Australian National University

Digital literacy training

ANU Research Data Management Manual

Managing your Research Data at
The Australian National University

2022

Libguide at ql.anu.edu.au/data

Table of Contents

1 Introduction

Research Data Management is an integral part of modern research. Almost all researchers manage various forms of digital data — measurements from instruments, survey records, multimedia, or documentation.

Research Data management involves activities such as backups, collaborative work, data security, and archiving.

Managing your data allows you to work more efficiently, produce higher quality data, achieve greater exposure for your research, and protect your data from being lost or misused.

This document gives an overview of research data management at The Australian National University.

1.1 Objectives

- Understand what research data is and why it needs to be managed.
- Appreciate legal, institutional and funding issues related to data.
- Learn how various data management methods can help you work more effectively with your data.
- Develop an awareness of existing data management services at ANU.
- Write a data management plan.

1.1.1 Research Data Management at ANU

The ANU Library Digital Literacy Training offers introduction to Research Data Management[1] at The Australian National University. The Research Data Management manual can be accessed at Research & Learn section of the ANU Library website[2].

1.1.2 Outline

While this document is intended to be read sequentially, it can also be used as a handy reference. For instance, if your aim is to quickly put together a Data Management Plan (see section 6), or begin using the ANU Data Management Services (see section 5), you can skip ahead to those sections and refer back to the earlier sections as needed.

[1] Workshop schedule and bookings at https://services.anu.edu.au/training/data-management
[2] http://anulib.anu.edu.au/research-learn/research-data-management

An outline of this document and a summary of the key points are as follows:

1.1.3 Research Data Management

- All researchers have digital data. While publications can be the most commonly encountered forms of digital data, researchers may also make use of data concerning measurements, survey responses, multimedia references, etc.
- Data Management can be loosely defined as "Anything outside of actually using the data." For example, organization, protection, and distribution of data.
- A Data Management Plan (DMP) is a document that describes what data will be created during a project, and how it will be managed.

1.1.4 Benefits and Requirements

- The key motivation for doing good data management is so you can spend more time using the data to comply with data management policies.
- There are a number of policies relating to data management: the *ANU Policy for Responsible Practice of Research*, the *Australian Code for the Responsible Conduct of Research* and the *ARC Funding Agreement for Discovery Projects*. Most relate to the ethics and long term storage (archiving) of data.

1.1.5 Methods of Research Data Management

- Data Organization: Description of various methods for working more efficiently with data.
- Data Administration: Discussion of methods to protect and improve the quality of data.
- Data Archiving and Sharing: Details of Data Archiving for preservation, and Data Sharing for exposure and open research.

1.1.6 ANU Research Data Management Services

- Local IT Support Staff (LITSS) – provide your computer and software. May also provide a fileserver for backups and a webserver.
- ITS Infrastructure as a Service (IaaS) – manage ANU central file storage (HomeDrive) and webserver.
- ANU Library Digital Literacy Training – provide training to support researchers.
- ANU Data Commons – ANU repository for long-term storage and dissemination of data.
- ANU Supercomputing Facility – High performance computing, visualization, and large data storage.
- Discipline specific archives – ADA (Social Sciences)

1.1.7 Writing a Data Management Plan

- An example can be found at the end of this manual (see section 6 and Appendix)

2 Data Management

This section defines key terms such as *data*, *data management*, and *data management plans*.

Other commonly used terms (such as fileserver, FTP, and Open Access) can be found in the Glossary (see section 7).

2.1 Data

Throughout this document, 'data' will refer to *digital research data*. Digital research data is any data that is created during research that can be stored on a computer. Data also includes field notes, analog recordings, and non-digital images as they can be converted to digital images. Physical data such as biological specimens, soil samples etc. are not considered.

Digital research data can additionally include:

- Numerical data: instrument measurements, survey responses.
- Documentation: Publications, experimental methods, field notes, analytical methods, technical reports, dataset descriptions.
- Digital images: photographs, diagrams, graphs.
- Digital audio: audio data, interviews, wildlife recordings, language recordings.
- Digital video: high-speed recordings, interviews.
- Configuration data: Configuration and optimization settings for simulation and *in-silico* experimentation.

Although not strictly data, you may also wish to consider the storage of any software developed to analyze the data, as an aid to being able to reproduce the results of your research.

2.2 Data Management

Data Management can generally be considered as any activity involving data outside of actually using the data.

Data management is best defined as any and all of the following examples:

- Organizing data into directories/folders and using meaningful filenames.
- Keeping backups of data in case you accidentally delete or lose data.
- Storing final state data in an archive.
- Making data available to others via an archive or website.
- Ensuring security of confidential data.
- Collaboratively creating and using data with other researchers.
- Synchronizing data between desktop, laptop, USB key, cloud storage, etc.
- Maintaining a bibliography and electronic copies of relevant literature.

Data management involves organizing, protecting and distributing the data. Data management does not produce results but is an unavoidable consequence of working with data. The aim is therefore to spend as little time doing data management as possible so that more time is spent using the data productively. Typically, people only do data management when it is needed and therefore tend to use the most obvious methods. The obvious methods are often the most inefficient, i.e., they are time-

consuming and error prone. Using more advanced and automated methods will reduce the amount of time spent managing data.

2.3 Data Management Plan

A 'data management plan' is a document that describes what research data will be created, what policies (funding, institutional, and legal) apply to the data, who will own and have access to the data, what data management practices (backups, access control, archiving) will be used, what facilities and equipment will be required (hard-disk space, backup server, repository), and who will be responsible for each aspect of the plan.

The best time to develop your data management plan is at the beginning of your research. Any time spent on creating a robust and easy to use data management framework will be rewarded many times over during your research.

Australian Research Council (ARC) requires that you provide an outline of your data management plan when applying for a grant. A guide can be found at Australian Research Data Commons (ARDC) website.[3]. The Australian National University is a participating institution of the DMPTool service[4], provided by the California Curation Centre of the California Digital Library, for the creation and management of data management plans.

