

AIRS

Advanced Information Research Skills

AIRS Module 8

Managing data



airs.library.qut.edu.au

© Queensland University of Technology (QUT) 2014-. Except where otherwise noted, content in this presentation is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike licence.

Attribution: Advanced Information Research Skills (AIRS), Library, Queensland University of Technology (QUT).

the university for the real world[®]



Contents

Module 8: Managing data	2
8.1 Getting started with research	3
Research integrity.....	4
Research misconduct.....	5
What is research data?.....	6
Research data lifecycle.....	8
8.2 Plan	9
Review secondary datasets.....	9
Your responsibilities.....	10
Documentation and guidance.....	11
Stage 2/Research Proposal and your Data Management Plan.....	12
8.3 Collect or capture data	13
Recommended storage.....	13
Metadata - describe your data.....	15
8.4 Process and analyse	16
Easy metadata for working data.....	17
Documentation.....	18
8.5 Publish and share	18
Why publish data?.....	18
Durable data formats.....	18
Licensing data.....	19
Data repositories.....	20
8.6 Preserve	22
Data retention.....	22
Archiving research data.....	23
8.7 References	24

Module 8: Managing data

[Workshop presentation \(YouTube, 41m\)](#)

'Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly' ([Nature, 2009, para. 1](#)).

Learning objectives

This module explores the strategies and issues that need to be considered for appropriate management of data throughout your research candidature. By the end of this module, you will be able to:

- consider research integrity and your responsibilities as a researcher at QUT under the provisions of the QUT Code for Responsible Conduct of Research
- apply data management strategies to organise and store data proficiently, ethically and legally
- commence your data management and primary materials plan.

Applying the content to your resource log

As you progress through this module, consider, and refer to the requirements of the resource log and apply your learning to each question.

Resource Log

Managing data (15 marks)

- a. Describe your data. Include the format and scope of your research data [\[See Module 8.1\]](#).
- b. Outline your plans to securely store and back-up your data.
- c. Identify and discuss **three** types of contextual information (metadata) that help to make data F.A.I.R. [\[See Module 8.3 and Module 8.5\]](#).
E.g. Types of metadata include descriptive, rights etc.
- d. Discuss the commercial, legal, and ethical considerations related to storing and sharing research data.
E.g. documenting conditions such as: ownership, copyright, re-use licence conditions. Also, privacy and access control considerations.
- e. Identify how long your data is legally required to be retained after completion of your research project/thesis. Refer to the [Data retention obligations](#) and **provide your reasoning** as to how this applies to your research data. [\[See Module 8.6\]](#)

8.1 Getting started with research

In this section, you will learn about research integrity and your responsibilities as a researcher at QUT under the provisions of the QUT Code for Responsible Conduct of Research. The [Australian Code for the Responsible Conduct of Research, 2018 \(also known as the Code\)](#) provides a range of guidelines for responsible research practices. Failure to comply may result in research misconduct.

The Code lists eight principles of responsible research conduct (honesty, rigour, transparency, fairness, respect, recognition, accountability and promotion):

P1 Honesty in the development, undertaking and reporting of research

- Present information truthfully and accurately in proposing, conducting and reporting research.

P2 Rigour in the development, undertaking and reporting of research

- Underpin research by attention to detail and robust methodology, avoiding or acknowledging biases.

P3 Transparency in declaring interests and reporting research methodology, data and findings

- Share and communicate research methodology, data and findings openly, responsibly and accurately.
- Disclose and manage conflicts of interest.

P4 Fairness in the treatment of others

- Treat fellow researchers and others involved in the research fairly and with respect.
- Appropriately reference and cite the work of others.
- Give credit, including authorship where appropriate, to those who have contributed to the research.

P5 Respect for research participants, the wider community, animals and the environment

- Treat human participants and communities that are affected by the research with care and respect, giving appropriate consideration to the needs of minority groups or vulnerable people.
- Ensure that respect underpins all decisions and actions related to the care and use of animals in research.
- Minimise adverse effects of the research on the environment.

P6 Recognition of the right of Aboriginal and Torres Strait Islander peoples to be engaged in research that affects or is of particular significance to them

- Recognise, value and respect the diversity, heritage, knowledge, cultural property and connection to land of Aboriginal and Torres Strait Islander peoples.
- Engage with Aboriginal and Torres Strait Islander peoples prior to research being undertaken, so that they freely make decisions about their involvement.
- Report to Aboriginal and Torres Strait Islander peoples the outcomes of research in which they have engaged.

P7 Accountability for the development, undertaking and reporting of research

- Comply with relevant legislation, policies and guidelines.
- Ensure good stewardship of public resources used to conduct research.
- Consider the consequences and outcomes of research prior to its communication.

P8 Promotion of responsible research practices

- Promote and foster a research culture and environment that supports the responsible conduct of research.

Research integrity

Integrity in research is critical to assuring research and scientific excellence and public trust. Research integrity is exemplified by 'a commitment to intellectual honesty and personal responsibility for one's actions and to a range of practices that characterize responsible research conduct' ([National Research Council \(US\) and Institute of Medicine \(US\) Committee on Assessing Integrity in Research Environments, 2002](#)).

