

AIRS

Advanced Information Research Skills

AIRS Module 8

Manage



Contents



Module 8: Manage	2
8.1 Storage	3
Common storage devices	3
Networked storage.....	3
Data repositories.....	4
Non digital research data	5
Stage 2 proposal and your data management plan	6
8.2 Legal issues	6
QUT MOPP D/2.8: Management of research data	6
Licensing data.....	8
8.3 Documentation and metadata	9
Kinds of documentation	9
Metadata requirements and responsibilities	9
Metadata schemas.....	11
Metadata standards and controlled vocabularies	11
8.4 Research Integrity	12
What is research integrity?.....	12
Ethical misconduct.....	14
References	15



Module 8: Manage

MANAGING RESEARCH DATA

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

This module explores the strategies and issues that need to be considered in appropriately managing data through your research candidature. The information that follows will guide you in managing your data effectively across the research lifecycle. By the end of this module you will:

- apply data management strategies to organise data proficiently, ethically and legally
- get started with creating your data management plan.

Good data management is the basis of successful research and it is important to plan how you will manage your data at the beginning of your project. 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), and are insurance against catastrophic loss of your raw data.

Research data is information another researcher might need to validate or replicate your research. Data might be facts, observations, images, computer program results, recordings, measurements or experiences upon which an argument, theory, test or hypotheses might be founded. Data can be numerical, descriptive, visual or tactile; it may be raw, cleaned or processed and may be held in any format or media. Research data is not journal articles, conference papers, books or other forms of published results unless this is the topic of your research.

Disciplines may have their own discipline-specific language to describe and interact with research data. This module is relevant both to data created in a digital form ('born digital') or data converted to a digital form (digitised), such as:

- documents (text, Word), spreadsheets
- laboratory notebooks, field notebooks, diaries
- questionnaires, transcripts, codebooks
- audiotapes, videotapes, photographs, films
- test responses
- slides, artefacts, specimens, samples
- collection of digital objects acquired and generated during the process of research
- data files
- database contents (video, audio, text, images)
- models, algorithms, scripts
- contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- methodologies and workflows
- standard operating procedures and protocols.

8.1 Storage

Research data should be stored securely in durable formats and backed up regularly from the commencement of a research project. Storage arrangements need to comply with the QUT statutory retention period (Section 7 of the [Guidelines for the Management of Research Data](#)). Things to consider when selecting where to store digital research data include:

- How long does the data need to be stored?
- Is security an issue?
- Could the hardware, software and media fail or become obsolete within the timeframe?
- Would the impact of such a failure be disastrous?
- Is support for the hardware, software or media available?

You may choose to store data on removable media, hard drives, and networked drives and store non-digital data. Refer to [Digital Data Storage Options for QUT Researchers](#) for additional information.

Common storage devices

Removable media

USB drives, memory cards, CDs and DVDs are convenient and affordable. Their smaller, portable size means they are at greater risk of being lost or damaged; removable media are not very robust; the data can be damaged by magnetic fields, water and high temperatures. For large quantities of data, you may need to use multiple disks or USB drives making it difficult to retrieve specific files (especially if the documentation is poor). 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. USB drives are not considered suitable for working copies of data.

Hard drives

Storing master copies of research data files on the main drive of individual desktop or laptop personal computers is a convenient option for working copies of data. However, protect data by backing it up regularly to networked storage.

Networked storage

[QUT's Research Data Storage Service](#) provides all QUT researchers (staff and Higher Degree Research students) with a secure data repository. To discuss or request Research Data Storage needs, submit an [IT request online](#) and an analyst will call you to further discuss your needs. You may require Research Data Storage because it:

- helps achieve compliance with research grant conditions
- maximises overall value from entire research lifecycle
- unlimited storage capacity for research data

- easy on-campus and off-campus access to your data
- controlled access to your data.

