QUT Library Collection Development Manual

3. Practices

3.4. Digital Repository

3.4.6 Research Data Collection

Identification Block

<table>
<thead>
<tr>
<th>QUT Library Collection Development Manual</th>
<th>3.4.6 Research Data Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Officer</td>
<td>Research Data Librarian</td>
</tr>
<tr>
<td>Approval Authority</td>
<td>Library Leadership Team</td>
</tr>
<tr>
<td>Approval Date</td>
<td>6 June, 2019</td>
</tr>
<tr>
<td>Date of Next Review</td>
<td>6 June, 2020</td>
</tr>
</tbody>
</table>

Policy Statement

1. Purpose and scope
2. University policy and practice
3. Background
4. Description of the research data collection
5. QUT data repository
6. External Research Data Service
7. Services
8. Responsibilities
9. Access

3.4.6 Research Data Collection

Purpose and scope

The purpose of this guideline is to outline the scope, policy and practice, background, collection, repository, services, responsibilities and access issues for the QUT Research Data Collection.

It acknowledges and guides the Library’s role in the management and administration of QUT’s collection of published research datasets, as produced by QUT staff and Higher Degree Research (HDR) students, as well as the safekeeping of said research data. In addition, it guides the Library’s
role in the development and maintenance of data management planning infrastructure, such as the Data Management Planning Tool, and safekeeping of those plans.

This guideline also acknowledges the Library’s role as information broker, mediating between data providers and QUT researchers.

**University policy and practice**

In November 2015, the University Academic Board approved the revised University Policy MOPP D/2.8 Management of research data, which states the value of research data as an output of the research process; outlines the University’s approach to best data management practices and storage; and mandates the use of QUT’s Data Management Planning Tool in the creation of data management plans by Higher Degree Research (HDR) students as well as researchers undertaking publicly-funded research (see also Guidelines on the use of the Data Management Planning Tool for Higher Degree Research Students).

The policy was previously updated in 2013 to include reference to the Research Data Finder data repository, which provides description, discovery and access services for research datasets, and again in 2015, to include reference to the Data Management Planning Tool.

With assistance from the eResearch Office, the Library has published information on the staff and student community webpages, which complement the Manual of Policies and Procedures, and provide detailed information on the practical issues of data ownership, storage and backup, data formats, documentation and metadata, privacy and confidentiality, retention and disposal, and access and re-use. The Data Management Planning Tool also provides guidance on the above issues in great depth.

Developed in 2016/2017, the Research Data Management Strategy 2017-2020 was revised in October 2018 by the newly established Director, eResearch. The purpose of the Strategy is to transform and direct research data management resources, promote best research data management practices, build capacity and skills, and craft a cultural narrative that recognises research data as a valuable asset. A Project Manager will be appointed in the second half of 2019. This role will guide the implementation of activities outlined in the Strategy.

**Background**

Research has moved into a data-centric era, which is becoming popularly known as the Fourth Paradigm of research. New methods of conducting research are commonly referred to as eResearch or eScience and the systems and networks referred to as cyberinfrastructure or eResearch infrastructure. The methods involve the marrying of discipline-specific knowledge, scientific method and information communication technology services e.g. high-performance computing, such that the skill sets become intertwined into a digital literacies skill set.

The storage, publication and re-use of research datasets suggests increased efficiency and productivity can be achieved, therefore accelerating the rate of discovery. The use of large datasets, and the mash-up of previously unrelated datasets allows not only for the discovery of new patterns in data which were previously not visible, but also enables peer validation of research data and outputs.

There is evidence that making research data openly available and linked to publications, increases interest in both the data and the associated publication(s), which in turn increases the citation count and therefore the potential impact of the research. Several international collaborations began in
2018/19 to provide a framework for organisations to expose and discover links between articles and datasets, thereby providing an indication of data citations e.g. Scholarly Link Exchange (Scholix).

A number of government-funded initiatives have developed over the last decade to make available data storage and sharing facilities, as well as virtual research environments. Some of the data and eResearch support agencies and initiatives include:

- Queensland Cyber-Infrastructure Foundation (QCIF) – data compute, storage and visualisation;
- The Australian Research Data Commons – an initiative working with higher education, government and industry partners to build a national and collaborative research data commons. Formed through the amalgamation of the Australian National Data Service (ANDS), National eResearch Collaboration Tools and Resources (NeCTAR) and the Research Data Services (RDS);
- Australia’s Academic and Research Network (AARNet) – research network infrastructure;
- Australian Access Federation (AAF) – an identity broker providing subscribers with a national single sign-on that allows individuals across many different organisations to collaborate and access online resources;
- Australasian eResearch Organisations (AeRO) – a national association of eResearch organisations;
- National Collaborative Research Infrastructure Strategy (NCRIS) – funding for research infrastructure projects; and
- Tinker – a prototype Data Enhanced Virtual Laboratory (DEVL) to support researchers in the Humanities, Arts and Social Sciences.