For the individual researcher, integrity embodies a range of good research practices and conduct, including:

- intellectual honesty in proposing, performing, and reporting research
- accuracy in representing contributions to research proposals and reports
- fairness in peer review
- accurate and fair acknowledgement of the work of others (referencing and citing)
- collegiality in scientific interactions, including communications and sharing of resources
- transparency in conflicts of interest or potential conflicts of interest
- protection of human subjects in the conduct of research
- humane care of animals in the conduct of research
- adherence to the mutual responsibilities between investigators and their research participants.

All staff and students who are involved in research practice should be familiar with and must comply with:

- the [QUT Code for Responsible Conduct of Research](#)
- [QUT Responsible Research Framework](#) which is based on:
 - the [Australian Code for the Responsible Conduct of Research \(2018\)](#)

Compliance with the principles outlined in the Code is a requirement for the receipt of funding from the Australian Research Council and the National Health and Medical Research Council.

QUT's [Office of Research Ethics and Integrity](#) has created Research Integrity Online (RIO), an online course in research integrity, that assists researchers to fulfill their obligations set out in the Code. A completion certificate is granted upon completion of the course.

Who must complete the course?

- All HDR students (PhD, Master of Philosophy, Prof Doc) who have enrolled since 1 January 2017 (within the first three (3) months of candidature)
- All staff involved in research, research management, or research support including, for example, academic researchers, lab assistants, research assistants, supervisors of HDRs, and research managers.

In addition to RIO, Graduate Research Education + Development (GRE+D) also facilitates a moderated online module called [eGrad Research Integrity in Practice](#). This course is voluntary and will assist researchers to apply research integrity principles and values to practical, scenario-based research activities.

Research misconduct

The University requires all QUT researchers to maintain the highest standards of research practice in accordance with the [QUT Code for Responsible Conduct of Research \(D/2.6\)](#). The QUT Code is consistent with the *Australian Code for the Responsible Conduct of Research, 2018*, legislation, policies, the accepted practices within a discipline and the codes of relevant external funding bodies.

A deviation from these obligations may be considered either a **breach** or **misconduct**.

Breach refers to a deviation from the QUT Code for Responsible Conduct of Research where the extent, seriousness, wilfulness and/or consequences of the deviation are not significant. The repetition or continuation of a breach may, however, constitute research misconduct, particularly if the QUT researcher has been counselled about the standards of research conduct required by QUT.

Research misconduct refers to deviations from the QUT Code for Responsible Conduct of Research that are intentional and deliberate, reckless, or amount to gross and persistent negligence, and result in serious consequences. This might include:

- fabrication, falsification, plagiarism or deception in proposing, carrying out or reporting the results of research
- failure to declare or manage serious conflicts of interest
- conducting research without ethics approval as required by legislation
- wilful concealment or facilitation of research misconduct by others.

The QUT MOPP [D/2.7 Managing and investigating potential breaches of the QUT Code for Responsible Conduct of Research](#) outlines QUT's procedures for dealing with allegations of research misconduct. In addition, specific procedures for research misconduct apply to student researchers, such as [E/8.1 Management of student misconduct](#).

Researchers, journal editors and scientific institutions worldwide also work together to improve communication about misconduct cases. The [Committee on Publication Ethics \(COPE\)](#) supports and offers training for editors and publishers about ethics relating to all aspects of publishing with particular focus on issues relating to misconduct around research and publication.

Plagiarism or scientific misconduct is formally identified and recorded in various ways including:

- Blogs such as Retraction Watch
- Referred to COPE
- PubMed and other databases.

Activity – View a retraction

View this [article retraction](#).

The retraction includes explanations, case studies and further readings.

What is research data?

Research data may refer to data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis is built. It relates to data generated, collected, or used, during research projects, and in some cases may include the research output itself. Data may be numerical, descriptive, visual or tactile. It may be raw, cleaned or processed, and held in any format or media.

Research data, in many disciplines, may by necessity include the software, algorithm, model and/or parameters used to arrive at the research outcome, in addition to the raw data that the software, algorithm or model is applied to [MOPP D/2.8 Management of research data and primary materials](#). The table adapted from University of Edinburgh’s (2022) [MANTRA, Research data in context \(Classification of research data\)](#) shows research data which has been classified according to processes used to gather or generate them.

Data Class	Process	Content examples	Data examples
Experimental	Generated by lab equipment	Gene sequences; chromatograms	Laboratory notes; specimens; samples; methodology; slides; artefacts
Computational / Simulation	Generated from computational models - the actual model (and its metadata) may be more important than the output data	Climate models; economics prediction models	Methodology; data files; models; algorithms; scripts; workflows; standard operating procedures and protocols; simulation software

Observational	Recordings of specific phenomena at a specific time or location	Seismic data, medical imaging, opinion polls, climate data, interview or survey results	Transcripts; audio or video recordings; field notebooks; diaries; photographs; films; slides; questionnaires; test responses; codebooks; text documents
Derived	Produced via processing or combining of other data	Data mining; compiled databases; GIS	Database contents; spreadsheet data; data files
Reference	Extracted from reference datasets	Genbank, HILDA, ABS CURF datasets	Spreadsheets; data files; contents of an application (schemas, input, output; log files for analysis software)

There are many forms of research data that do not fit into traditional definitions. Watch the recording below to hear about different ways in which data can be understood and used in research.



[Non-Traditional research data: Interview with QUT's Dr Karike Ashworth \[4:04 mins\]](#)

Research data lifecycle

Good data management is the basis of successful research. It is important to plan how you will manage your data from the beginning of your project and throughout the research data lifecycle.