[3] https://www.ands.org.au/guides/data-management-plans
[4] https://dmptool.org

3 Benefits and Requirements

This section describes the benefits to researchers of data management as well as some of the institutional and funding requirements related to data management.

3.1 Benefits of Data Management

Research data is a valuable asset and data management should be seen as a necessary part of good research. The benefits of data management are:

- **Efficiency** of research through good organization, collaboration and documentation of data.
- **Protection** of data against becoming lost, unusable, forgotten, or improperly released.
- **Quality** of data through procedures to ensure data is accurate and authentic.
- **Exposure** of research outcomes through collaborations with others and dissemination of results and publications.
- **Citation** of data through identifiers to enable data metrics, transparency and data acknowledgement.
- **Reproducibility** of experimental and computational outcomes enables easy validation and verification of results.

3.1.1 Efficiency

Data management can improve the efficiency with which you work with your data. Typically organization and documentation of data are only done when they are absolutely necessary. Using software for version control and collaboration, and documenting data when it is created, will save time and allow you to work more efficiently with your data.

3.1.2 Protection

Data is a valuable asset so it is worthwhile protecting it from accidental loss or improper release.

Most people recognize the risk associated with losing data through accidental deletion and equipment failure, theft or destruction. Multiple and backup copies are therefore often kept for important data, but researchers should also consider using automated backup facilities to back up all their data.

Data management also protects the data from being improperly released. This is important where the data contains confidential or commercially valuable information.

Improperly releasing data can violate privacy laws, confidentiality agreements, and possibly void intellectual property claims. It is therefore important to have well defined access rules for your data.

3.1.3 Quality

It is important to ensure the quality and authenticity of data that will be used for analysis and generating conclusions. Inaccurate data can invalidate results and conclusions resulting in lost time and damaging reputations.

Likewise, making any software or data analysis scripts available along with your data can help substantiate your results. This is particularly important when dealing with large datasets or complex analyses.

It is also important to ensure the authenticity of data to avoid claims of plagiarism and ownership disputes.

3.1.4 Exposure

Placing your publications and research data in an archive greatly increases the exposure of your research. Research has shown that Open Access (OA) publications receive 2-3 times as many citations as articles that are only available via journal subscription[5].

3.1.5 Citation

Data citation refers to the practice of providing a reference to data in the same way as researchers routinely provide a bibliographic reference to outputs such as journal articles, reports and conference papers. Citing data is now recognised as one of the key practices leading to recognition of data as a primary research output[6].

3.1.6 Reproducibility

Reproducibility of results and independent verification is an important criterion for research[7]. Ensuring a proper record of provenance and context trail facilitates recreation and analysis of critical research hypotheses and data. Again, making any software developed available can aid the verification process.

3.2 Benefits of Data Archiving and Sharing

Data archiving and sharing makes for good research as it allows for independent verification of results and conclusions and further analysis through the reuse of data.

An excellent list of the benefits of data archiving and sharing is given by the ICPSR's *Guide to Social Science Data Preparation and Archiving*[8]:

- Reinforces open scientific inquiry. When data are widely available, the self-correcting features of science work most effectively.
- Encourages diversity of analysis and opinions. Researchers having access to the same data can challenge analyses and conclusions.
- Promotes new research and allows for the testing of new or alternative methods. Examples of data being used in ways that the original investigators had not envisioned are numerous.
- Improved methods of data collection and measurement through the scrutiny of others. Making data publicly available allows the scientific community to reach consensus on methods.

[5] Stevan Harnad and Tim Brody, "Comparing the Impact of Open Access (OA) Vs. Non-OA Articles in the Same Journals," *D-Lib Magazine* 4, no. 6 (2004).
[6] http://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation
[7] Ewa Deelman et al., "Workflows and E-Science: An Overview of Workflow System Features and Capabilities," *Future Generation Computer Systems* 25, no. 5 (2008).
[8] ICPSR. *Guide to Social Science Data Preparation and Archiving (4th Edition).* Inter University Consortium for Political and Social Research (ICPSR), 2009 www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

- Reduces costs by avoiding duplicate data collection efforts. Some standard datasets, such as the General Social Survey and the National Election Studies, have produced literally thousands of papers that could not have been produced if the authors had to collect their own data. Archiving makes known to the field what data have been collected so that additional resources are not spent to gather essentially the same information.
- Provides an important resource for training in research. Secondary data are extremely valuable to students, who then have access to high-quality data as a model for their own work.

3.3 Funding and Legislative Requirements

There are some key funding and legislative requirements relating to data management. Basic data management is required by the *Australian Code for the Responsible Conduct of Research*. Compliance with the *Code* is already a requirement for ARC and NHMRC funding and is likely to be mandated by other funding bodies, government and institutions in the near future.

The following sections summarize the policies relating to data management.

3.3.1 ANU Responsible Practice of Research Policy (summary)

- ANU policy document[9]
- Data management should comply with the *Privacy Act (1988)*[10]
- Research results should be open to scrutiny. Non-confidential data related to publications must be made available
- Data must be retained for at least 5 years. Retention must comply with the *Archives Act (1983)*
- Researchers should not unnecessarily enter into research agreements that limit access to information
- Researchers are responsible for data security
- Supervisors must ensure the validity of data gathered by their students
- Research misconduct includes: fabrication, falsification, and interference with data

3.3.2 Australian Code for the Responsible Conduct of Research

- The Code document[11]
- Researchers should retain research data and primary materials for sufficient time to allow reference to them by other researchers and interested parties. For published research data, this may be for as long as interest and discussion persist following publication.
- Research data should be made available for use by other researchers unless this is prevented by ethical, privacy or confidentiality matters
- Research data should be retained for at least the minimum period specified in the institutional policy.