A number of options exist to help QUT researchers store research data:

- **Networked drives:** provides different data storage drive locations:
 - The H drive provides 10GB of personal storage to QUT Higher Degree Research students and staff for personal use. Access to the H drive is restricted to your account only. This is not considered a good place for your research data.
 - The U drive is restricted to your peers and colleagues within your organisational unit and is handy if you wish to share data within your faculty. There are different areas with U:, some are suitable for faculty work, others more suitable to research.

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 the IT Helpdesk for more information). Decisions about storage for highly confidential or highly sensitive research data should be made on a case-by-case basis in consultation with faculty IT support staff. The storage method and location may also need to be approved by a research ethics committee.

High Performance Computing & Research Support (HPC) provides QUT staff and Higher Degree Research students with specialised advanced computing facilities, storage and support, including [HPC File Store](#), [Request an HPC account](#) or [contact the HPC team](#).

- **Cloud Storage:**
 - [QRIScloud](#): is a merit-based cloud computing and data storage service hosted by the Queensland Cyber Infrastructure Foundation (QCIF).
 - [CloudStor](#) is an AARNet web service that allows you to store and share files in the cloud. It gives individual researchers at AARNet-connected institutions (e.g. QUT) up to 100GB of storage free of charge.
- [QUT Media Warehouse](#) is a rich media repository that houses re-usable multimedia. Types of media to house in Media Warehouse include:
 - **Images** may be original creative digital works or digital photographs of physical creative works, or digital photographs of places, people and events.
 - **Videos** may be original creative works, such as movies or animations, or video recordings of places, people, events or performances.
 - **Sound** may be original creative digital sound works or recordings of performed works or sound recordings of places, people or events.

You can view an extended list of storage options under [Data storage and sharing](#). Arrange a consultation with [HPC Advisory Service](#) to discuss your storage needs.

Data repositories

At the conclusion of your research you should consider the possibility of storing your data in a data repository (if it can be made publicly available), so it can be shared.

Use repositories for datasets accompanying manuscripts, where you can make the data underlying

scholarly publications discoverable, accessible, understandable, freely reusable, and citable for all. Multidisciplinary repositories include [QUT Research Data Finder](#), [FigShare](#) and [Dryad](#). Find other repositories at the [Registry for Research Data Repositories](#).

Why share data?

Publishing your data and citing its location in published works can allow others to replicate, validate and ensure accuracy of results. 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. Sharing data has many benefits to researchers including:

- promotes the research and the researchers, who created the data
- publically available data is associated with an increase in citation count by 69% (see [Sharing detailed research data is associated with increased citation rate](#))
- increased collaboration and networking opportunities between researchers
- greater opportunities for grant funding as grant funding bodies encourage data sharing.

Australian National Data Service (ANDS) provides further reasons to share data in [Sharing Data Ethically](#).

Consider making the metadata about your data available to the public (e.g. via QUT's data registry, QUT Research Data Finder). The metadata should include a description of the data, conditions of access, your contact details, and the location of the data.

In [QUT Research Data Finder](#), you can list the following information about your dataset:

- your data collection (e.g. a description of the data, conditions of access, and the location of the data)
- the activity (e.g. details about the research activity/project that produced the data)
- the service (e.g. details of software or specialised instruments used to collect your research data)
- parties (e.g. information about you – research interests, contact details, collaborating partners).

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, and institute or off-campus research facility. Refer to the [Records Management](#) section of the [QUT Governance Services](#) web page for more information about dealing with non-digital research records.

Activity – data protection

Go to the [Data Management Planning Tool](#) and complete Sections 4 and 6 relating to data protection, sharing and reuse.

Stage 2 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 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)

For more information on your Stage 2 requirements refer to the [Research information on the QUT Students site](#).

8.2 Legal issues

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. Module 2 introduced you to the [Australian Code for the Responsible Conduct of Research](#); this module focuses on the legal requirements of good data management including the [QUT policy MOPP D/2.8 Management of research data](#) and licensing your work via [Creative Commons](#).