Research data lifecycle. Adapted from [UK Data Service \(2019\) \[1:40 mins\]](#)

The data management lifecycle signposts the different stages your data goes through from the beginning of the research project right through to publishing and reuse. It is important to note that data often has a longer lifespan than the research project that creates them. Good data management practices ensure:

- compliance with the [Australian Code for the Responsible Conduct of Research, 2018](#)
- legal requirements and relevant policies
- facilitate data reuse (for yourself and others)
- insure against catastrophic loss of your raw data.

This module is relevant to data created in a digital form ('born digital'), data converted to a digital form (digitised) or primary material.

8.2 Plan

As a researcher at an Australian university, you are required to collect, manage and retain your research data and primary materials in accordance with relevant legislation and University policies which are consistent with the [Australian Code for the Responsible Conduct of Research, 2018](#). In the planning stage, consider the following:

- conduct a review of secondary datasets and determine whether your research will use secondary data, new data or a combination of both
- consult relevant policies, procedures and legislation
- investigate confidentiality and privacy requirements, and any contractual arrangements with funders or industry partners
- begin a data management plan.

Review secondary datasets

Consider whether secondary datasets are appropriate to answer your research question. With more research data being made openly available for further research and analysis through data repositories and archives, there is a real possibility that useful datasets may be available to either fully or partially answer your research question.

There are two main approaches to using secondary datasets:

1. **Secondary analysis** uses data that has been collected or produced by another researcher to answer a separate and different research question.
2. **Ancillary studies** involve the use of secondary data while adding one or more other measurements to the existing study - adding a questionnaire to an existing study would be an example of an ancillary study.

Case study

QUT Researchers Kavel Deilami, Liton Md. Kamruzzaman and John Hayes used secondary datasets from *NASA's Landsat program* to test causal relationships between land cover/use and surface urban heat islands. Why use Landsat? "...Landsat images are freely available for download, and their extended temporal coverage enabled the researchers to obtain datasets meeting some specific criteria", including five-year time differences. To obtain images with similar air temperatures, data from Archerfield airport weather station was used to identify relevant dates for searching in Landsat. In addition, ABS socio-demographic datasets, IKONOS images and Google Earth images were used in the research.

Deilami, Kaveh, Kamruzzaman, Md., & Hayes, John F. (2016) Correlation or causality between land cover patterns and the urban heat island effect? Evidence from Brisbane, Australia. *Remote Sensing*, 8(9), Article Number-716.
<https://eprints.qut.edu.au/221640/>

Sources of secondary data and published datasets:

- [QUT subscribed datasets](#) including Carbon Disclosure Project datasets, Bloomberg, and

Bureau of Meteorology climate datasets

- [Datasets for research in social science](#)
- [Datasets for Health and STEMM](#)
- [QUT Research Data Finder](#) - datasets published by QUT HDR students and researchers.

Your responsibilities

QUT recognises research data as a valuable product of research activity, which can assist in promoting open enquiry and debate, complementing research outputs and publications, providing research transparency, and justifying research outcomes. You are responsible for appropriately managing research data and primary materials throughout the research lifecycle, by addressing issues of confidentiality, security, ownership, storage, recordkeeping, retention and re-use.

The [QUT MOPP \(D/2.8\) Management of research data and primary materials](#) applies to the management of all research data created by researchers associated or affiliated with QUT including postgraduate students engaged in research activities. The policy states that 'researchers should ensure their data are as FAIR (findable, accessible, interoperable and reusable) as possible, subject to funding body, contractual, commercialisation, ethical, privacy or confidentiality considerations' (QUT, 2020, para. 3).

All QUT researchers and HDR students must complete a [data management plan](#) for all those research projects for which they are responsible. The [Research Data Management and Primary Materials Checklist](#) will help in this task. Documentation of your data management plan can be done with the [QUT Data Management Planning Tool](#) (QUT VPN is required to access the DMP Tool off campus). Remember to consider your responsibilities in relation to confidentiality, privacy, contracts and ownership of data.

Confidentiality

A dataset may contain confidential or classified information that should not be released to the public. If this is the case, you have a responsibility to guard against unauthorised access by using physical locks or digital encryption/password controls. Authorised access should be managed via the use of a signed confidentiality/non-disclosure agreement. Non-disclosure agreements are also used to prevent potentially patentable information from being leaked into the public domain. Inadvertent disclosure of data can prevent a patent from being granted as the product or process is no longer considered 'novel'. If your research involves confidential information, this should be recorded in the data management plan.

For more information, see [ARDC Research Data Rights Management Guide](#).

Privacy

If your data includes person identifying information (names and addresses), health information about identifiable individuals, or information about the movements of identifiable individuals, privacy restrictions will apply. Any personal information should be kept secure. Before collecting

and using personal information in your research, you should seek consent from the person to whom it relates or use a process of de-identifying the data, so it no longer reveals their identity. If your research involves data restricted by privacy, this should be recorded in the data management plan.

See the ARDC [Research Data Rights Management Guide](#) and the ARDC resources on [identifiable data](#) and [sensitive data](#) for more information.

Contracts

When a funding body or industry partner externally funds research, there may be conditions attached to the funding and/or obligations related to ownership, access and dissemination. This information will be recorded in a contract that will be signed by the Chief Investigator, but the obligations apply to all researchers involved in the project. The information should be included in the data management plan.