[9] https://policies.anu.edu.au/ppl/document/ANUP_007402
[10] *Privacy Act (1988)*, www.comlaw.gov.au/Series/C2004A03712
[11] https://nhmrc.gov.au/research-policy/research-integrity/release-2018-australian-code-responsible-conduct-research

- If the results from research are challenged, all relevant data and materials must be retained until the matter is resolved. Research records that may be relevant to allegations of research misconduct must not be destroyed.
- The institutional policy on the secure and safe disposal of primary materials and research data must be followed

3.3.3 ARC Funding Agreement for Discovery Projects

- Data from research in the social sciences should be archived with the Australian Data Archive (ADA, see section 20.6 of the funding agreement commencing 2019) within two years for secondary use by other investigators[12]
- The ARC strongly encourages the depositing of data arising from a Project in an appropriate publicly accessible subject and/or institutional repository, such as ANU Data Commons (see section 5.5).

3.3.4 NHMRC Principles for Accessing and Using Publicly Funded Data for Health Research

- The Principles provide information and guidance for researchers and organisations when researchers seek permission from organisations to access and use data for their research[13]
- The Principles have been developed by NHMRC on the advice of researchers, consumer representatives and organisations that hold data. They represent a common view about sharing the data, and the roles and responsibilities of all parties

[12] https://www.arc.gov.au/grants/grant-application/fundinggrant-agreements/discovery-program-fundinggrant-agreements
[13] https://nhmrc.gov.au/sites/default/files/documents/reports/principles-publically-funded-data.pdf

4 Methods of Data Management

This section outlines some of the methods of data management.

4.1 Data Organization

Data organization is about working more efficiently with data. Creating and using data requires some level of data organization. Often this organization becomes time consuming and error prone, in which case automated data organization methods should be considered.

Each section lists the standard methods of dealing with data organization and their drawbacks. Some automated and more efficient alternatives are suggested, but keep in mind that they often require some configuration and familiarization with the software. If the standard methods are adequate for your needs, then it is best to continue using them. If you think you are spending too much time organizing your data, then you should consider looking into the advanced methods.

4.1.1 Bibliography Management

Creating a bibliography manually is time consuming and error prone. Journals and conferences will usually specify a particular citation style, so it is best to generate citations automatically to save time and avoid errors. Furthermore, researchers often have hundreds of academic articles stored on their computers as part of the literature review. Finding a particular article can become time consuming.

There are a number of software-based reference management tools that automate citations and bibliography creation when writing an article. They also organize references into a database, making it easy to sort and search. Most of these programs also offer the ability to search online academic databases, such as *CiteSeer*, *PubMed, Scopus, ProQuest and others*.

EndNote is a popular reference management tool and ANU has an institutional license which allows staff and students to install *EndNote* on their office and home computers, laptops and mobile devices[14]. The ANU Library Digital Literacy Training offers courses in *EndNote*.

Unfortunately, *EndNote* does not run on Unix and cannot manage BibTex bibliographies; therefore, LaTeX authors and Unix users can use *JabRef*, which is a free program, runs on all operating systems[15]. *JabRef* can also import and export BibTeX's and *EndNote's* database formats.

Mendeley[16] is a free reference management application with an emphasis on researcher collaboration. The ANU Institutional Edition[17] offers enhanced Mendeley access for all ANU staff and students, including increased personal and shared storage, and the capacity to create an unlimited number of private, shared groups with up to 100 members.

[14] *Endnote*, www.endnote.com
[15] *JabRef*, http://jabref.sf.net
[16] *Mendeley*, https://www.mendeley.com
[17] *ANU Institutional Edition of Mendeley* https://anulib.anu.edu.au/research-learn/writing-referencing/citing-referencing-your-writing

Zotero[18] is a free and open-source bibliography management tool which runs on all operating systems and has support for exchanging bibliographic databases with both BibTex and *EndNote*. Zotero can be downloaded or connectors and plug-ins can be installed[19]. The Zotero bibliographic database can also be shared using access to ANU Homedrive (see section 5.2.1).

4.1.2 File Transfers and Remote Access

Researchers collaborating on projects will often need to share primary data and preliminary results; hence, it is often necessary for them to transfer data between computer systems. Researchers may also wish to transfer data stored on their university computer from outside the university, such as when overseas.

The most common method for transferring files is with email attachments, but there are limits to the size of files that can be transferred. Removable data storage media, such as USB keys and portable hard disks can transfer large amounts of data, but require the researcher to physically carry the data to its destination.

Specialist services such as CloudStor[20], provided by AARNET can be used to transfer files without needing to use portable media devices. While primarily a service for the exchange of data between AARNET member institutions you can generate a voucher or magic code which you can send to a colleague at a non AARNET institution to allow them to use Cloudstor to either send file to you or receive data from you.

To assist good data management, the ANU provides local area network and Internet access to Homedrive, a central storage space on which each member of the University is allocated file space to store personal files (4.5GB for students and staff). Homedrive is accessible from any campus computer or computers outside campus[21] (see section 5.2.1).

ANU-provided SharePoint (see section 5.2.2), allows data to be accessed and modified with a web browser.

4.1.3 File Synchronization

Often researchers will work on their university desktop as well as a laptop, and possible a home computer. Typically files are just copied back and forth between the computers. This is the most obvious method but has a number of drawbacks.

- It is time consuming to manually copy files.
- You have multiple copies of data and you can easily lose track of which copy is the latest version.
- If both copies have been modified, then it is easy to overwrite some changes without knowing.

If you are synchronizing regularly or have lots of files to synchronize, then you should consider using file synchronization software. File synchronization software offers the following advantages over manual synchronization:

- Faster and requires less thought (usually just click a button).
- Automatically detects when two files have been modified and lets the user choose which one to keep. Some programs can also display the difference between the files.

[18] *Zotero*, www.zotero.org
[19] *Zotero download*, https://www.zotero.org/download
[20] https://cloudstor.aarnet.edu.au
[21] https://myfiles.anu.edu.au/

One of the most popular file synchronization programs is *WinSCP*, which is primarily for SSH and FTP transfers, but can also synchronize data[22].

rsync is another widely used open-source utility for incremental file transfer and synchronization[23]. It is cross-platform and can be used to generate 'snapshots' and regular backups.

