

# AIRS

## Advanced Information Research Skills

### AIRS Module 8

### Managing data



<http://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).

a university for the **real** world<sup>®</sup>



# Contents

|  |           |
|--|-----------|
| <b>Module 8: Managing data</b> .....                         | <b>2</b>  |
| <b>8.1 Getting started with research</b> .....               | <b>3</b>  |
| Research integrity.....                                      | 4         |
| Research misconduct.....                                     | 5         |
| Research data lifecycle.....                                 | 7         |
| <b>8.2 Plan</b> .....  | <b>9</b>  |
| Review existing datasets.....                                | 9         |
| Your responsibilities.....                                   | 10        |
| Documentation and guidance.....                              | 11        |
| Stage 2/Research Proposal and your Data Management Plan..... | 11        |
| <b>8.3 Collect or capture data</b> .....                     | <b>12</b> |
| Recommended storage.....                                     | 12        |
| Metadata - describe your data.....                           | 15        |
| <b>8.4 Process and analyse</b> .....                         | <b>16</b> |
| Easy metadata for working data.....                          | 16        |
| Documentation.....   | 16        |
| <b>8.5 Publish and share</b> .....                           | <b>17</b> |
| Why share data?.....   | 17        |
| Durable data formats.....                                    | 18        |
| Licensing data.....  | 19        |
| Data repositories.....                                       | 20        |
| <b>8.6 Preserve</b> .....                                    | <b>21</b> |
| Data retention.....  | 21        |
| Archiving research data.....                                 | 22        |

# Module 8: Managing data

[Workshop presentation \(YouTube, 40m15s\)](#)

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

## Learning objectives

---

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

- consider research integrity and your responsibilities as a researcher at QUT under the provisions of the QUT Code of Conduct for Research
- apply data management strategies to organise data proficiently, ethically and legally
- commence your data management 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 administrative, commercial, legal and/or ethical considerations related to storing and sharing research data.  
*E.g. accurate description of data, file format, data structures, storage location, etc.*  
*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. [\[See Module 8.6\]](#)  
*Consult the relevant section in the 'Guidelines for the Management of Research Data at QUT'.*

## 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 of Conduct for Research. The [Australian Code for the Responsible Conduct of Research](#) 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' ([Committee on Assessing Integrity in Research Environments, National Research Council, Institute of Medicine, 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.

The QUT Blueprint 5 articulates a cultural value of research and innovation, aimed at "Promoting a strong culture of research integrity and ethical research practices via appropriate training for academics, researchers and students". These values are consistent with national and international conventions on research integrity including the Singapore Statement on Research Integrity.

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

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

The *Australian Code for the Responsible Conduct of Research* provides a framework for responsible research conduct set out in a number of high-level principles and responsibilities, which apply to researchers and research institutions. The Code outlines expectations for the responsible, ethical and integral conduct of research in Australia. A failure to meet the principles and responsibilities is a breach in the Code.

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.

## Research misconduct

The University requires all QUT researchers to maintain the highest standards of research practice in accordance with the QUT Code of Conduct for Research ([D/2.6](#)). The QUT Code is consistent with the *Australian Code for the Responsible Conduct of Research*, 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 of Conduct for 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 Research Code 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.

QUT policy [D/2.7 Procedures for dealing with allegations of research misconduct](#) 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 \(COPE\)](#) "provides advice to editors and publishers on all aspects of publication ethics and, in particular, how to handle cases of research and publication misconduct".

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.

The Integrity Online for Students (RSC\_RIO) must be completed by All QUT HDR (PhD, Master of Philosophy & Prof Doc) students within the first three (3) months of their candidature. Failure to complete this course may affect your candidature status.

### Activity – Code of Conduct quiz

The [Research Integrity Online \(RIO\) quiz](#) must be completed by all QUT HDR must be completed by all QUT HDR students who have commenced on or after January 1, 2017

## 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 Model 2017.](#)

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
- 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 physical material.

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](#)). The table adapted from [MANTRA](#) shows research data which has been classified according to processes used to gather or generate them.

| Data Class                        | Process  | Content examples  | Data examples   |
|-----------------------------------|--|---|---|
| <b>Experimental</b>               | Generated by lab equipment   | Gene sequences; chromatograms   | Laboratory notes; specimens; samples; methodology; slides; artefacts  |
| <b>Computational / Simulation</b> | 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                       |
| <b>Observational</b>              | 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 |
| <b>Derived</b>                    | Produced via processing or combining of other data   | Data mining; compiled databases; GIS  | Database contents; spreadsheet data; data files   |
| <b>Reference</b>                  | Extracted from reference datasets  | Genbank, HILDA, ABS CURF datasets   | Spreadsheets; data files; contents of an application (schemas, input, output; log   |

files for analysis software)

## 8.2 Plan

Plan is the first stage of the data lifecycle. 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](#). In the planning stage, consider the following:

- conduct a review of existing datasets and determine whether your research will use existing 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 existing datasets

Consider whether existing 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 existing 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 Kaveh Deilami, Liton Md. Kamruzzaman and John Hayes used existing 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/98528/>

Sources of secondary data and published data sets:

- [QUT subscribed datasets](#) including HILDA, Bloomberg, ABS CURFs and Bureau of Meteorology
- [Datasets for research in social science](#)
- [Datasets for STEM](#)
- [QUT Research Data Finder](#) - datasets published by QUT 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 [MoPP D/2.8 Management of Research Data](#) 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 'research data will be made available for access and re-use subject to any contractual, ethical, privacy or confidentiality matters'. It emphasises that 'Research which is supported by public funding, and higher degree research student projects, must use the online research data management planning tool'. 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 confidentially/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 [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 [Research Data Rights Management Guide](#) and the [ANDS Guide to publishing and sharing 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 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](#)).