Ownership of data

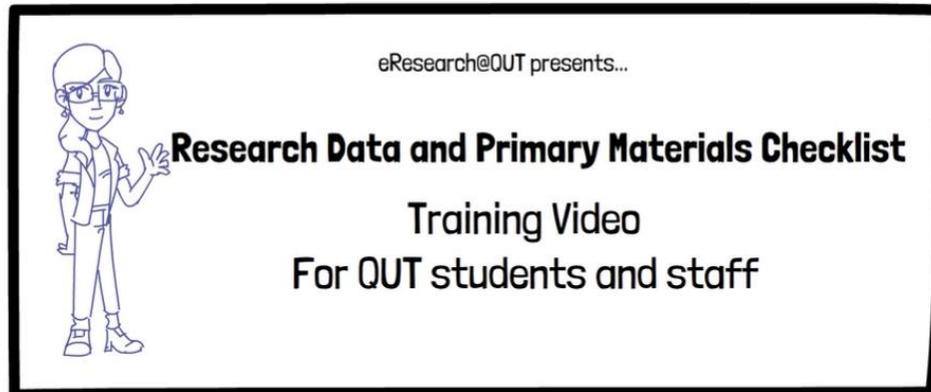
In general, QUT is the owner of research data and primary materials generated in the conduct of research at the University. Research students can retain a copy of the research data and primary materials for future use, subject to any contractual, statutory, ethical or funding body requirements ([MoPP D/2.8.5 Management of research data and primary materials – Ownership and custodianship](#)).

Documentation and guidance

The QUT [Research Data Management and Primary Materials Checklist](#) guides you to ensure that your data and primary materials are managed correctly during the course of your research.

This simple online Checklist provides guidance to you that fulfils the principles set out in the Australian Code for the Responsible Conduct of Research, 2018 and in other international consensus policies, such as the [F.A.I.R. Data Principles](#). The Checklist acts as a precursor to documenting the details of your research data management requirements in the QUT Data Management Planning Tool (see below), where more detail and a plan can be generated. The Checklist also allows your HDR Supervisors to review your progress in developing your data management plans, leading to informed, meaningful conversations.

Watch the video below to understand how to use the Checklist.



[Research data and primary materials checklist: Training video for QUT students and staff \[4:21 mins\]](#)

The [Data Management Planning \(DMP\) Tool](#) will help you to identify and document the legal, confidentiality, privacy, ownership or contractual issues that may apply to your research data (QUT VPN is required to access the DMP Tool off campus). The DMP tool is based on the model created by the UK Digital Curation Centre and includes QUT specific advice to help you write data management plans.

Documentation to capture at the planning stage includes:

- Confidentiality agreements
- Participant consent forms
- Ethics approvals
- Copies of contracts.

Stage 2/Research Proposal and your Data Management Plan

As a researcher you need to be confident about the data you are collecting, ensuring that you store it in a way that complies with the requirements of your project, confidentiality considerations and potential access requirements. As part of your Stage 2/Research Proposal, you are required to indicate whether you have completed a data management plan.

8.3 Collect or capture data

In this stage of the research data lifecycle, researchers must ensure that all research data is stored securely and backed up or copied regularly.

Consider the following when planning to store your research data:

What data formats will be used? The choice of format will determine how the data may be used, analysed, backed up, stored and reused in the future. When deciding which format to use consider:

- Could the hardware, software and media fail or become obsolete within your project timeframe?
- Would the impact of such a failure be disastrous?
- How long does the data need to be stored?
- Is support for the hardware, software or media available at QUT?
- Is security an issue?
- Who needs access to the data? Are team members local at QUT or collaborators from other institutions?

Activity – Data formats

In your [Data Management Plan](#) list the data formats you will use, include the software and any access requirements (QUT VPN is required to access the DMP Tool off campus).

Recommended storage

The information provided here is introductory and brief. You can view an extended list of storage options under [Store digital research data](#).

Please note: CloudStor is no longer an option for storing research data.

Network storage

When storing data, please follow the **3-2-1 rule**:

Keep 3 copies of your files in 2 different locations, with 1 copy off-site (ideally in a different geographic zone).

- Master copy: Keep at secure location
- Working copy: Keep on a reliable/safe device or locations
- Back up copy: Keep off-site

You should store master copies of your data in [QUT's Research Data Storage Service \(RDSS\)](#), which provides all QUT researchers (staff and Higher Degree Research students) with secure storage for data at the different stages of the research lifecycle, including raw, working and archived data. To access to RDSS submit an [IT request online](#). The Research Data Storage Service provides:

- large storage capacity for research data
- easy on-campus and off-campus access to your data
- controlled access to your data.

Decisions about storage for highly confidential or highly sensitive research data should be made on a case-by-case basis.

Network drives may be accessible to a large number of people or can be configured for use by a single user or group of users (contact HiQ for more information). **QUT's H and U drives are not recommended for the storage of master copies of research data.**

The Office of eResearch provides QUT staff and Higher Degree Research students with specialised advanced computing facilities, storage and support. [Submit a request](#) for access and advice to the Help Centre, or [apply for an account](#).

Cloud storage

Please note: CloudStor is no longer an option for storing research data.

QUT OneDrive

Separate from a personal OneDrive account, your [QUT OneDrive](#) account may be used for research data storage.

Benefits:

- Up to 1TB of storage is available to all researchers at QUT.
- Remote access to research data.
- Data can be shared with QUT and external researchers.
- Deleted files in QUT OneDrive can be restored using the self-service previous versions feature.
- Access permissions easily managed.