The use of dedicated Version Control Software (see section 4.1.5) is another option for file synchronization.

Commercial, user-friendly, file synchronization services are becoming increasingly popular. *Dropbox*, for example, provides 2GB of data storage for free and also provides good user management tools to support collaborative work[24].

4.1.4 Collaboration

Many research projects are carried out collaboratively: between postgraduates and their supervisors; within departmental research groups; as cross-discipline research, and as inter-university research.

This is mutually beneficial as it improves access to funding; avoids repeating costly experiments; increases recognition through co-authorship; and can help lead to new research ideas.

For simple tasks this is usually done by transferring data by email, USB-key, or a network drive. Publications with multiple authors are often written this way – authors will take turns editing the document and email it to their colleagues, or the primary author will periodically email the latest version and their colleagues will reply with corrections and additions.

These methods are adequate for simple work and if there is only a small number of collaborators. It is worth considering using collaborative software tools, such as the ANU-provided *SharePoint* (see section 5.2.2) and/or version control software (see section 4.1.5)[25].

SharePoint[26] is a web-based collaborative platform that integrates with Microsoft Office 365. It's used to co-author documents, manage versions, apply workflows, and discover and share information.

4.1.5 Revision Control

When the data is constantly being edited, especially by multiple users, it is a good idea to implement some form of version control to keep track of changes. This can be as simple as appending a number to the end of a file after each major edit. For example:

- Journal_v1.0.tex, Journal_v1.2.tex
- Journal_Feb12.tex, Journal_May5.tex
- Journal_Feb12_John_DRAFT_WithSallysEdits_NewDiagram.tex

Such conventions are good for simple work but quickly become unmanageable when you have multiple authors or make lots of edits.

[22] *rsync*, en.wikipedia.org/wiki/Rsync
[23] *WinSCP*, en.wikipedia.org/wiki/WinSCP
[24] *Dropbox*, www.dropbox.com
[25] *TortoiseSVN*, tortoisesvn.net/
[26] https://anu365.sharepoint.com/teams/collab/help

The alternative is to use revision (or version) control software[27]. These programs are used extensively for software development but are also excellent for documentation, such as writing a paper with several authors. Version control software also provides access control, a collaborative work environment, synchronization between home/office/laptop computers, and a degree of data safety (although not as good as proper backups).

Such programs offer several advantages:

- The software requires you to input a description of the changes made, which makes it easier to pick up where you left off and for collaborators to see what you are doing
- You can be confident with making major changes as you can revert to an old version if you make a mistake. You can also easily compare two versions to help you find errors
- Useful for people who use more than one computer. It implicitly provides synchronization and is good for resolving conflicting changes

While the time required to learn the software may seem like a drawback, it is highly recommended in order to avoid regular problems with simple filename version control.

TortoiseSVN, for example, is a popular program that uses the *Subversion* system of version control[28]. It integrates with Windows Explorer making it one of the easiest version control programs to use.

GitHub is a version control repository and internet hosting service that 'fosters a fast, flexible, and collaborative development process that lets you work on your own or with others'. It offers distributed version control, source code management (SCM), access control and several collaboration features such as bug tracking, feature request, task management, and wikis for every project.[29] A similar product is *Bitbucket*[30].

While version control software is in some cases harder to set up, it provides more advanced version tracking. A distributed version control system like *Bazaar*[31] can be used with to manage documents and data.

Such tools also make it easier for any number of people to work on a document or code. It is also more efficient as everyone has access to the latest version and can make edits without conflicting with other people's changes. The entire history of the document is also stored, making it easier to revert to an older version and for users to see what has changed they last looked at the data.

4.2 Data Administration

This section covers methods for protecting data and ensuring its quality. Many of these methods are necessary for compliance with the requirements of Data Management (see section 3.3). Some are *policies*, such as security and access, and others are *practices*, such as backups, quality control and documentation.

[27] *Revision Control*, en.wikipedia.org/wiki/Revision_control
[28] *TortoiseSVN*, tortoisesvn.net/
[29] *GitHub*, https://en.wikipedia.org/wiki/GitHub
[30] *Bitbucket*, https://en.wikipedia.org/wiki/Bitbucket
[31] *Bazaar*, bazaar.canonical.com/en/

4.2.1 Backups

Making regular backups of data is probably the most important and, fortunately, one of the easiest tasks to manage.

While most people are aware of the risk and cost of losing data through hard drive failure or accidental deletion, it is still good to have a policy and schedule in place for maintaining data backups.

When considering your backup strategy, you need to consider:

- How often will you make backups
- How long will backups need to be stored
- How much hard-drive space, or number of DVDs, will be required to maintain this backup schedule
- If the data is sensitive, how it will be secured and (possibly) destroyed
- What backup services are available that meet these needs, if none, then what will be done
- Who will be responsible for ensuring backups are available

Backup security requires further mention. If the data is sensitive then it should not be stored on a computer that is connected to the internet, and preferably not connected to any network. If the data needs to be destroyed at the end of a project then consider what level is required – a hard drive will need to be overwritten several hundred times to ensure that no data can be recovered. Very high-level security institutions, such as Defence, require hard-disks to be physically destroyed and optical discs to be shredded.

The lifetime of backups should also be considered. Burned optical discs have average lifetime of two years, and five years if kept in a cool dark place.

If you are using a network drive then your data is probably already being backup up for you by IT staff. It is still a good idea to check with them to find out how often they backup, what is the maximum amount of data they can backup and how long they keep old backups.

You may need to maintain your own backups if:

- There are no services available to you
- You have valuable data that you do not trust with other people
- You have sensitive data that you cannot store on unsecure computers (medical records, data for defence projects, etc.)

If you do maintain your own backups you need to ensure that all media is correctly labelled with the date and contents of the backup. Keeping a diary of what you backup and when can be helpful if you have to reconstruct a data set after a drive failure.

