



TROPICAL DATA HUB

Best Practice Guidelines for Research
Data Management at JCU

The exponential growth in digital data requires new mechanisms and services to manage data effectively. The benefit of effective data management practices include the increased re-use of publically funded research data increasing the profile of the data provider. This guideline provides steps for effectively managing your research data in accordance with the *JCU Code for the Responsible Conduct of Research* and the *JCU Research Data Management Policy (DRAFT)*.

Making data discoverable requires attaching some descriptive information (or metadata) to the dataset. The Tropical Data Hub is JCU's metadata repository which attaches standardized metadata fields to each deposited research dataset. A dataset with no metadata is analogous to a tin can with no label in a cupboard. The Tropical Data Hub will serve as a showcase for JCU's research output and enhance our tropical position.

Step 1 Data Management Plan

Effective data management starts with a data management plan. Download a Data Management Plan template from the *Data Management at JCU* eSpace at <https://espaces.edu.au/data-tools> or use the online data management planning tool in the JCU Research Data site (<http://research.jcu.edu.au/researchdata/dashboard>). This document will assist you in undertaking your data management responsibilities at JCU. The template is informed by the requirements for managing research data set out in the *JCU Code for the Responsible Conduct of Research*.

A data management plan is a document outlining how research data and associated materials will be managed, stored, documented and secured throughout a research project as well as planning for what will happen to the data and materials after completion of the project. This includes retention and disposal, archiving, accessing, sharing or publishing the data, and any conditions or restrictions for sharing the data. The plan is intended to provide descriptive details of the data, the processes, decisions, as well as identifying roles and responsibilities.

Funding bodies are increasingly requiring a data management plan and you can use the JCU Data Management Plan as a template.

Step 2 Data Storage and Back-Up

JCU policy requires that all digital research data be backed-up securely on the HPRC (High-Performance and Research Computing). This is the recommended storage solution for data management at JCU as it offers a safe and secure way to store your research data. To create a HPRC account please go to <https://plone.jcu.edu.au/hpc/getting-an-account> and follow the instructions under HPRC Account Information.

The HPRC account is a private computing account and requires a JCU Username and Password to login. Follow responsible practice and do not share your username and password with anyone else. If you want to share a directory/folder on the HPRC, each of the people in your group will have to create their own HPRC

account and you will have to forward a list of those HPRC account holders that are to be given access to that folder to the HPRC Systems Manager, Dr Wayne Mallett, on ext 15084 or hpc.operator@jcu.edu.au

Accessing the HPRC

For transferring your files to your HPRC there are several options but the most common are either to set up a network drive from your computer or to use a file transfer program.

Set up a Network Drive

- For PCs go to the Tools drop down menu in an Explorer window and select Map Network Drive. Pick a drive (e.g. Z) and type in the address box: `\\dmf.jcu.edu.au\jc123256`
- Mac OSX: `smb://dmf.jcu.edu.au/jc123456`

Use a secure file transfer program

- Many SFTP clients exist. Many researchers use FileZilla for secure transfer of files. You can download the FileZilla client (for all platforms) from <http://filezilla-project.org/>

The program requires you to provide the following details to Quickconnect:

1. Host: `sftp://zodiac.hpc.jcu.edu.au`
2. Username and password: your JCU credentials
3. Port: 8822 (or 22)
4. Click Quickconnect

The window will split with your computer on the left and your HPRC account on the right. Just drag and drop files.

For more information or assistance see the online documentation at <http://plone.jcu.edu.au/hpc/user-guides> or contact the eResearch Centre.

Options for backup storage

To keep your data secure, we recommend you keep 3 copies of your data and keep them in separate places. Some suggested storage solutions are listed below.

Internal Storage Solutions:

High Performance Research Computing (<https://plone.jcu.edu.au/hpc>): Available to all researchers at JCU. Large storage capacity and is the preferred location for all working research data at JCU.

eGRS (<https://egrs.jcu.edu.au>). This is for postgraduate students wanting to store their research documents and share them with their supervisor(s).

RDSI JCU (<http://eresearch.jcu.edu.au/data-management/rdsi-storage-at-jcu>): JCU hosts an additional node of the Research Data Storage Infrastructure (<http://www.rdsi.edu.au>). Storage is available by application for research data collections of significance.

Research Data Repository (<https://research.jcu.edu.au/researchdata/dashboard>): This is for data records and research data that is to be archived e.g. datasets associated with publications and data from funded projects. Smaller datasets (< 100 MB) can be stored directly in this site; Larger datasets, the Research Data team will assist with setting up long term storage.

External File Sharing and Storage Solutions:

eSpaces (<http://www.espaces.edu.au>). A researcher's website where they can collaborate on small projects with external partners.

Cloudstor+ (<https://cloudstor.aarnet.edu.au/plus/>). Cloud storage offered by AARNet. Up to 100GB of storage available free to Australian-based researchers. Operates in a similar fashion to Dropbox and is for individual researchers who want to manage their own sharing. Apps are available to synchronise between cloud, computers and mobile devices.

QRIScloud (<http://www.qcif.edu.au/>). A cloud-based storage and compute solution offered by the Queensland Cyber Infrastructure Foundation (<http://www.qcif.edu.au>). See the QCIF website for more information.

Cloudstor (<https://cloudstor.aarnet.edu.au/filesender/>). This is a secure way to share large files. All that is required is your JCU Username and Password to login. The maximum file size is 100GB. Individual file storage time is limited to a maximum of 20 days. Maximum number of recipients for a file is 100.