Limitations:

- Not suitable for master copies of research data.
- Not suitable for sensitive data storage.
- Access to QUT OneDrive storage will cease when you leave QUT.
- External researchers need a Microsoft account to access OneDrive.

QUT SharePoint

Where your work forms part of a larger research project that might have its own Teams/SharePoint site, it might be appropriate for some documents to go there. However, that may be subject to any arrangements in the team or project.

Local hard drives

Local hard drives are helpful when using applications that require files to be stored locally with the application, for example, NVivo. Copies of these working or analysis files should be securely stored in [QUT's Research Data Storage Service as a backup](#).

Removable media (not recommended for storage of research data)

While USB drives, external hard drives and memory cards are useful for transporting and capturing data in the field, data collected should be transferred to stable, secure storage as soon as possible. These devices along with CDs and DVDs are at greater risk of being lost or damaged, are not very robust and can be damaged by magnetic fields, water and high temperatures. If portable media are used for transporting copies of data, use only high-quality products and ensure that any confidential data is encrypted or password protected.

Non-digital research data

Data in non-digital formats (e.g. biological samples, analogue recordings, paper artefacts) should be stored in secure facilities located in your school, faculty, institute or an off-campus research facility. Refer to MOPP [D/2.8.7 Management of research data and primary materials](#) for more information about dealing with non-digital research records. Visit [Store non-digital research data and primary materials](#) for your faculty's Space, Assets and Logistics contact.

Metadata - describe your data

Documenting data and capturing or describing metadata at all stages of the research lifecycle, enables you and others to find, access, interpret, validate and reuse the data. Documentation should provide contextual information for the data so that it can be understood in the future.

Metadata (data about data) can describe individual items or groups of items. Requirements may vary depending on the discipline and type of research being conducted. At the **collection stage**, documentation and metadata to be captured include:

Study level documentation:

- context of data collection: project history, aims, objectives and hypotheses
- collection/generation/capture methods, sampling methods
- instruments used and calibrations required
- hardware or software used
- data scale and resolution, temporal, time or geographic coverage

- questionnaire copies, interview questions and instructions, test samples
- quality control processes

Data level documentation including:

- description of file structures and relationships between files
- file types and file naming conventions chosen
- names, labels and descriptions of variables, records and values
- explanation of codes and classification schemes used
- provenance information about sources of derived or reference data

Adopted from UK Data Service, (2022).

It will be useful to make a copy of the raw data at collection/generation and use this in the working/analysis stage of your project. Retain a read-only copy of the raw data as both back-up against loss or corruption and evidence of the work you have undertaken.

Types of metadata

- **Descriptive:** metadata required for discovery and assessment of the collection, including title, contributors, subject or keywords, study description, list of publications the dataset contributes to and location and dates of the study.
- **Provenance:** metadata about the data source, instruments used to collect or generate the data, version tracking and transformations (often including the steps that were applied to produce the data product).
- **Technical and Structural:** metadata about file types, software, file size and contents of components e.g. variable names. How the data or its database is configured and how it relates to other data or how components within a set related to each other.
- **Rights and Access:** metadata to enable access and licensing or usage rules e.g. negotiated access by contacting the owner or open access via a creative commons licence.
- **Citation:** metadata required for someone to cite the data, including a persistent identifier such as a DOI or stable URL, e.g. Creator(s), Publication Year, Title, Publisher, DOI.

8.4 Process and analyse

Data processing and analysing is an important stage in the research data lifecycle. **Data Analysis** is about how raw data is chosen, evaluated, and interpreted into meaningful and significant conclusions that other researchers and the public can understand and use ([Responsible Conduct in Data Management \(PDF, 530KB\)](#)). **Data processing** is the manipulation and conversion of data by computer in order to format or transform raw data into machine-readable data ([Britannica Academic](#)).

It is recommended that working data (the copy of your master data that you are using for analysis) is stored for processing and analysis in other locations such as QUT OneDrive but

