

Griffith University

Guidelines for Effective Management of Research Data

Description of the Guideline

This guideline recognises the need to manage the increasing volume and complexity of data created by researchers at Griffith University and sets out a framework for best practice data management.

Research data is valuable for a number of reasons. In the first instance, it has value to researchers for the duration of their research. It may also have residual value to those researchers after results have been published, as well as value for other researchers or the wider community. Because of this, and given the investment Griffith University has made in the research, data needs to be carefully managed to be accessible for use and reuse, in ways that satisfy legal, statutory and funding bodies' requirements.

This guideline outlines how researchers can comply with the requirements of the [Griffith University Code for the Responsible Conduct of Research](#) with regards to management of data (including retention and preservation where appropriate).

Other Griffith Related Policies, Procedures & Forms:

[Griffith University Privacy Plan](#)
[Griffith University Intellectual Property Policy](#)
[Griffith University Intellectual Property Procedures](#)
[IT Code of Practice](#)
[Records Management Policy](#)
[Information Resources Management Guidelines](#)
[Research Computing Services offer Data Management Support](#)
[Data Management Plan Questionnaire](#) (online)

1. Principles

See Appendix 1 for a list of principles that are used as the guiding principles adopted by this guideline. In addition, or more specifically, the following principles are applicable:

1.1. Centralised Data Storage

Where possible, research data is to be stored centrally, in a facility provided by the Information Services Division. Enterprise backup procedures will be implemented that involve off-site copies using contemporary standards.

1.2. Security of Data

Research data will be held in a manner that allows for controlled access where this is required.

1.3. Sharing of Research Data

Research data should normally be made available to be shared with the wider community wherever possible, giving consideration to the complete life cycle of that data to determine what is appropriate at what stages in the life cycle.

1.4. Preservation of Data when Required

Data will be preserved for the long term if it is deemed by researchers or the University to be of value.

1.5. Sustainability of Storage

A business model that allows for affordable, increasing storage allocation to researchers is to be provided.

1.6. Information Lifecycle Management

Provides a framework within which to build a system of capacities and procedures to manage data across the whole of their lifecycle from creation to disposal or permanent preservation (see examples in Appendix 2).

1.7. Use of Open Standard Formats

The use of open standard formats to create and store [data](#) is encouraged, to facilitate long term use and availability of research data.

2. Definitions

2.1. Controlled vocabulary

A controlled vocabulary is a set of pre-selected words or phrases used by people in documents or databases or file systems to describe the subject content of the items cited or submitted to the database. Typically experts in specific knowledge domains come up with their own controlled vocabulary

2.2. Copyright

The exclusive right given by law for a certain term of years to an author, composer, designer, etc., to print, publish, and sell copies of their original work.

2.3. Curation

The act of curating, of organizing and maintaining a collection of works or artifacts.

2.3.1. Digital Curation

Digital curation is the curation, preservation, maintenance, and collection and archiving of digital assets

2.4. Data

Information stored in digital form, including text, numbers, images, audio, video, software, algorithms, equations, animations, model, simulations, digital surrogates of physical objects (eg. sculpture, jewelry, 3D scans/models, motion capture data, and historical documents), etc.

2.4.1. Observational Data

'Many scientific records derive from observations of natural phenomena. Their subjects are single events that cannot be repeated, e.g. a volcanic eruption or a cyclone, or they may be long term examinations of natural cycles and patterns, e.g. climatic change, changes in sea level or vegetation communities, or impacts of human activities.'¹

¹ Wendy Southern, 'Appraising science and technology records', Disposal Manual Technical Guideline E8, National Archives of Australia, June 1992, p. E8.5

2.4.2. Experimental Data

'Experimental studies answer questions and test hypotheses by manipulating measurable parameters. Experiments may be set up in laboratories, workshops or in other controlled environments. If all of the original parameters are replicable and the experimental design and set up is documented, experiments and their results can be repeated.'²

2.4.3. Analysed Data

Data, models or records that are based on raw data (usually observational or experimental). Raw data is reduced, manipulated or processed to reach research conclusions.³

2.4.4. Creative and Performing Arts Data

Data, digital surrogates, and records of creative practice, research and teaching activities. This material may be an input of, a record of or an output of the creative process and research and teaching associated with the creative process. Meta-data is both a primary and secondary output of this category of academic activity.⁴

2.4.5. Associated Documentation

Data by itself may be of little value and is itself highly compromised without associated documentation containing relevant information about the data (eg information about the format of the data to assist in the migration to new technologies over time).