Remember that backups can sometimes fail due to faulty media, and it is important to keep more than one copy of important data.

It is also useful to keep a duplicate version of your backup at different sites e.g. home and office.

4.2.2 Data Validation and Authentication

Your data will be used to obtain the results and conclusions of your research, so it is important to ensure its accuracy. Your data may also become an important dataset that is used by many others, so errors have the potential to hinder many research efforts.

It is therefore important to set up and fully document all policies and practices to ensure the accuracy and authenticity of your data. This can include:

- Calibration of instruments.
- Use of Computer Assisted Interviews (CAI)[32]
- Securing master copies to avoid accidental/intentional tampering.
- Data entry checks, such as two pass verification and range checking[33]

4.2.3 Documentation

It is important to document the experimental or data gathering methods. Other researchers may question your results or want to repeat/extend your research, so it is important to document this. The sciences already have a culture of keeping good lab notes and the social sciences often record their survey methodology. This is often done in a notebook, but you should also consider recording this information digitally or converting it manually[34]. This is important as notebooks are easily lost or put into storage when an academic or postgraduate student leaves. This information is far more useful if it is archived with the data it refers to. Scanners are available in most ANU library buildings. Some ANU research schools[35] provide ANU eNotebooks[36] (a rebranded name for LabArchives[37]). It is a cloud-based electronic filing system for managing files and datasets, and designed to protect data and comply with regulatory funding requirements. eNotebooks is currently available to all ANU users.

It is also valuable to document analytical methods. For example, if you write a script/macro/program to help analyze the dataset by producing graphs or statistics from your dataset.

4.2.4 Access Controls

Well-defined access controls help you comply with privacy and confidentiality policies and help maintain data authenticity by limiting who can modify data. The access controls may change throughout the life of the research project. Initially all data will usually be restricted to the research group, when the results are published the data may then be made available to other researchers.

Access controls can be defined on a per-user or per-data basis. When the data is active and there are a small number of people using the data then you will probably use per-user access permissions:

- None – has no access to the data
- Read – can read the data, but not modify
- Write – can read and modify data

[32] *Computer Assisted Interviewing*, en.wikipedia.org/wiki/Computer-assisted_telephone_interviewing
[33] *Data Validation*, en.wikipedia.org/wiki/Data_validation
[34] You can scan a notebook and use software to extract the writing, but it may not work well unless you have very good handwriting.
[35] Research School of Biology, John Curtin School of Medical Research and The Australian Cancer Research Foundation (ACRF) Biomolecular Resource Facility (BRF)
[36] ANU eNotebooks, https://enotebooks.anu.edu.au
[37] LabArchives, https://www.labarchives.com/

- Administrator – has the ability to modify others' access permissions

As an example, the principal researcher would have *Administrator* permissions over all data and may be the only one with *Read* permissions of confidential survey data. Research collaborators would have no access to the confidential survey data, *Read* access to de-identified survey data, and *Write* access to data analyses and publications.

How you set access permissions varies between operating systems. In Windows they are usually set by right-clicking on a file or directory and editing the security properties. Some versions of Linux use a similar method.

4.2.5 IT Security

It is important to consider the security of your own data to prevent:

- Theft of valuable data
- Breach of confidentiality agreements and privacy laws
- Premature release which can void intellectual property claims
- Release of data before it has been checked for accuracy and authenticity

Security of digital research data is a remit of Information Technology Security and ANU has extensive range of policies and information related to IT security[38].

The topic of IT Security is too large to cover here, but, at the very least, you should install up to data antivirus software on your computer. ANU staff and students are encouraged to select and install a suitable Anti-Virus product on their home computers[39], see Anti-Virus comparison[40]. ANU-managed machines should have Intercept X installed[41].

If you have sensitive data that is covered by privacy laws or confidentiality agreements it is best to store it on a computer that is not connected to any network. If this is not possible then you can also consider encrypting your data.

Encrypting data is not a trivial exercise. Your Local IT Support Staff (LITSS) may be able to help[42], some archiving services offer encryption e.g. CloudStor[43]. It is important to remember that if you encrypt your data you need to be able to decrypt it and you need to ensure that you store the access code or password safely and securely.

The final issue to consider is physical security. A computer that is not connected to a network is still vulnerable to someone removing the hard-drive and installing it in their own computer where they can bypass passwords and access restrictions. For highly sensitive data you can use an external hard-drive and store it in a locked safe overnight.

[38] *ANU IT Security*, services.anu.edu.au/information-technology/it-security
[39] *Anti-Virus Software*, https://services.anu.edu.au/information-technology/software-systems/sophos-anti-virus
[40] Anti-Virus comparison, https://au.pcmag.com/antivirus/44373/the-best-free-antivirus-protection
[41] Intercept X, https://services.anu.edu.au/information-technology/software-systems/sophos-anti-virus
[42] services.anu.edu.au/information-technology/help-support/contact-local-it-support-staff-litss
[43] https://support.aarnet.edu.au/hc/en-us/articles/235972928-Is-my-CloudStor-FileSender-data-secure-Can-I-encrypt-the-data-

4.3 Data Archiving and Sharing

Data archiving and sharing is considered an important part of academic research that encourages open inquiry into research results and conclusions, as well as promoting data reuse and repurposing.

ANU supports the FAIR data principles (Findable, Accessible, Interoperable, Reusable) drafted by the FORCE11 group in 2015. The principles are a useful framework for thinking about sharing data in a way that will enable maximum use and reuse[44]

Translating the FAIR principles in practice will be different for different disciplines, however the below guidelines set out the broad principles:

- **F**indable

This includes assigning a persistent identifier (like a DOI or Handle), having rich metadata to describe the data and making sure it is findable through disciplinary discovery portals (local and international).