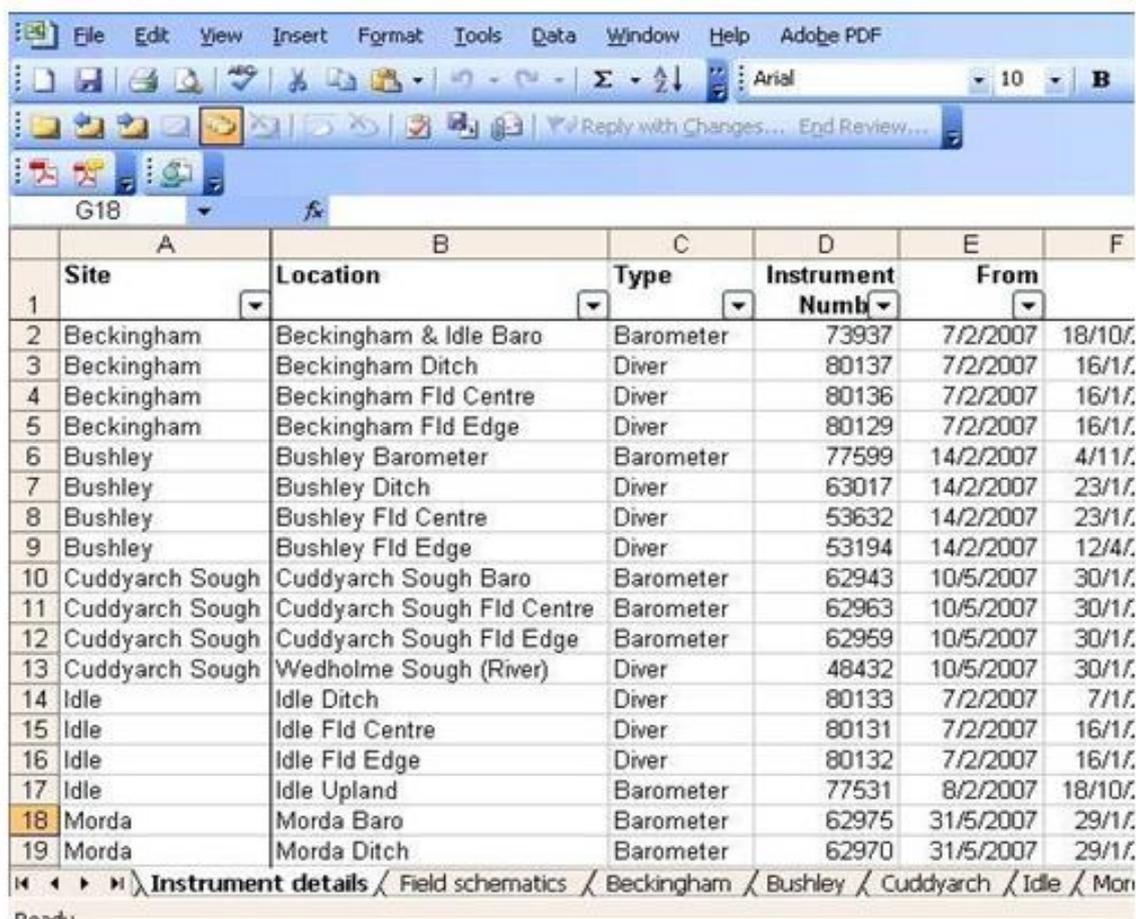
ensure that your master copy is stored in a secure location such as [QUT's Research Data Storage Service](#).

Easy metadata for working data

A simple way to capture metadata during the working phase of your research is to create a readme.txt type file, a collection of simple metadata that describes the details of the datasets and improves the long-term usability of the data. Save the readme.txt file with the data files in the same folder within your research storage. Also store a copy of your data management plan, ethics approval and other relevant documents here too.

Download a Readme.txt template from [Cornell University's guide](#).

In some cases, metadata can be generated or extracted from digital files automatically. For example, a digital camera records the date, time, exposure setting, and file format. Software programs sometimes allow structured metadata such as title, author, organisation, subjects or keywords to be added via 'Properties'. View [disciplinary metadata standards](#).



	A	B	C	D	E	F
1	Site	Location	Type	Instrument Num	From	
2	Beckingham	Beckingham & Idle Baro	Barometer	73937	7/2/2007	18/10/07
3	Beckingham	Beckingham Ditch	Diver	80137	7/2/2007	16/1/07
4	Beckingham	Beckingham Fld Centre	Diver	80136	7/2/2007	16/1/07
5	Beckingham	Beckingham Fld Edge	Diver	80129	7/2/2007	16/1/07
6	Bushley	Bushley Barometer	Barometer	77599	14/2/2007	4/11/07
7	Bushley	Bushley Ditch	Diver	63017	14/2/2007	23/1/07
8	Bushley	Bushley Fld Centre	Diver	53632	14/2/2007	23/1/07
9	Bushley	Bushley Fld Edge	Diver	53194	14/2/2007	12/4/07
10	Cuddyarch Sough	Cuddyarch Sough Baro	Barometer	62943	10/5/2007	30/1/07
11	Cuddyarch Sough	Cuddyarch Sough Fld Centre	Barometer	62963	10/5/2007	30/1/07
12	Cuddyarch Sough	Cuddyarch Sough Fld Edge	Barometer	62959	10/5/2007	30/1/07
13	Cuddyarch Sough	Wedholme Sough (River)	Diver	48432	10/5/2007	30/1/07
14	Idle	Idle Ditch	Diver	80133	7/2/2007	7/1/07
15	Idle	Idle Fld Centre	Diver	80131	7/2/2007	16/1/07
16	Idle	Idle Fld Edge	Diver	80132	7/2/2007	16/1/07
17	Idle	Idle Upland	Barometer	77531	8/2/2007	18/10/07
18	Morda	Morda Baro	Barometer	62975	31/5/2007	29/1/07
19	Morda	Morda Ditch	Barometer	62970	31/5/2007	29/1/07

Documentation

Methods of processing must be rigorously documented to ensure the utility and integrity of the data.

Documentation at the processing and analysing stage should include:

- derived data created, with code, algorithm or command file used to create them
- weighting and grossing variables created and how they should be used
- data list describing cases, individuals or items studied, for logging qualitative interviews
- all structured, tabular data should have cases or records and variables well documented
- other documentation may be contained in user guides, reports, publications, working papers and laboratory books.

8.5 Publish and share

Why publish data?