Commercial Storage Solutions:

There are many cloud providers offering some amount of free space with additional capacity available for purchase. Most include apps to enable synchronisation between the cloud, your desktop, laptop and mobile devices. These include but are not limited to:

- **Dropbox** (<http://dropbox.com>): 2GB for free.
- **Google Drive** (<http://drive.google.com>): 15GB free shared across Google Drive, Gmail and Google+ Photos. Need to sign up for a Google account to access. Will work with apple and android
- **OneDrive** (<http://onedrive.live.com>): 7 GB available for free. Need to sign up for a Microsoft account to access. Used to be known as SkyDrive.
- **iCloud** (<http://www.icloud.com>): 5GB for free. Need to sign up for an Apple account to access.
- **Box** (<http://www.box.net>): 10GB for free.
- **ADrive** (<http://www.adrive.com>): 50GB for free.
- **Bitcasa Infinite Drive** (<http://www.bitcasa.com>): 10GB for free.

There are many more – these are the current (2014) popular ones. A quick web search for “free cloud storage” will yield lots of results and include reviews to help you decide which might be best for you.

Step 3

Recommended File Formats

During the research project follow consistent data collection and file naming conventions. This will require pre-planning and be outlined in your Data Management Plan. Use high quality, open file formats where possible. Important formats include JPEG 2000, PDF/A, TIFF/EP, TIFF/IT, Open Document Format (ODF), and Office Open XML. Table 1 provides a list of acceptable file formats for sharing, re-use and long term preservation of research data.

Table 1: Recommended file formats for data sharing and preservation (from the UK Data Archive, <http://www.data-archive.ac.uk/create-manage/format/formats-table>).

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
<p>Quantitative tabular data with extensive metadata</p> <p>a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data</p>	<p>SPSS portable format (.por)</p> <p>delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information</p> <p>some structured text or mark-up file containing metadata information, e.g. DDI XML file</p>	<p>proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)</p>
<p>Quantitative tabular data with minimal metadata</p> <p>a matrix of data with or without column headings or variable names, but no other metadata or labelling</p>	<p>comma-separated values (CSV) file (.csv)</p> <p>tab-delimited file (.tab)</p> <p>including delimited text of given character set with SQL data definition statements where appropriate</p>	<p>delimited text of given character set - only characters not present in the data should be used as delimiters (.txt)</p> <p>widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)</p>
<p>Geospatial data</p> <p>vector and raster data</p>	<p>ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn)</p> <p>geo-referenced TIFF (.tif, .tiff)</p> <p>CAD data (.dwg)</p> <p>tabular GIS attribute data</p>	<p>ESRI Geodatabase format (.mdb)</p> <p>MapInfo Interchange Format (.mif) for vector data</p> <p>Keyhole Mark-up Language (KML) (.kml)</p> <p>Adobe Illustrator (.ai), CAD data (.dxf or .svg)</p> <p>binary formats of GIS and CAD packages</p>
<p>Qualitative data</p>	<p>eXtensible Mark-up Language (XML) text according to an appropriate Document Type</p>	<p>Hypertext Mark-up Language (HTML) (.html)</p>

textual	Definition (DTD) or schema (.xml)	widely-used proprietary formats, e.g. MS Word (.doc/.docx)
	Rich Text Format (.rtf)	
	plain text data, ASCII (.txt)	some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
Documentation and scripts	Rich Text Format (.rtf)	plain text (.txt)
	PDF/A or PDF (.pdf)	some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx)
	HTML (.htm, .html)	
	OpenDocument Text (.odt)	XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0
Digital image data	TIFF version 6 uncompressed (.tif)	JPEG (.jpeg, .jpg) but only if created in this format
		TIFF (other versions) (.tif, .tiff)
		Adobe Portable Document Format (PDF/A, PDF) (.pdf)
		standard applicable RAW image format (.raw)
		Photoshop files (.psd)
Digital audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) but only if created in this format
		Audio Interchange File Format (AIFF) (.aif)
		Waveform Audio Format (WAV) (.wav)
Digital video data	MPEG-4 (.mp4)	
	motion JPEG 2000 (.mj2)	

Step 4

Metadata Management and Data Publishing

Research Data Repositories are growing in number and are being harvested by Thomas Reuters for their Data Citation Index.

For NHMRC and ARC grant applications you are now required to include a paragraph on what you plan to do with your data. This statement must be unique to your application and cannot be a form response that is included in all JCU applications.

If your data is not to be made public for ethical, confidentiality or contractual reasons, your research data still needs to be described and stored securely and privately, but a link to it will not be provided to the public.

Create a metadata record in JCU's Research Data (<https://research.jcu.edu.au/researchdata>) describing your data and pointing to its location for sharing and re-use. All research data can have a metadata record created for it and be made publicly available. Research Data is harvested regularly by the Australian Research Data Commons, and the data record will also be available on Research Data Australia (<http://researchdata.andso.org.au>). Publications and websites relating to the research data can be linked to the metadata record.

Writing up your data management plan, storing your data on the HPC, and creating a metadata record of it in the Tropical Data Hub will fulfill all of your research data management requirements.

Contact the Research Data team (researchdata@jcu.edu.au) at eResearch Centre for advice on how to store and preserve your digital research data at JCU and how to publish your data on the Tropical Data Hub, JCU's central research data repository. The Research Data Management website is at <http://espaces.edu.au/data-tools> and provides a range of tools, policies and guidelines for data management.

Reference Relevant Policies, Documents and Links

The JCU Code for the Responsible Conduct of Research

http://www.jcu.edu.au/policy/research/allresearch/JCUDEV_009786.html

The JCU Research Data Management Policy (DRAFT)

The JCU Intellectual Property Policy

http://www.jcu.edu.au/policy/research/allresearch/JCUDEV_009787.html

The JCU Data Management Information

<http://eresearch.jcu.edu.au/data-management>

<http://espaces.edu.au/data-tools>