- **A**ccessible

This may include making the data open using a standardised protocol. However the data does not necessarily have to be open. There are sometimes good reasons why data cannot be made open, for example privacy concerns, national security or commercial interests. If it is not open there should be clarity and transparency around the conditions governing access and reuse.

- **I**nteroperable

To be interoperable the data will need to use community agreed formats, language and vocabularies. The metadata will also need to use a community agreed standards and vocabularies, and contain links to related information using identifiers.

- **R**eusable

Reusable data should maintain its initial richness. For example, it should not be diminished for the purpose of explaining the findings in one particular publication. It needs a clear machine readable licence and provenance information on how the data was formed. It should also have discipline-specific data and metadata standards to give it rich contextual information that will allow for reuse.

4.3.1 Data Sharing Methods

Data dissemination is actively and permanently making your data accessible to others.

Archiving is the preferred option as most archives have the dual purpose of data preservation and dissemination. Archives usually have a search utility and are often indexed by the major web search engines, thus increasing the chances of other researchers finding, using and citing your datasets and publications. Archiving datasets also means the dataset owner can specify a wide range of access controls.

If your dataset is online, then including a citation of it including a permanent link in your publications will greatly increase its use, citation metrics and exposure.

[44] https://www.ands.org.au/working-with-data/fairdata

4.3.2 Copyright and Licensing

In general, ANU owns the copyright of material generated by staff in the course of their employment. The researcher, however, owns the copyright of their academic publications.

The owner is usually the creator, but some publisher, funding and research agreements require copyright to be handed over to another party.

Licenses grant permission for others to use the copyrighted data. A comprehensive guide[45] has been developed by the ARDC to assist researchers in working through this process. Open content licenses are also an easy way for researchers to license their data for others to use. Researchers can choose the most suitable license for their needs rather than develop a custom license themselves. The most notable open content licenses are:

- Creative Commons[46] - most popular open content licenses.
- Science Commons[47] - similar to Creative Commons but tailored for scientific data and publications.
- GNU Free Documentation License[48] - used by Wikipedia.
- GNU General Publication Licence v3 – the most common licence for open source software

4.3.3 File Formats and Standards

Before creating research data you need to consider what formats and standards you should use as it is sometimes difficult to convert between file formats[49]. Using an inappropriate or bespoke file format will also make your life more difficult in the long run.

Where possible, it is best to use preservation-friendly formats as they are more likely to be readable in the future and are easier to share with others[19]. It is usually safe to use a proprietary format if it is very widespread, as free programs will most likely exist to read these formats.

Some examples of preservation-friendly formats are:

- PDF – document format
- Open Document Format (ODF) – used by LibreOffice Writer among others
- PNG, TIFF, JPEG – Image formats

Your LITSS (see section 5.1) and ANU Open Research staff (see section 5.4) can give you advice on what file formats to use. For archiving, PDF (Portable Document Format) for documents and TIFF (Tagged Image File Format) for images are recommended. Note that most document and image formats can be converted to PDF and TIFF, respectively, but there may be some loss in quality.

[45] *Research Data Rights Management Guide* https://ands.org.au/guides/research-data-rights-management
[46] *Creative Commons*, creativecommons.org
[47] *Science Commons*, sciencecommons.org
[48] *GNU Free Documentation License*, www.gnu.org/copyleft/fdl.html
[49] *File Format*, en.wikipedia.org/wiki/File_format

4.3.4 Access Restrictions[50]

When data is in its final state and ready for dissemination or archiving, you should create its metadata (see section 4.3.6) and define the *Access Restrictions* on each item of data:

- Unrestricted – Anyone can download.
- Registered – Users must give their name and affiliation so the data owner can track who is using their data.
- Requested – Users must submit a request outlining how they will use the data.
- Closed – No access (i.e., confidential data).

4.3.5 Archiving

Archiving of your final research data is encouraged and in some cases is required (see section 3.3) by your employer, funders and / or journal editors.

Archiving your data also ensures it will not be lost, forgotten, or become unusable due to being stored in legacy file formats or storage media (see section 3.1.6). Archiving also takes care of preservation, dissemination, access control and security.

Archives or institutional repositories, generally only accept final state data. The objective of the archive is to preserve the data and – if the data owner allows it – make the data available for future research. The owner of the data can specify a range of access restrictions (such as those described in section 4.3.4); although each archive may use different terminology. It is also possible to place an embargo on the dataset such that the data cannot be accessed until after a specified date. This is often done to give the data creators time to publish their results before making their data public. Sensitive datasets can also be archived after the data has been de-identified[51].

An archive provides long-term storage of data and therefore prefers file formats that are unlikely to become obsolete. Most file formats can be converted to a suitable archiving format but some loss in quality (such as images or audio) or distortion (such as converting PowerPoint to PDF) may occur. Most archives are able to perform the conversion but it is best if you do the conversion to ensure that you are happy with the result.

The time and costs associated with archiving are often underestimated. Each item of data deposited will need to have metadata written for it, which can be very time consuming if your data consists of several hundred images which were taken some years ago. It is therefore best to write the metadata at the same time the data is created and to archive that metadata and data continuously, rather than leaving it until the end of the project. It is recommended that you include the costs of archiving in your grant application.

The University's archiving/repository services can be found in sections 5.5, 5.6, 5.7.

[50] *Updated Access Restrictions:* http://services.ands.org.au/documentation/rifcs/1.6.1/vocabs/vocabularies.html#Access_Rights_Type
[51] https://www.ands.org.au/working-with-data/sensitive-data/de-identifying-data

4.3.6 Metadata

Metadata is often described as "data about data"[52], and helps people to discover, understand and re-use data. It can be used to describe physical items as well as digital items (files, documents, images, datasets, etc.). It includes descriptive metadata, such as title, time, author, keywords, relations to other data objects; administrative metadata such as ownership and use permissions; and history of changes to the data, versions etc.

Good metadata is also vital for F.A.I.R research data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (See section 4.3)

Metadata are usually found in a separate file with several text fields that describe the attributes of another piece of data, such as an experimental dataset, image, or video. Sometimes, metadata is written into source data file itself, such as in jpeg files from digital cameras.

