Collaborating to develop research data management services and collections

Martin Borchert
Associate Director, Library Services (Information Resources and Research Support)
Queensland University of Technology Library

Stephanie Bradbury
eResearch Access Coordinator

Philippa Broadley
Research Data Librarian

Thanks also to the Australian National Data Service (ANDS) for project funding and support
Agenda

- Facilitating open access to institutional research data “data out”
- Policy, skills, infrastructure, identification, description and discovery, storage, services
- Facilitating institutional access to external research data “data in”
- Collections
- Library as data broker
The Researcher Support Strategy is a framework for the development and delivery of a range of services.
Broader policy and funding environment

- Australian Code for the Responsible Conduct of Research
- National Collaborative Research Infrastructure Strategy (NCRIS)
- Australian National Data Services (ANDS)
- National Health and Medical Research Council (NHMRC) OA policy
- Australian Research Council (ARC) OA policy
“Managing data” and
“Data out”
Advocacy can happen at all levels.

It is important for open access and the institutional repository to be made core business, reflected in committee papers.
The QUT Management of research data policy was released in early 2010.

It was one of the earlier such policies in Australia.

The policy is structured to mirror, and ensure compliance with, the requirements of the Australian Code for the Responsible Conduct of Research and in particular Section 2 Management of research data and primary materials:

Responsibilities of institutions – to retain research data and primary materials – provide secure research data storage and record keeping facilities – identify ownership – ensure security and confidentiality.

Assist researchers to – retain data - store data.

It deals with issues of: research data planning, intellectual property, record keeping, storage, retention and disposal, privacy and confidentiality, and access and re-use. It promotes a centralised data registry and decentralised data storage.
Library roles

- Research Support Coordinator
- Scholarly Communications Librarian
- Repository Resource Librarian
- Data Librarian
- Applications and Development Team Leader
- Programmer
- Liaison Librarians with a lead role in research support x 8
- Research Data Team
- Associate Director

There are specialised positions supporting research and the repository. It is important to make research support core business with key positions and working groups.

The Data Librarian has the responsibilities to:

- Develop QUT Research Data Finder
- Coordinate service delivery around it
- Provide training and support
- Advocate good practice

The Research Data Team is led by the Data Librarian and includes also three Liaison Librarians with a lead role in research support.
Staff development

- Research support seminars
- Data Team community of practice
- Train the trainer
- Repository training
- RIF-CS metadata training
- Co-trainer
- Peer assessment
- Resources wiki

Staff development is ongoing over 4 years now.

We introduced our Liaison Librarians to a broad range of research support issues:

Seminars
Workshops for researchers

Liaison Librarians first attended in the audience, then co-presented and then were primary presenters. We use a system of peer assessment / buddying system.
Promotion, training, information and support is ongoing.

Promotion is via Divisional and Library newsletters.

Monthly data management workshops are held, since 2010. About 200 researchers attend each year. These workshops use the Data Management Guidelines and Data Management Plan as the basis for conversation.

Workshops can be registered for online and via the Research Students Centre (graduate office) website.

There is a supporting research support web site with data management resources, including all training resources.

The site also includes contacts for further support. The Liaison Librarians are front-line support. The Research Support Team provided service development and second tier support.
IFN001 Advanced Information Research Skills (AIRS) is a mandatory unit for PhD students, and optional for Masters Research and Honours.

It is available in face to face and online modes, or in combination, also self-paced.

It covers a wide range of topics and there is a data management module included also.

About 1,000 higher degree research students complete AIRS each year.

This is a great strategy to skill from the students up.
As per my other presentation at this conference on ensuring academic engagement in QUT ePrints, the institutional open access repository of research outputs, QUT has a 10 year history of supporting and promoting green open access.

We are hoping to leverage this broad base of support for open access and apply it to open data. It’s early days still, but there exists a culture and a conversation.
Here is a high level illustration of research data infrastructure at QUT and its connect with national infrastructure.

QUT Research Data Finder is at the centre. It is the institutional data registry service.

Below are data storage, metadata capture points, and intersections with other university systems to source information about projects and people.

Storage comes from QUT and also in future from cloud services such as research service providers in Australia.

Above are external agents – National Library of Australia for people records, and Research Data Australia (national data registry).
The University is developing approaches to make storage on university supported and partner services more usable and robust, so researchers are not attracted to informal solutions.
We have a four tier data architecture.
Managing the University’s research data assets is taking on more of a focus. Research funding organisations like the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) are starting to expect data created by funded research, to be made openly managed.

QUT Library and ITS are developing QUT Research Data Finder which will be a data repository – registry service.

It is based on the VIVO open source system from Cornell University.

It pulls information about people and groups, research activities and grants from existing sources of truth.
There are numerous contact points to discover opportunities for identifying and describing datasets.

Notification of intent, grant application, ethics application etc.
This diagram is the workflow for registering new data collections at QUT.

It is difficult to see on screen here, but perhaps you can look at it from the slides later.

Thanks go to Philippa Broadly, Data Librarian, for creating this.
Open data records are sent to ANDS Research Data Australia national data catalogue.

The Registry Interchange Format – Collection Services (RIF-CS) metadata schema is used which uses a network of four inter-related records:

Party record – people and groups
Activity record – research projects
Services – software and machines
Collections – dataset description
The Library and HPC partner to develop and deliver services. We learning about each other’s services so we can provide referrals. We are located in the same buildings.

HPC provides a range of data services.

The image here is of an optiportal which is a large data visualisation system.
“Data in”
Engagement on data management happens at numerous levels.

Here are some examples of LibGuides on using data in research; one of datasets available in the social sciences, and another of data mining.

The aim is to introduce researchers to data centric research.

Researchers want to know about good practice and storage options, but no so much about sharing options (yet).
Another means to being data into the university to support research is to incorporate online commercial data services into the library’s online collections just like we do with publisher databases etc, and to support them in similar ways.

The investment can be large.

QUT started out first with working with Business on financial and economic datasets.

Most are remotely hosted by the publishers.

It requires the same collaboration with faculty to ascertain their needs.

Some are primarily supported by key academics, although librarians are learning the new business.
The library offers an “external dataset service” by which we act as data brokers on behalf of researchers. We pay, license and facilitate access.

We have about 70 usages of the service each year.

The service promotes efficiencies as the university licenses each dataset only once.

Researchers can then spend their time researching, instead of administering the procurement.

Finally, I’ll also add at this point that libraries can add open datasets to their collections, web site, catalogue etc to facilitate advocacy and use of these in the same way as we do for subscribed database content.

This is another aspect to promoting open access – open data.