QUT MOPP D/2.8: Management of research data

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. This policy covers data management planning, recordkeeping, storage, retention, disposal, privacy, confidentiality, access and reuse of data. The policy states that ‘Research data will be made available for access and re-use subject to any contractual, ethical, privacy or confidentiality matters’. [MOPP D/2.8 Management of research data](#) emphasises that ‘Research which is supported by public funding, and higher degree research student projects, must use the online research data management planning tool’.

The [Guidelines for the Management of Research Data at QUT](#), together with the [Data Management Planning \(DMP\) Tool](#), will help you to identify any legal issues that may apply to your

research data. The DMP is a tool has been developed by QUT Library based on the model created by the UK Digital Curation Centre to help you write data management plans.

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).

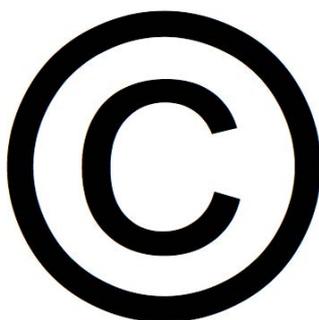


Image attribution: [Mike Seyfang \(2005\)](#)

Ownership of data

In general QUT owns original research data and primary materials. Intellectual property (IP) is different and you should understand this difference. Data could be considered the 'thing', IP is the special knowledge in obtaining it.

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 retention periods for research data. The [Guidelines for the Management of Research Data at QUT](#) gives an overview of the retention schedule for research data and research records.

Confidentiality

A dataset may contain confidential 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. See [Practical Data Management: A Legal and Policy Guide](#) (p.22) for more information.

Privacy

If your data includes personally 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 confidential information, this should be recorded in the data management plan. See [Practical Data Management: A Legal and Policy Guide](#) (p.24) for more information.

Contracts

For research that is externally funded by a funding body or industry partner, 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.

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. You would retain ownership of the data but license others to use the work on liberal terms.

The most liberal CC licence requires attribution (BY) which means that the author is attributed, that the work is not falsely attributed to another author and that the work is not altered so as to prejudice the author's reputation.

An additional five CC licences can be created by combining attribution with other conditions, such as no commercial use (NC), no derivatives (ND) which means adaptations/mashups are not allowed, and share alike (SA) which means that derivative works must be shared under the same terms.

Example: The creative commons licence 'CC-BY-NC' means that others may use the work so long as the author is attributed and it will not be used for commercial purposes.

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. Decisions relating to sharing and licensing your data at the conclusion of your research should be included in the data management plan. For more information on licensing data, refer to the Australian National Data Service (ANDS) guide to [Copyright, Data and licensing](#).

Alternatively, you may wish to specify more restrictive conditions on the reuse of your data. The AUSGoal [Restrictive Licence Template](#) has been developed specifically for material that contains personal or other confidential information. It may also be used for other reasons, including material that is to be licensed under some form of limiting or restrictive condition (such as a time limit on use, or payment arrangements other than an initial once-only fee).

8.3 Documentation and metadata

In this section you will learn about documentation and metadata, including requirements of each, responsibilities, standards and the value of controlled vocabularies. Good documentation ensures that your data will be easier to understand and its quality easier to judge.

Readily accessible, detailed and clear data is trusted and usable. Reproduction and validation of the results is faster with good documentation. If that documentation is not available, the data runs the risk of being collected but not used.

Kinds of documentation

The documentation should provide contextual information for the data so that it can be understood in the future. Requirements will vary depending on the discipline and type of research being conducted. Producing good documentation is easier if it is planned from the start of a project and considered throughout the lifecycle of the data.

Documentation should include:

- project aims and objectives (to provide context)
- catalogue of data collected
- description of lifecycle of key data elements (procedures for collection/creation, validation, transformation, processing, analysis, publication, archiving/destruction)
- description of instruments, calibrations etc.
- description of how data is structured (data model, coding schemes, controlled vocabularies etc.)
- details of any quality control processes
- confidentiality agreements and consent forms
- laboratory notebooks and experimental protocols
- questionnaires, codebooks, data dictionaries
- software syntax and output files
- methodology reports
- provenance information about sources of derived data.