Metadata, as a minimum, should contains at least the following information about the data:

- Filename
- File size (kilobytes, megabytes, etc.)
- File type (LaTeX document, JPEG image, etc.)
- Date of creation
- Author or copyright holder
- Brief description
- Keywords

You can think of the metadata, in relation to the data it describes, as being analogous to the abstract or keywords of a paper – it is there to help people find your data and quickly decide if it is what they need. If you want people to find and reuse your data (and therefore help you by citing your work), then it is worth your while making good metadata in order to 'sell' your data.

Metadata is critical for archiving; most archives will not accept data that does not have adequate metadata. Creating metadata at the end of a project is also extremely difficult, as you may have to go through several hundred photographs or audio files. Metadata should therefore be made as the data is created.

Metadata schemas/standards include what are considered essential elements to describe any data. A list of commonly used disciplinary standards can be found in the ANU Data Management LibGuide[53].

[52] *Metadata*, en.wikipedia.org/wiki/Metadata
[53] ql.anu.edu.au/data

5 ANU Data Management Support Services

This section discusses the various data management services available at the ANU.

ANU IT Services and your Local IT Support Staff (LITSS) provide the majority of the data management services. They provide the day-to-day data management services like backups and shared drives.

Data archiving is more specialized and is available through: *ANU Data Commons*, NCI[54], and the Australian Data Archive (ADA)[55].

The IT Services website[56] contains extensive information on ANU information services.

5.1 Local IT Support Staff (LITSS)

The majority of your data management needs will be provided by your Local IT Support Staff, or LITSS[57].

The services that they provide are usually determined by the Head of College, so may vary between colleges and even schools. Most colleges will provide a file server and web hosting and employ staff as LITSS.

A file server is a computer that stores data and makes it accessible to your computer via a network connection. Normally it will appear as a mounted drive and behave the same as any other directory on your computer. It is recommended that you keep all your files on your mounted drive as the file server is automatically backed up at regular intervals and is less likely to fail. If you choose to keep your work on the hard-drive of your desktop/laptop, then you will need to perform the backups yourself.

If your data is sensitive then you should either store the data on your computer's hard-drive or encrypt the data before placing it on the mounted drive. Your LITSS may be able to set up your account to automatically encrypt a directory within your mounted drive. If you forget the password it will be impossible to recover the data.

Most colleges have web servers for hosting faculty and department websites. They may also allow academic staff to use these webservers for personal or research group websites. This is an easy way to make your publications and datasets available online, but it is recommended to store the data in an archive, such as ANU Data Commons, and link to it from your website.

Your LITSS may also provide a mounted drive for data storage and collaborative work. Normally a fileserver provides a directory that only you can access, but your LITSS may be able to set up a mounted drive that several people can contribute to. Using a mounted drive creates problems such as keeping track of changes and simultaneous edits. If the data is being edited often or there are a large number of people using the data, then it is best to use version control software.

Finally, your LITSS are your first point of contact if you need software for data management, such as EndNote, Office software, and conversion tools[58].

[54] http://nci.org.au/
[55] http://www.ada.anu.edu.au/
[56] *ITS*, https://services.anu.edu.au/information-technology
[57] *LITSS Information*, services.anu.edu.au/business-units/information-technology-services/local-it-support-staff-litss
[58] *Software Information*, https://services.anu.edu.au/information-technology/software-systems

5.2 Information Technology Services (ITS)

Information Technology Services at ANU is responsible for the undergraduate computer laboratories, the major mail servers, and more. In regards to data management, they provide your Homedrive fileserver, OneDrive, *Wattle* and other webservers.

5.2.1 Homedrive

Homedrive is a personal network storage space[59] where all ANU research staff and students can store university related documents and files. Research staff and students are allocated 4.5GB of space. Home drives are for individual use only. All students and staff have a Homedrive. You can access your data from any campus computer. You can also access your home drive from off-campus[60].

5.2.2 SharePoint

Office 365[61] Collaboration Toolset includes SharePoint, Teams, Groups and One Drive. They are all used to manage documents, but have different purposes at ANU.

SharePoint[62] is used to co-author documents, manage versions, apply workflows, and discover and share information. Teams is focused on real-time communication, including with people outside ANU. Groups provides mailing list capability, and OneDrive stores your personal documents.

5.2.3 OneDrive

OneDrive[63] is a free online personal storage available to all ANU staff, students and affiliates via your Office 365 account. OneDrive allows you to upload, sync, and share files, and then access them from a web browser or local device. Up to 25GB of storage is available using this service.

5.3 Statistical Support Network

The Statistical Support Network[64] is a research collaborative unit that specialises in statistical data in academic research across the University. They keep abreast of fast-evolving research methodologies and statistical software in a data-driven research environment, across a broad range of disciplines. They can provide advice on data management, exploratory data analysis, statistical modelling and data presentation.

[59] https://myfiles.anu.edu.au
[60] https://myfiles.anu.edu.au
[61] https://services.anu.edu.au/information-technology/software-systems/microsoft-office-365
[62] https://anu365.sharepoint.com/teams/collab/help
[63] https://services.anu.edu.au/information-technology/software-systems/onedrive
[64] https://services.anu.edu.au/business-units/dean-higher-degree-research/statistical-support-network

5.4 ANU Library Digital Literacy Training

ANU Library Digital Literacy Training[65] provide instructor-led training courses and online training courses[66]. Several training courses offered can assist in data management:

- NVivo (manage and analyse qualitative data)
- SPSS (statistical analysis)
- Working with data in Excel
- Reference management software

5.5 Open Research

ANU Open Research is the designated institutional repository of the University[67].

It provides long-term storage, preservation and dissemination of ANU research output. Open Research has its own search tools (accessible via web interface) and is indexed by major search engines such as Google, thus increasing the likelihood of your research and associated data being found and re-used.