2.5. Data Management

All of the actions needed to maintain data over their entire lifecycle and over time for current and future generation of users. This embraces curatorial care, including preservation and archiving.

2.6. Formats

The ways that data are captured digitally, including types of software programs (.pdf, .doc, .jpg).

2.7. Intellectual Property

Property which is the product of invention or creativity, and which does not exist in a tangible, physical form (e.g. patents, trademarks, and copyright material).

2.8. Metadata

Information about data, including author, owner, date of creation, context, source

2 Wendy Southern, 'Appraising science and technology records', Disposal Manual Technical Guideline E8, National Archives of Australia, June 1992, p. E8.5

3 Wendy Southern, 'Appraising science and technology records', Disposal Manual Technical Guideline E8, National Archives of Australia, June 1992, p. E8.5

4 ACLS - American Council of Learned Societies. (2006). Our cultural commonwealth: Report of ACLS Commission on cyber infrastructure for humanities and social sciences. Retrieved November 2, 2006, from <http://www.acls.org/cyberinfrastructure/>

AEHRN - Australian e-Humanities Research Network. (2004). Leveraging digital scholarship in the humanities. Retrieved November 15, 2006, from <http://www.ehum.edu.au/arc-report/context.html>

Heery, R., & Anderson, S. (2005). Digital repositories review. Retrieved January 12, 2006 from <http://ahds.ac.uk/preservation/digital-repositories-review-2005.rtf>

etc.

2.9. Open Standard Format

An open format is a published specification for storing digital data that is free from legal restrictions on use. The use of open formats helps to ensure long-term access to data without current or future uncertainty with regard to legal rights or technical specification.

2.10. Originators

People who create data (authors, inventors etc.), are termed 'originators'.

2.11. Preservation

The act of making data available for use over time, including managing/migrating formats and ensuring security of data from corruption or deletion.

2.12. Research Records

Records are documents containing data or information in any form, created or received by an organisation or person in the course of their work, and kept as evidence of that work or because of the value of the information contained within those documents in the interpretation of the data. Research records include project files, correspondence (including e-mail), laboratory notebooks and consent forms.

2.13. Research Management Records

Include grant management records (applications, reports and acquittals) and ethics committee records (applications, deliberations and approvals).

2.14. Storage

The act of storing digital data, where there is no accompanying activity to also manage usability and accessibility of such data over time.

2.14.1. Local Storage

Predominantly researcher's local disc drives, or data collectors attached to experimental instruments. To be used for short-term storage only.

2.14.2. Centrally Managed Network Storage

A facility operated by the Information Services Division, which can provide reliable, secure long-term storage for research data.

2.15. Archival Significance

Data is considered to have archival significance if the value of benefits for research and the broader community are assessed as outweighing the costs of its continued retention.

3. Data Management Plan

All research proposals should be accompanied by a data management plan that will

outline how data flowing from the research is to be created, managed, accessed and stored over the life of the research and, if applicable, in the long term. This plan should be an integral part of each Research Program and each Research Project. An example data management plan is attached as Appendix 3. An interactive version of this data management plan is available as a [training resource in Blackboard](#).

A data management plan ideally will be created before the research has begun. Researchers moving their data to central storage will be required to complete a data management plan before the data is accepted and transferred.