Publishing your data and citing its location in published works allows others to replicate, validate and ensure accuracy of results. Some research data would be impossible to collect again e.g. recordings of a specific seismic event. Sharing data improves the scientific record and increases scientific integrity. The *Australian Code for the Responsible Conduct of Research, 2018* advises that researchers should share their data wherever possible. To support best practice in publishing data, QUT has adopted the [F.A.I.R data principles](#) to make research data Findable, Accessible, Interoperable and Re-usable. The benefits of making data publicly available include:

- greater attention to the research and the researchers, who created the data
- increased citation counts by up to 69% [see study by Piwowar, Day and Fridsma \(2007\)](#)
- increased collaboration and networking opportunities between researchers
- greater opportunities for grant funding as grant funding bodies encourage data sharing.

Durable data formats

Use standard interchangeable formats that most software is capable of interpreting, or at the end of the project convert completed data to these formats. Check data after conversion for errors or changes. Acceptable formats for long term access to data include the following options.

Data type	Acceptable formats for sharing, reuse & preservation
Containers	TAR, GZIP, ZIP
Computer aided design	DWF, DXF, DWG, DWS, DWT, X3D, STEP, STP

Databases	XML, CSV
Geospatial	SHP, DBF, GeoTIFF, NetCDF
Moving images	MOV, MPEG, AVI, MXF
Sounds	WAV, AIFF, MP3, MXF, FLAC
Statistics	ASCII, DTA, POR, SAS, SAV, STATA, SPSS
Still images	TIFF, JPEG 2000, PDF, PNG, GIF, BMP, RAW
Tabular data	CSV, XLS XLSX, ODS,
Text	XML, PDF/A, HTML, ASCII, UTF-8, RTF, HTML, NUD*IST, NVivo, ATLAS.ti
Web archive	WARC

Based on information from [UK Data Service \(2017\)](#).

Copyright and research data

Compilations of data are protected by copyright law as 'literary works', provided the compilation involved intellectual effort, was not copied from another source, and supplies 'intelligible information' (i.e. is human/machine readable). Research students own the copyright of their works unless there is a written agreement stating otherwise.

See the Australian Research Data Commons' [Research Data Rights Management Guide](#) for more information about whether copyright subsists in your data.



Image attribution: [Mike Seyfang \(2005\)](#).

Licensing data

If you decide to share your data when you have completed your research, you may use a [Creative Commons \(CC\)](#) licence to specify the conditions that apply to reuse. When you apply a CC licence to your data, you retain ownership of the data but license others to use the work on liberal terms. The four different licence terms are:



Attribution: (BY) You must always provide credit to the original author.



Share-Alike: (SA) If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.



Non-Commercial: (NC) You may not use the material for commercial purposes.



No-Derivatives: (ND) You may not distribute modified versions of the work.

Image attribution: [Creative Commons](#)

Case study

QUT Post-doctoral researcher Christopher Nouné has assigned a CC-BY-ND licence to his dataset, which means that others may use, copy and redistribute his work, even for commercial purposes, so long as they give appropriate credit. If they remix, transform or build upon the work, they may not distribute the modified material without seeking permission from Christopher.

See: Nouné, Christopher; Hauxwell, Caroline. (2017): HaSNPV-AC53 Genotyping and Abundance Datasets. [Queensland University of Technology].



<https://doi.org/10.4225/09/595f00367d2cf>

Check which terms apply to the works you want to use via [Creative Commons Australia's licence chooser](#).

Combining and analysing existing data from multiple sources is common practice. The conditions attached to CC licences only apply when an entire, or substantial part of, a CC licensed dataset is reused. You should not apply a CC licence to a dataset if you are not the copyright owner or if the data contains secret, private or confidential information.

Use the [Data Management Planning Tool](#) to document the decisions relating to sharing and licensing your data at the conclusion of your research (QUT VPN is required to access the DMP Tool off campus). For more information on licensing data, refer to the Australian National Data Service (ANDS) guide to [Copyright, data and licensing](#) and QUT's information about [publishing research data](#).

Data repositories

Consider sharing your completed datasets or subset with other researchers, publishers and the public via QUT's data repository [Research Data Finder](#), or an open access repository such as [FigShare](#), [Dryad](#) or [Github](#). [Explore other discipline-specific and multidisciplinary repositories](#) and at the [Registry for Research Data Repositories](#).

In [Research Data Finder](#), you can list the following metadata about your dataset to make it Findable, Accessible, Interoperable and Reusable (F.A.I.R):

- **Descriptive:** metadata required for discovery and assessment of the collection, including title, contributors, subject or keywords, study description, list of publications the dataset contributes to and location and dates of the study.
- **Provenance:** metadata about the data source, instruments used to collect or generate the data, version tracking and transformations (often including the steps that were applied to produce the data product).
- **Technical & Structural:** metadata about file types, software, file size and contents of components e.g. variable names. How the data or its database is configured and how it relates to other data or how components within a set related to each other.
- **Rights & Access:** metadata to enable access and licensing or usage rules e.g. negotiated access by contacting the owner or open access via a creative commons licence.
- **Citation:** metadata required for someone to cite the data, including a persistent identifier such as a DOI or stable URL e.g. Creator(s), Publication Year, Title, Publisher, DOI.

Case study

See how Phd candidate MD. Lifat Rahi has applied the F.A.I.R data principles to his published [research data set](#).