Metadata requirements and responsibilities

Metadata (data about data) is standardised information about a resource, presented in a structured format that is both machine-readable and human-readable. Metadata can describe individual items or groups of items (individual files, images or datasets etc). The items described by the metadata

may be physical or digital.

Example: A library catalogue includes metadata about books and eBooks held by the Library plus the electronic journals to which the Library subscribes. A library catalogue record is simply a collection of metadata elements (for example, title or author of a book), linking to items in the library collection through the call number.

Metadata helps the Library to manage its resources and assists users in the discovery and use of those resources.

Metadata helps researchers to manage and reuse data after its creation.

Ideally, as much metadata as possible should be gathered at the beginning of a research project, with ways devised to collect metadata (automatically if possible) throughout the life of the project.

Types of metadata

There are different types of metadata:

- *Descriptive metadata* provides 'descriptive terms' that will facilitate search and retrieval of files.
- *Rights metadata* is information about ownership of the data.
- *Administrative metadata* includes preservation requirements, confidentiality requirements, access restrictions and timelines (e.g. release dates).
- *Provenance metadata* provides information about the data source, version tracking and transformations (often including the steps that were applied to produce the data product).
- *Technical metadata* gives information about file types, software, file size and contents of components (e.g. variable names, contributing performers of each track in audio recordings etc).
- *Structural metadata* indicates how components of a set relate to one another (e.g. a detailed list of all the tables in a database or the details of how one object relates to another (e.g. is an earlier version of...)).

Identifying documentation and metadata requirements

To choose the best metadata for your research first consider your own needs (or the needs of the research project team). If the plan involves depositing the data in a data repository or archive, you should consider the data documentation and metadata requirements of the relevant repository.

There are three key questions to answer:

1. *Responsibilities:* Who will be responsible for what?
2. *How will the metadata be stored?* Some metadata can be stored within the digital object (e.g. the metadata in a digital image file) but often this is not the case. For external metadata, consider using a data repository. Repository software stores 'digital objects' which are made up of metadata plus one or more related files.
3. *How will the metadata be created?* In some cases, metadata can be generated or extracted

from digital files automatically. For example, a digital camera records the date, time, exposure setting, file format etc. In other cases, human effort will be required to create documentation and metadata. Software programs sometimes allow structured metadata (e.g. include title, author, organisation, subjects and keywords) to be added via 'Properties'.

Metadata schemas

A 'metadata schema' defines a standard *set of terms* (e.g. a controlled vocabulary) that will be used to describe a resource and a *set of rules* that define the syntax and language (e.g. XML). Wherever possible, metadata should be created using schema which are in common use, as this will facilitate the process of contributing a metadata record to a data repository at the completion of the project.

Metadata standards and controlled vocabularies

Metadata standards and controlled vocabularies provide a means for standardising descriptions within your metadata schema so that you are always able to retrieve a set of items that have been allocated the same descriptor.

Additionally, standardisation enables computers to retrieve and merge metadata from multiple sources. Some standards are suitable for many different kinds of material and across disciplines; others are more discipline-specific. Examples include:



[Dublin Core](#) is used to index websites. You can check if a particular website uses Dublin Core by whether their source code includes 'dc' fields.



[Registry Interchange Format-Collections and Services](#) (RIF-CS) was developed as a data interchange format for supporting the electronic exchange of data collection, parties, activities and service descriptions. It organises information about data collections and services into the format required by the [Australian National Data Service \(ANDS\)](#) Collections Registry, [Research Data Australia \(RDA\)](#). RDA is designed to be a metadata registry for research data collections specifically collected in Australia and/or relevant to Australian research interests.