A typical data management plan would encompass the following elements (which will be explained below):

- i. Originators and owners of the data
- ii. Description of project
- iii. Metadata (schema / standards)
- iv. Types of data to be collected
- v. Volume of data to be managed and storage requirements
- vi. Retention of research data and records and legal controls
- vii. Quality assurance and controls
- viii. Format/s of and software used in creation and use of the data
- ix. Security, access policies and provisions
- x. Roles and responsibilities of researchers and database managers
- xi. Confidentiality requirements
- xii. Storage, preservation and archiving of data
- xiii. Ethics Clearance
- xiv. Supporting Documentation

3.1. Originators and owners of the data

3.1.1. Intellectual Property and Copyright

The originators of research data at Griffith University have certain rights to use and disseminate the data which will be in accordance with the Griffith University Intellectual Property Policy⁵.

3.1.2. Special Arrangements and exceptions

Any exceptions to the Intellectual Property Policy or any special agreements involving intellectual property rights or copyright (e.g. research projects involving multiple institutions) should be documented in the Plan including references to any supporting documentation such legal agreements etc.

3.2. Description of Project

The description of the research project provided by the researcher is an important tool for managing the research data itself and the information about that research data (metadata).

This section links very closely with section iii. of the Plan (Metadata), in that the description of the project will determine much of the metadata required to manage the research data that is created.

3.3. Metadata (schema / standards)

Metadata is a vital tool to enable researchers to identify and re-use data after its creation.

⁵ Griffith University Intellectual Property Policy §2.2

Wherever possible, metadata should be created using an existing schema to assist in interoperability and the ability to share data. Standard schemata may be part of a suite of information management tools provided by the University to assist in data management (via the proposed data curation function – see appendix 4), or the researcher may use another schema if the standard schemata are unsuitable for their project.

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

In general, the types of metadata collected will consist of:

Access metadata	Information governing rights, permissions, restrictions and timelines (e.g. release dates)
Organisational metadata	Administrative and management information
Provenance metadata	Information about the source, version tracking and transformations
Descriptive metadata	Descriptors for search and retrieval (bibliographic/descriptive terms)
Technical metadata	Information about file types, software used, size of file, contents of components of data such as variable names and descriptions or contributing performers of each track in audio recordings etc.

3.4. Types of data to be collected

Data generated by researchers will have a wide diversity in format and origin. Generally the data collected will fall into one of three categories, based on its origin: observational data, experimental data or analysed data.

Decisions on data preservation, format management and version control will be assisted by knowing what type of data is being created by the researcher.

3.5. Volume of data to be managed and storage requirements

If possible, the researcher should provide an estimate of the amount of data their research project may create, to allow the University to plan data storage facilities in an efficient manner.

If the data is to be integrated with other data being collected or currently held, then the researcher should outline this including the physical location of where the data is to be held.

3.6. Retention of research data and records and legal controls

Research data and records need to be retained for as long as required to:

- meet statutory or regulatory obligations (records legislation, funding agency guidelines, contractual arrangements with research partners);
- meet current needs of researchers;
- meet future needs of researchers where these can reasonably be anticipated; and
- satisfy community expectations of the University in documenting research activity.

Decisions about how long research data and records are to be retained should be documented in the data management plan prepared for the research project.

Decisions to retain research data can be reviewed from time to time, especially

at the point when data must be copied, migrated or converted in order to preserve it. As there are costs associated with retaining, managing and preserving research data over long periods of time, the benefits for the research and broader community must be assessed as outweighing the costs of retention.

Research data may go through several versions from its initial creation to a final outcome. Not all research data will be preserved for the long term, as not all data will have lasting research value. Differing versions of data may also be discarded, where the original data has been retained.

Generally, research data and records are retained for the length of time required in the centre, school, department or faculty where they have been produced. If data and records are to be retained permanently, responsibility for storage and preservation will rest with the University centrally. See Appendix 2 for a table of retention periods for research data and records.

3.7. Quality assurance and controls

Identify the measures to be put in place for quality assurance and controls including provisions for the auditing of compliance with data management and access requirements.

3.8. Format/s of software used in creation and use of the data

It is important that the data management plan records information about what file format originators of research data expect to use to capture and store their research data, and what software programs will be used to create, manipulate the data and retrieve the data (e.g. view, playback).