F.A.I.R data principles - applied

- ❖ **Findable**
 - ❖ F1 (meta)data assigned a globally unique and eternally persistent identifier.
 - ❖ F2 data are described with rich metadata.
 - ❖ F3. (meta)data are registered or indexed in a searchable resource.
 - ❖ F4. metadata specify the data identifier.
- ❖ **Accessible**
 - ❖ A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - ❖ A1.1 the protocol is open, free, and universally implementable.
 - ❖ A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
 - ❖ A2 metadata are accessible, even when the data are no longer available.
- ❖ **Interoperable**
 - ❖ I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
 - ❖ I2. (meta)data use vocabularies that follow FAIR principles.
 - ❖ I3. (meta)data include qualified references to other (meta)data.
- ❖ **Re-usable**
 - ❖ R1. meta (data) have a plurality of accurate and relevant attributes.
 - ❖ R1.1. (meta)data are released with a clear and accessible data usage license.
 - ❖ R1.2. (meta)data are associated with their provenance.
 - ❖ R1.3. (meta)data meet domain-relevant community standards.

www.force11.org/group/airgroup/airprinciples

F3 & A2 R1.3 = Research Data Finder registry which uses interoperable metadata schema.

F2 = title, description, creator.

I3 = link to dataset.

R1.1 = CC Licence and access rights.

**I1 = includes a readme.txt file
R1.3 = FASTA file format is the standard in bioinformatics.**

R1 = equipment & software used to generate data, scope of data, limitations.

I3 = connections to researcher records with ORCIDs, publications with DOIs.

F2 & I2 = keywords, subject headings, FoR codes.

F1, F4, A1 & A2 = DOI and https.

R1.2 = Data dates, connections, project record links, geographical area.

8.6 Preserve

Data retention

You are obligated to retain research data and research records for as long as required by legislation, statutory requirements, funding agency guidelines and contractual arrangements with research partners. Decisions about data retention and disposal should be documented in your data management plan and stored with the data. [The Queensland State Archives University Sector Retention and Disposal Schedule](#) specifies the statutory retention periods for research data (see section 601.2/A50) as set out below. For more information on statutory retention periods see: [About retention and disposal schedules](#).

How long do I need to keep research data?

Data retention periods vary depending on the kind of data involved, its significance to the research discipline and the public and whether it may be of interest to future researchers for re-use. Retention periods are defined in section 2.1 of the [Australian Code for the Responsible Conduct of Research, 2018](#).

The minimum retention periods are:

Research data	Description	Status	Action
Research data - other (does not result in patent)	Research data created in the conduct of research which does not fit into the other categories, which does not result in a patent.	Temporary	Retain for 5 years after last action e.g. publication
Research data - other (results in patent)	Research data created in the conduct of research which does not fit into the other categories below and results in a patent.	Temporary	Retain for 7 years after expiry of patent
Research data - clinical trials	Research data created in the conduct of clinical trials	Temporary	Retain for 15 years after completion of clinical research/trial AND 10 years after last patient service provision or medico-legal action

Research data - significant	Research data created in the conduct of a research project, including clinical trials, which is of high public interest or significance.	Permanent	Retain permanently
Short term research projects that are for assessment purposes only, such as research projects completed by students	Research project completed by undergraduate students as part of course assessment	Temporary	12 months after the completion of the project

How long do I need to keep signed consent forms?

All records relating to participant consent in research activities has to be kept for 15 years after the project is concluded or abandoned, according to the Queensland government's [University Sector Retention and Disposal Schedule](#).

Archiving research data

[QUT's Research Data Storage Service](#) has designated archival storage for the retention of research data from completed projects. Contact [Office of eResearch](#) for access. Retain a copy of all metadata and documentation with the completed data sets in the archived project folder.

8.7 References

- Alberts, B. (2011). Retraction. *Science*, 334(6063), 1636.
<https://www.science.org/doi/10.1126/science.334.6063.1636-a>
- Data's shameful neglect. (2009). *Nature*, 461(7261), 145.
<https://www.nature.com/articles/461145a>
- Deilami, Kaveh, Kamruzzaman, Md., & Hayes, John F. (2016) Correlation or causality between land cover patterns and the urban heat island effect? Evidence from Brisbane, Australia. *Remote Sensing*, 8(9), Article Number-716. <https://eprints.qut.edu.au/98528/>
- Encyclopaedia Britannica. (n.d.). Data processing. Britannica Academic. <https://academic-eb-com.eu1.proxy.openathens.net/levels/collegiate/article/data-processing/472172>
- Force11. (2014). *FAIR data principles*. <https://www.force11.org/group/fairgroup/fairprinciples>
- National Research Council (US) and Institute of Medicine (US) Committee on Assessing Integrity in Research Environments. (2002). *Integrity in scientific research: Creating an environment that promotes responsible conduct*. Washington (DC): National Academies Press (US), 2002.
<https://www.ncbi.nlm.nih.gov/books/NBK208706/>
- Piowar, H. A., Day, R. S., & Fridsma, D. B. (2007) Sharing detailed research data is associated with increased citation rate. *PloS One*, 2(3), e208.
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>
- Queensland University of Technology. (2020). *D/2.8 Management of research data and primary materials*. Brisbane: QUT. https://www.mopp.qut.edu.au/D/D_02_08.jsp
- UK Data Service. (2022.) *The importance of managing and sharing data*. <https://ukdataservice.ac.uk/learning-hub/research-data-management/>
- UK Data Service (2019). Research data lifecycle. Youtube. <https://www.youtube.com/watch?v=-wjFMMQD3UA>
- University of Edinburgh. (2022). *Research data MANTRA*. <https://mantra.ed.ac.uk/>