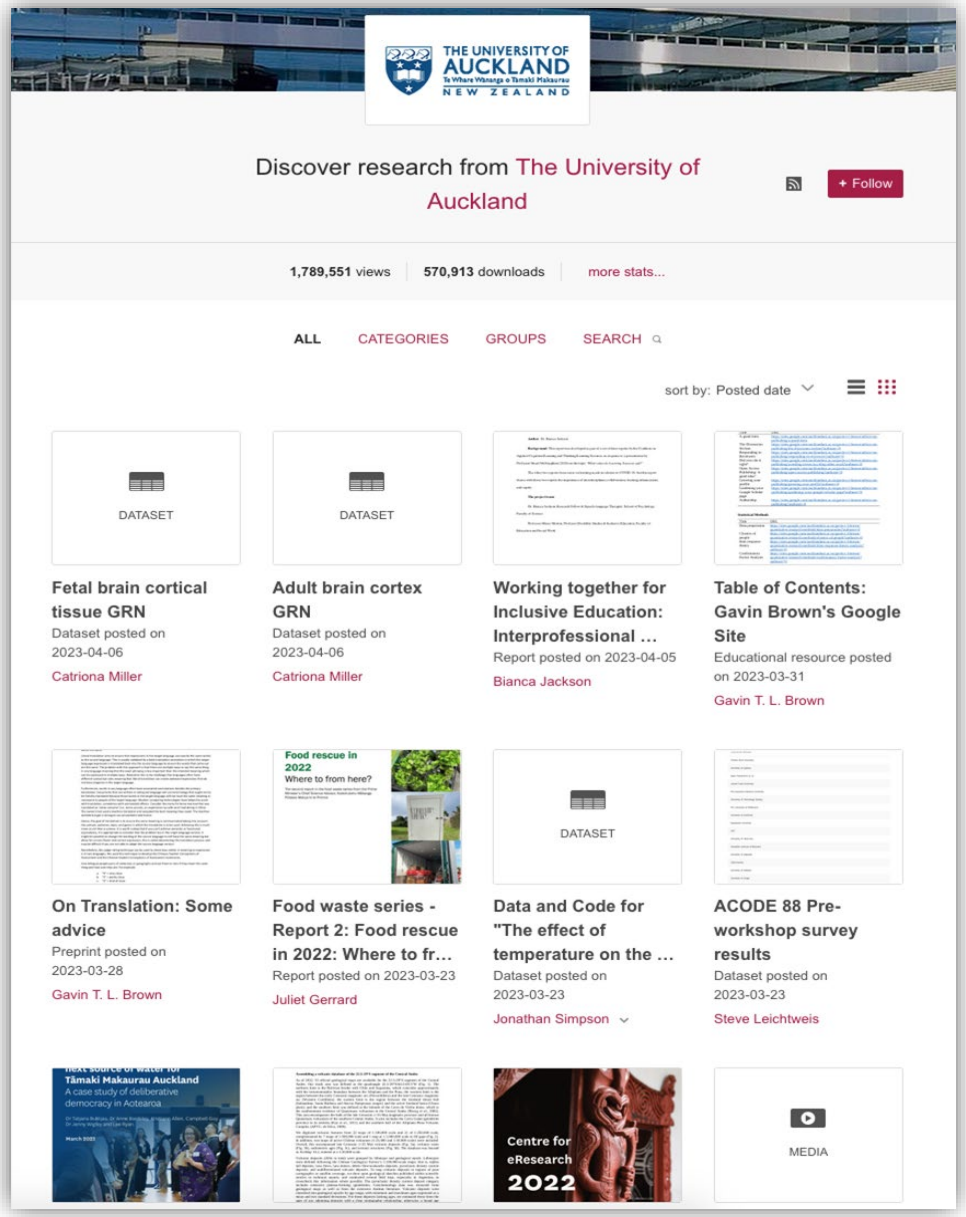
The effects of environmental stimuli, such as temperature, oxygen and nutrient availability, on microbial growth and behaviour are well known and are carefully managed in commercial applications [1–3]. In contrast, sound as an environmental stimulus has received less research attention and receives scant attention in commercial applications. Published results indicate that audible sound (20 Hz– 20 kHz) [4] stimulation can directly affect growth and other

Data repositories

- Designed to store, preserve, and provide access to research data (enables FAIR) or a metadata record (balance of FAIR/CARE)
- Multidisciplinary or discipline specific options

figshare zenodo

- Items are assigned a permanent, resolvable and citable Digital Object Identifier (DOI)
- Track views, downloads, citations for impact.



Data Repository Platforms

Discovery & Reuse (by others)

Data that are available for discovery and access may be reused, either to substantiate findings or to generate new insights.

- Governance
  - Long term stewardship
  - Data access committees
- Impact
- Future collaboration

Re-use of data

Data that are made available for discovery and access may be reused by other researchers, either to substantiate or reproduce original findings or to generate new insights.

Where appropriate, infrastructures have been developed to review proposed uses of data prior to data being released. One example is **Data Access Groups** / **Data Access Committees.**

Governance helps to:

- Ensure compliance with original ethics restrictions
- Prevent damage to the original researcher’s IP
- Prevent harm to study participants (e.g., reidentification)

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.

10. What haven’t we discussed?
(Add your questions into the chat or raise your hand)

UNIVERSITY OF AUCKLAND
Waipapa Taumata Rau
NEW ZEALAND

Questions? Get in touch…

researchdata@auckland.ac.nz

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

Thank you