There are many different electronic formats (some specific to particular research disciplines) used to capture, store, manipulate and retrieve data. Where possible, an open standard format should be chosen to capture and store data (for example, .pdf).

There can be no guarantee that data will be useable once a format becomes outdated, but it is more likely to be so if it was originally captured with an open standard format.

It is recommended that any raw data required to be held for the long-term but not held currently in an open standard format are either migrated to an open standard format or held in such a way as to enable their long-term use.

While it is recommended that the software used to create and manipulate research data is a widely used and recognised program, the key requirement is that data be saved in an open standard format wherever possible.

Where this is not possible, the software and data may be stored together but owing to changes in operating systems, no guarantees can be made as to long-term usability.

Advice and support for managing formats and software will be available via a suite of information management tools provided by the University to assist in data management.

3.9. Security, Access Policies and Provisions and Confidentiality Requirements

While most data is initially private to an individual researcher or group, it is possible that either now or later such data may be useful to other researchers or the wider community.

Free and open access to data is encouraged wherever feasible subject to access regimes, confidentiality agreements, and copyright and intellectual property issues.

The University will develop a scalable Research Repository infrastructure for the storage and management of research data. Support for wider community access to enable research collaboration is system requirement. Discovery mechanisms will provide access to this data through federated searching of repositories, metadata harvesting or through a catalogue of the data available. This catalogue may include information about ownership and access controls.

Any access to research data would be governed in the first instance by the researcher or research group, and be subject to privacy and intellectual property considerations as outlined below and in sections 3.i and 3.xii of this guideline.

Griffith University Privacy Plan

<http://www.griffith.edu.au/ua/aa/vc/pp/>

Griffith University Intellectual Property Policy

<http://www62.gu.edu.au/policylibrary.nsf/mainsearch/0a1e4df4c2be64584a256bc100631d7d?opendocument>

Griffith University Intellectual Property Procedures

<http://www62.gu.edu.au/policylibrary.nsf/mainsearch/8a581b3be85b69e34a2570530063fdcf?opendocument>

IT Code of Practice

<http://www.griffith.edu.au/computer/coc/>

Griffith University Code for the Responsible Conduct of Research

<http://www62.gu.edu.au/policylibrary.nsf/xupdate/month/e7852d226231d2b44a25750c0062f457?opendocument>

3.10. Roles and responsibilities of researchers and database managers

Clearly identify who is responsible for what, including who is responsible for the management of the databases, the authorisation and access and the actual data repository. Where research is conducted across institutions, the roles and responsibilities of external parties must also be identified in the data management plans.

A list of the creators, collectors, compilers, “consumers/user” and the parties that funded the data collection would also be useful.

3.11. Confidentiality Requirements

While Griffith’s policies and procedures in relation to confidentiality requirements are covered in Section 3.ix above, the person who is responsible for the data (see Section 3.x.) is responsible for ensuring an external requirements in relation to confidentiality are met.

3.12. Storage, preservation and archiving of data

The originator of research data has primary responsibility for arranging storage and preservation of data (refer to NHMRC and AVCC guidelines). Where the originator of the data is external to the University and that data is stored on University equipment, a copy of the terms of the agreement must be stored with the data and attached to data management plans.

The University will provide a research data storage service, managed by the Information Services Division and funded by a combination of central funds and through cost recovery.

This facility will provide reliable preservation including the attributes of physical

diversity (all data replicated across multiple sites) and non-repudiation (applied where appropriate to ensure that classes of data nominated by a researcher cannot be changed, whether accidentally, deliberately or maliciously).

4. Identified Responsibilities

4.1. Responsibilities of Researchers

Completion of a data management plan at the beginning of a research program or project.

4.2. Responsibilities of the Information Services Division

Provision of storage and retrieval for research data, advice on storage costs and technologies available.

4.3. Responsibilities of the Data Curation Function

Support of researchers in data management, including offering advice on completion of data management plans, forensic curation of data, and provision of advice to researchers and the University on issues to do with research data management.

4.4. Responsibilities of the eResearch Committee

Oversight and approval of guideline.

4.5. Responsibilities of the Office for Research

Ensure that Data Management Plans are completed during applications for grants etc.