## Documentation and guidance

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. 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' (QUT Research Student Centre). As part of your Stage 2/Research Proposal, you are required to indicate one of the actions listed below:

- I have not yet considered data management issues but will do so before confirmation
- I have read the data management planning information and discussed the need to complete the Data Management Plan (online) with my supervisor
- I have read the data management planning information, discussed it with my supervisor, and have completed the Data Management Plan (online)

### 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.

Watch the video: [Data Management \(YouTube video, 2m57s\)](#)

### Recommended storage

#### 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](#) 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 services for 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.

You can view an extended list of storage options under [Collect and store research data](#).

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). **The 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. [Contact the team](#) for access and advice, or [apply for an account](#).

## Cloud storage

---

### QCIF's QRISdata

[QRISdata](#) is a data storage service hosted by the [Queensland Cyber Infrastructure Foundation \(QCIF\)](#), designed to complement well-managed data storage provided by your institution. Depending on the storage option you select, QRISdata may be replicated on tape storage, and/or stored in multiple locations. For more information about QRISdata, read the [QRIScloud FAQs](#).

Benefits:

- High volume storage for research data, both active and long-term use.
- Support for working and archived data is provided.
- Researchers are provided with a range of tools to help manage data, enable collaboration

and enable fine-grained control access.

- Large datasets can be shared with researchers/collaborators worldwide.
- Access to a wide range of existing eResearch services, tools and applications is possible.
- Integrated access to Queensland-based high-performance computing facilities and specialised cloud services is available.

Limitations:

- Data storage allocation is based on merit, taking into account the data's significance and value to the wider research community. Applications processing could take up to one month.
- Long-term storage plans are required.

### **AARNet's CloudStor**

CloudStor is a file transfer and storage service for easily and securely storing, sending and receiving large files, available to all Australian researchers via the Australian Access Federation.

Benefits:

- Up to 100GB of storage is available to individual researchers. 'Low-cost' additional data storage is available on request.
- Up to 100 recipients can share/receive a file.
- Australian Access Federation credentials used to login, providing access to researchers at other universities and research institutions.

Limitations:

- Maximum file size for transferring data is 100GB.
- Not suitable for master copies of research data.
- Not suitable for sensitive data, as encryption only occurs during transmission.

### **Local hard drives**

---

Local hard drives are helpful when collecting and storing data in the field. However, master copies of data need to be securely stored in [QUT's Research Data Storage Service](#).

### **Removable media (not recommended for storage of research data)**

---

While USB 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) should be stored in secure facilities located in the school, faculty, institute or an off-campus research facility. Refer to the [QUT Governance Services](#) web page or [D/2.8.7 Management of research data and primary materials](#) of the MoPP for more information about dealing with non-digital research records.

## 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 Archive, (2017). Create and manage data, <https://www.ukdataservice.ac.uk/manage-data/document/data-level>)

It will be useful to make a copy of the raw data at the collection stage and use this in the working stage of your project. Retain a read-only copy of the raw data in the [RDSS Acquisition space](#) 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 cycle. **Data Analysis** is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data ([Responsible Conduct in Data Management](#)). **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 (copy of your master copy) is stored for processing and analysis in other locations such as QCIF 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.

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'. See more [tools for understanding digital files](#).

### 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

- 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.

### Embedding metadata in an MS Excel spreadsheet

|    | A               | B                          | C         | D                 | E         | F        |
|----|-----------------|----------------------------|-----------|-------------------|-----------|----------|
|    | Site            | Location                   | Type      | Instrument Number | From      |          |
| 1  |                 |                            |           |                   |           |          |
| 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  |

## 8.5 Publish and share

### Why share 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 scientific record and increases scientific integrity. The Australian Code for the Responsible Conduct of Research advises that researchers should share their data wherever possible. To support best practice in sharing data

QUT has adopted the [F.A.I.R data principles](#) to make research data Findable, Accessible, Interoperable and Re-usable. The benefits of sharing data publicly include:

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

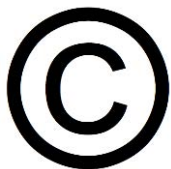
### Durable data formats

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 Stanford University Libraries (2017) and UK Data Archive (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.

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].

<http://dx.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 [DMP Tool](#) to document the decisions relating to sharing and licensing your data at the conclusion of your research. For more information on licensing data, refer to the Australian National Data Service (ANDS) guide to [Copyright, data and licensing](#).

## Data repositories

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

In [QUT 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/fairprinciples](http://www.force11.org/group/airgroup/fairprinciples)

**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 ORCID, publications with DOIs.**

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

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

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

## 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.

| Description (nature of research)  | Retention Period  |
|---|---|
| Research data which does not result in a patent   | Retain for 5 years after last action.   |
| Research data which results in a patent   | Retain for 7 years after expiry of patent.  |
| Clinical trials   | Retain for 15 years after completion of clinical research/trial AND 10 years after last patient service provision or medico-legal action. |
| <p><b>Significant</b> - research data which is of high public interest or significance to the discipline such that it has or will change a commonly held view or approach.</p> <p>Factors which may determine significance include projects which:</p> <ul style="list-style-type: none"> <li>• are controversial</li> <li>• are subject of extensive debate</li> <li>• arouse widespread scientific or other interest</li> <li>• have the potential to cause major adverse impacts on the environment, society or human health</li> <li>• involve eminent researchers</li> <li>• involve the use of major new or innovative techniques.</li> </ul> | Permanent   |

### Archiving research data

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