[Text Encoding Initiative \(TEI\)](#) is used to maintain a standard for representations of text chiefly in the humanities, social sciences and linguistics.



The [Visual Resources Association Core \(VRA\)](#) is used to describe original works of visual art and also images of them.



The Content Standard for [Digital Geospatial Metadata](#) (CSDGM), ISO 19115:2003. This is used for geographic digital data such as Geographic Information System files (GIS).



[The ANZLIC Metadata Profile](#) is a metadata standard for students working with Australian and New Zealand geographic data (AS/NZS ISO 19115:2005) and meets the ISO 19115:2003 standard.



[Data Documentation Initiative](#) (DDI) is commonly used for social and behavioural science data.

[Core Scientific Metadata Model](#) (CSMD) is used chiefly in the 'structural sciences', i.e. chemistry, material science, earth science and biochemistry where researchers are concerned with analysing the structure of substances and perform systematic experimental analyses on samples of those materials. The CSMD is being used as the core metadata within data management infrastructure being developed for large scale scientific facilities (e.g. the [ISIS Neutron Source](#) and the [Diamond Light Source](#)).

There are many discipline-specific metadata standards available. Investigation of the commonly used standards in your discipline should be part of the data management planning process.

Activity – metadata and your project

Complete section 6 of the Data Management Plan on metadata.

8.4 Research Integrity

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. QUT researchers are also bound by the Australian Code for the Responsible Conduct of Research, which provides a range of guidelines for responsible research practices. Failure to comply may result in research misconduct.

What is 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 [Singapore Statement on Research Integrity](#) is a global guide to the responsible conduct of research.



As a researcher at QUT, you are required to conduct your research with honesty, fairness, respect and professionalism. At QUT, the key policy on research integrity is the [QUT Code of Conduct for Research](#) (D/2.6 QUT Manual of Policies and Procedures). This policy outlines the key elements for the responsible conduct of research, including advice on data management, ethics, authorship, peer review and collaborative research. It also outlines the consequences of academic misconduct.

QUT researchers are also bound by the [Australian Code for the Responsible Conduct of Research](#). This Code has relevance across all research disciplines, guiding institutions and researchers in responsible research practices. It explains how researchers and institutions should:

- manage breaches of the Code and allegations of research misconduct (including the responsibilities and rights of researchers if they witness research misconduct)
- maintain research data and materials
- publish and disseminate research findings, including proper attribution of authorship
- conduct effective peer review
- manage conflicts of interest.

Ethical misconduct

At QUT, research misconduct is constituted by a failure to comply with the specific provisions of the [QUT Code of Conduct for Research](#) and the broader *Australian Code for the Responsible Conduct of Research*. QUT policy [D/2.7 Procedures for dealing with allegations of research misconduct](#) outlines QUT's procedures for dealing with allegations of research misconduct.

Researchers, journal editors and scientific institutions worldwide also work together to improve communication about misconduct cases, such as published retractions of journal articles where misconduct in the form of plagiarism or scientific misconduct has been formally identified. These published retractions are logged by:

- PubMed and other databases
- Blogs such as [Retraction Watch](#)

Activity – view a retraction

View [this journal article retraction](#).

The retraction includes explanations, case studies and further readings. Using this resource will help build confidence in interpretation and application of the [QUT Code of Conduct for Research](#).

It is in your best interests to undertake [workshops offered by the Research Students Centre](#). The workshops will introduce you to your responsibility for managing and assuring your academic conduct and research integrity. If you commenced your HDR after January 1, 2014 you are required to complete the Code of Conduct quiz as soon as possible (at least within three months of commencement of candidature).

Activity – Code of Conduct quiz

Complete the [Code of Conduct for Research Higher Degree Students Quiz](#).

References

Data's Shameful Neglect. (2009.) *Nature*, 461(145). doi:10.1038/461145a

Sharing detailed research data is associated with increased citation rate. (2007.) *PLOS ONE*. doi:10.1371/journal.pone.0000308