**Description of the research data collection**

QUT’s research data collection uses a hybrid model of both centralised e.g. the Research Data Storage Service, and decentralised e.g. faculty and institute physical data asset registers, repositories and data storage solutions. Additionally, solutions managed by external agencies, such as QRISdata and CloudStor are used.

QUT’s Research Data Finder consists of metadata records that describe published research datasets, with or without (in the case of mediated datasets) links to the research datasets themselves. Metadata records are stored centrally in Research Data Finder, with the associated datasets stored across a range of storage solutions, including CKAN (Comprehensive Knowledge Archive Network). CKAN was implemented in 2016 as underpinning data storage software for Research Data Finder (small to medium-sized datasets only).

QUT metadata records are harvested by the Australian Research Data Commons (ARDC) and published in their national data registry, Research Data Australia.

**QUT data repository**

The Researcher Data Finder repository is based on the VIVO repository system which provides an interface for ingest and resource discovery.

Research Data Finder uses a local modification of the Registry Interchange Format – Collection and Services (RIF-CS) metadata schema supported by the ARDC. Metadata records are created for collections (data), activities (research projects), parties (people and institutions), and services (e.g. sensor machines) with bi-directional links created between the records to facilitate information discovery.
Research Data Finder is compliant with the FAIR Data Principles and can support publisher requirements for researchers to publish data in conjunction with manuscript submissions. Dataset records in Research Data Finder can link to related publications in QUT ePrints, as well as external websites for other types of information. Digital Object Identifiers (DOIs) are assigned to research datasets to provide a permanent link to the dataset and to facilitate open access of data and encourage data citation. Creative Commons and other licenses may be applied to datasets, facilitating rights management for data owners.

External Research Dataset Service

QUT Library provides an External Research Dataset Service by which it facilitates access to published datasets for researchers and HDR students for the purpose of QUT research. The Library will negotiate open access if possible or may use Library Resource Allocation funds to purchase commercially available research datasets. More information, and the list of datasets, is available from the QUT Library web site.

Services

Research data management services supporting the data repository and data collections include:

- Data management information and documentation via QUT websites, including the Digital Workplace (staff) and HiQ (research students)
- Data management workshops and drop-ins
- Advice and assistance with creating data management plans using the Data Management Planning Tool
- External Research Dataset Service
- IFN001: Advanced Information Research Skills (AIRS)

Responsibilities

Research data management is a shared responsibility across senior management, researchers (including Chief Investigators), Heads of School and Institute Directors. This responsibility lasts throughout the research data management lifecycle. The Roles and Responsibilities Table attached to the policy describes the responsibilities.

Access

Research Data Finder provides two levels of access:

- QUT-wide access for restricted collections; and
- Open access for sharable collections.

Metadata records for these collections are indexed in Research Data Australia, the national registry of research data and can also be made available to an international audience through Clarivate Analytics’ Data Citation Index.

Access to the External Research Dataset Service is via the Research Data Librarian based at the Kelvin Grove Campus. Some datasets to which access is facilitated by an external body, such as the Household, Income and Labour Dynamics in Australia Survey (HILDA), the Longitudinal Study of Indigenous Children (LSIC) and the Longitudinal Study of Australia’s Children (LSAC) require the
QUT researcher to sign a Confidentiality Deed Poll prior to access being granted. Some datasets are used on campus only, for example, Morningstar Direct, while others may be borrowed from the Library collection by negotiation with the Library Resource Services team.

**Related Documents**
- University Policy MOPP D/2.8 Management of research data
- Data Management Planning Tool – Help and FAQs
- About Research Data Finder
- The Fourth Paradigm: Data-Intensive Scientific Discovery
- Guidelines on the use of the Data Management Planning Tool for Higher Degree Research Students

**Modification History**

<table>
<thead>
<tr>
<th>Date</th>
<th>Sections</th>
<th>Source</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 December 2010</td>
<td>New guideline</td>
<td>ADIRRS</td>
<td>Final</td>
</tr>
<tr>
<td>7 December 2012</td>
<td>Formatting update</td>
<td>PA to Associate Directors, Library Services</td>
<td>Format overhaul in accordance with new Library look and feel</td>
</tr>
<tr>
<td>6 November 2013</td>
<td>Section 3 Background and Section 8 Responsibilities</td>
<td>ADIRRS</td>
<td></td>
</tr>
<tr>
<td>20 December 2013</td>
<td>All</td>
<td>IRC DLS (Acting)</td>
<td>Endorsed Executive approval</td>
</tr>
<tr>
<td>23 December 2013</td>
<td>All</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 February 2017</td>
<td>All</td>
<td>Research Data Librarian</td>
<td>Revised and submitted to LLT for approval</td>
</tr>
<tr>
<td>30 May 2019</td>
<td>All</td>
<td>Research Data Librarian</td>
<td>Revised and submitted to IRC for approval</td>
</tr>
</tbody>
</table>