4.6. Responsibilities of the University Legal Officer

Ensuring intellectual property guidelines are up-to-date. Provide advice as needed.

4.7. Responsibilities of the University Archivist

Provide advice on statutory requirements or standards for retention, and archival significance of research data and records.

Appendix 1

Information Management Principles for Research

- 1. Corporate Importance**
Information is a valued resource and Griffith's strategic and operational plans define how it should be used and managed.
- 2. Information Sources**
Because we manage information it can be easily found, and its source reliably identified.
- 3. User-centredness**
The information services and systems at Griffith are designed and provided with the needs of users in mind.
- 4. Availability**
Information should have the potential to be available to anybody, anytime, anywhere and anyhow as appropriate.
- 5. Staff and Student Development**
We continuously improve our skills in working with information.
- 6. Productivity and Efficiency**
Because we manage information, we are more productive as a community.
- 7. Statutory Requirements**
The way we manage information complies with legal and administrative requirements.
- 8. Trustworthy Information and Systems**
Griffith information is relevant, accurate and timely.
- 9. Retention and Disposal**
We keep the right information for Griffith and the wider community.
- 10. Information and Technology**
Our information needs are supported by our Information and Communication Technology services and systems.

Appendix 2

Retention and Disposal Arrangements for Research Data and Records

Refer to [the Griffith University Code for the Responsible Conduct of Research -- Schedule of Retention Periods for Research Data and Primary Materials for deciding on retention periods for research data and primary materials.](#)

	Description	Qualifier	Retention Period	Storage / Custody
1	Research data and records from research project or investigations of archival significance *	- Controversial - Innovative experimental technique used for the first time - Research of known public interest	Retain permanently	Griffith central research repository or distributed system.
2	Research data and records from other research projects or investigations	For short term research projects that are for assessment purposes only, such as research projects completed by students	Retain for 12 months after the completion of the project.	Retain in secure, accessible form in research location (faculty, school, department, centre)
		Not involving clinical trials	Retain for 5 years after publication of results; may then be destroyed	Retain in secure, accessible form in research location (faculty, school, department, centre)
3		Medical research <i>involving clinical trials</i>	Retain for 15 years after publication of results; may then be destroyed	Retain in secure, accessible form in research location (faculty, school, department, centre)
4		Involving psychological testing or intervention with <i>adults</i>	Retain for 7 years after publication of results; may then be destroyed	Retain in secure, accessible form in research location (faculty, school, department, centre)
5		Involving psychological testing or intervention with <i>children</i>	Retain for 25 years after date of birth of child; may then be destroyed	Retain in secure, accessible form in research location (faculty, school, department, centre)
6	Research data collections where there is an obligation to lodge data with a national or international repository for the discipline		Transfer to repository. Retain copy for Griffith where significant local benefit in doing so	Retain in secure, accessible form in research location (faculty, school, department, centre)
7	Research data collections of archival significance where the copy has <i>not</i> been lodged with a national or international repository for the discipline	- Generally observational data - Costly or impossible to reproduce - Of interest to researchers in same or other disciplines - If work has community or heritage value	Retain permanently	Griffith central research repository or distributed system.
8	Research data collections, not of archival significance, where the data is to be used in Griffith teaching programs		Retain until no longer required for teaching	Retain in secure, accessible form in research location (faculty, school, department, centre)
9	Other research data collections		Retain for 15 years after publication of results; may then be destroyed	Retain in secure, accessible form in research location (faculty, school, department, centre)

* What should be kept:

- Design
- Data and explanations of data such as codebooks etc
- Results
- Approvals and consent

Retention periods for research data and records are aligned with the following sources:

- *Retention and Disposal Authority for the Records of Higher and Further Education Institutions* (PROS 02/01) issued under section 12 of the Public Records Act 1973.
- *National Statement on Ethical Conduct in Human Research* (2007) developed jointly by National Health and Medical Research Council Australian Research Council Australian Vice-Chancellors' Committee which consists of a series of Guidelines made in accordance with the *National Health and Medical Research Council Act 1992* ('the Act').
- Classes 2, 3, 4 and 7 are based on *Joint NHMRC / AVCC Statement and Guidelines on Research Practice* (1997) and the revision to these guidelines, the *Australian code for the responsible conduct of research*, Reference No R39, 2007
<http://www.nhmrc.gov.au/funding/policy/researchprac.htm> and
<http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>
- Section 6.10 of *General Retention and Disposal Schedule for Queensland Universities*, Queensland State Archives, QSA02/365, Version 1, June 2005.
<http://www.archives.qld.gov.au/government/disposal.asp> and
<http://www.archives.qld.gov.au/downloads/Universities.pdf>

Appendix 3 Sample Data Management Plan

A version of this data management plan is available as a [training resource in Blackboard](#). Research Computing Services also offer a [Data Management Support Service](#).

GRIFFITH UNIVERSITY DATA MANAGEMENT PLAN

- 1. ORIGINATORS AND OWNERS OF THE DATA**
(Who will create the data? How will the data be generated and collected? Who does the data belong to? Does it belong to the creators only, or in combination with government/funding body? If the data is owned by an external party what are the terms of agreement it is being stored under?)
- 2. DESCRIPTION OF PROJECT**
(What are the projected outcomes of the research and how are they to be achieved?)
- 3. METADATA SCHEMA / STANDARDS**
(Which Metadata schema, if any, will be used to manage the research data created by this project?)
- 4. TYPES OF DATA TO BE COLLECTED**
(What types of data that is it envisaged will be collected by the research? For example, experimental data, interviews, video, music etc.)
- 5. VOLUME OF DATA TO BE MANAGED AND STORAGE REQUIREMENTS**
(How much data will be stored over the life of the project, if known? Is the data to be integrated with other data being collected or currently held? Where is the data to be held including the physical location?)
- 6. RETENTION OF RESEARCH DATA AND RECORDS AND LEGAL CONTROLS**
(Are there any statutory requirements dealing with the retention and management of the data created by this research? If so, what bodies are concerned? How long does the data need to be kept to fulfill these requirements? Are there any other legal constraints over the data held such as access and security? Are there any ethical issues involved over the complete lifecycle of the data?)
- 7. QUALITY ASSURANCE AND CONTROLS**
(What measures are to be put in place for quality assurance and control? What provisions are included for the auditing of compliance with data management and access requirements?)
- 8. FORMAT/S OF THE DATA, AND SOFTWARE USED IN CREATION AND USE OF DATA**
(What format/s are expected to be used to collect and retain data? For example, .pdf files, Excel spreadsheets, etc. What software programs will be used in connection with the research, if know?)
- 9. SECURITY, ACCESS POLICIES AND PROVISIONS**
(Who has rights to access the data? What data access restrictions might apply over the life of the data? Who should people seeking access to the data contact? How will the data be disseminated and managed? Is there a timeline for the distribution of the data?)

10. ROLES AND RESPONSIBILITIES OF RESEARCHERS AND DATABASE MANAGERS

(Who is responsible for what? For example, INS Database Management Services (CTI) may manage databases under a SLA, Identity & Messaging Services (EIS) is responsible for authorising who has access including regular audits etc. Provide a list of creators, collectors, compilers, “consumers/users”, parties that funded the data collection. Who is managing the actual data repository?)

11. CONFIDENTIALITY REQUIREMENTS

(What confidentiality requirements apply to the data? For example medical information, commercial confidentiality requirements, confidential interviews, etc.)

12. STORAGE, PRESERVATION AND ARCHIVING OF DATA

(Should this data be preserved for longer than any statutory requirements might dictate. Should it be kept in perpetuity?)

13. ETHICS

(Are there any ethics issues involved with the storage of the data. Is there an ethics clearance number? If so, quote the clearance number?)

14. SUPPORTING DOCUMENTATION

(Are there any documents that need to be considered in relation to the sections above. If so, attach the documents to the plan or indicate where these documents can be located during the time that the data is to be held by the University e.g. Central Registry, stored in the database with the data etc).