Open Research holds, for example, ANU Digital Theses, journal articles, creative works, podcasts, conference papers and more. Contact *Open Research* staff at repository.admin@anu.edu.au if you would like to add your research to the online collection.

Increasingly, funding agencies are requiring research results be made openly accessible through institutional repositories such as ANU Open Research. We can add PURL IDs for ARC and NHMRC grants to assist discovery[68] as well as making the Author Accepted Manuscript (AAM) available to meet funder compliance. For further information, contact Open Research staff on x59729.

5.6 ANU Data Commons

ANU Data Commons is a specialist data archiving service, providing long term storage and preservation of research data. Data can be uploaded in a number of ways and specialist upload facilities are available to both upload large datasets, large numbers of files, or data generated by research instruments.

Data stored within ANU Data Commons can be either public, private or available by request, i.e. people can search for and find a dataset's metadata but have to ask you for permission to access the data itself.

ANU Data Commons datacommons.anu.edu.au/DataCommons can create permanent Digital Object Identifiers (DOI[69]) as an aid to data citation and also publishes your metadata record to Research Data Australia to aid discovery. For further information, contact Open Research staff on x59729 or repository.admin@anu.edu.au

[65] *ANU LIbrary*, anulib.anu.edu.au/research-learn
[66] *How to guides*, anulib.anu.edu.au/find-access/how-guides
[67] *ANU Open Research@ANU*, http://openresearch.anu.edu.au/
[68] https://www.ands.org.au/working-with-data/citation-and-identifiers
[69] https://www.ands.org.au/guides/doi

5.7 National Computational Infrastructure (NCI)

The National Computational Infrastructure (NCI[70]) is Australia's national research computing facility, providing world-class services to Australian researchers, industry and government.

NCI is home to the Southern Hemisphere's most highly-integrated supercomputer and filesystems, Australia's highest performance research cloud, and one of the nation's largest data catalogues—all supported by an expert team.

The NCI provides:

- Compute facilities for intensive processing or analysis
- Mass Data Storage Facilities (MDSS) for projects with large data requirements
- Dataset hosting for very large datasets of national significance
- Consultancy and training to help people solve complex problems and to help them use the facilities

Use of NCI is usually free to ANU researchers upon application for an account[71]. Resources are allocated through the National Computational Merit Allocation Scheme (NCMAS)[72], or The Australian National University Allocation Scheme. Applications are submitted online.

5.8 Australian Data Archive (ADA)

The Australian Data Archive (ADA)[73] — now incorporating the Australian Social Science Data Archive — is a consortium of leading national Australian universities, managed by The Australian National University (ANU).

ADA employs a team of professional data archivists, advised by a panel of leading social scientists, provides both stewardship and outreach services to the Australian community. The archive:

- acquires, documents, preserves and disseminates data online to a broad range of social science researchers in the university, government, and other sectors
- provides the only comprehensive social science data collection in Australia, with a catalogue of over 2000 data sets
- holds data from Australian surveys, opinion polls and censuses and includes data from other countries within the Asia Pacific region
- provides specialist services within specific subject areas, including Indigenous studies, electoral behaviour, criminology and some humanities disciplines, and within specific data types, including quantitative, qualitative, time series and panel data, and historical statistics
- locates and manages access to overseas social science data sets required by Australian based researchers
- plays an important role on behalf of the Australian Research Council (ARC) through the management and dissemination of ARC funded data collections arising from Discovery and Linkage grants

[70] NCI (National Computational Infrastructure), nci.org.au
[71] my.nci.org.au/mancini
[72] http://nci.org.au/access/getting-access-to-the-national-facility/national-merit-allocation-scheme/
[73] https://www.ada.edu.au/

5.9 The Australian Research Data Commons (ARDC)

The Australian Research Data Commons (ARDC), formerly known as Australian National Data Service (ANDS), is a transformational initiative that enables Australian researchers and the eResearch community access to nationally significant, leading edge data intensive eInfrastructure, platforms, skills and collections of high-quality data.[74]

- Australian Data collections can be searched using the *Research Data Australia (RDA)* service[75].
- ANU students and staff can publish their data into RDA via ANU Data Commons (see section 5.5).

[74] *The ARDC (Australian Researchl Data Commons)*, https://ardc.edu.au/
[75] *RDA (Research Data Australia)*, https://researchdata.ands.org.au/

6 Writing a Data Management Plan (DMP) via DMPTool

The Australian National University is a participating institution of the DMPTool service[76], provided by the California Curation Centre of the California Digital Library, for the creation and management of data management plans. ANU recommends the use of the default DMPTool template.

To complete a plan, you will need to consider the following:

- Data Collection
  - What data will you collect or create?
  - How will the data be collected or created?
- Documentation and Metadata
  - What documentation and metadata will accompany the data?
- Ethics and Legal Compliance
  - How will you manage any ethical issues?
  - How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?
- Storage and Backup
  - How will the data be stored and backed up during the research?
  - How will you manage access and security?
- Selection and Preservation
  - Which data are of long-term value and should be retained, shared, and/or preserved?
  - What is the long-term preservation plan for the dataset?
- Data Sharing
  - How will you share the data?
  - Are any restrictions on data sharing required?
- Responsibilities and Resources
  - Who will be responsible for data management?
  - What resources will you require to deliver your plan?

[76] https://dmptool.org/

7 Glossary of Terms

AARNet Provides Internet services to the Australian education and research communities and their research partners

ADA The Australian Data Archive provides a national service for the collection and preservation of digital research data and to make these data available for secondary analysis by academic researchers and other users

AIATSIS Australian Institute of Aboriginal and Torres Strait Islander Studies

ANDS The Australian National Data Service now ARDC

APAC Australian Partnership for Advanced Computing

ARC Australian Research Council

ARDC The Australian Research Data Commons, formerly ANDS

Archive Digital Archive

ASEDA Aboriginal Studies Electronic Data Archive

ASSDA Australian Social Sciences Data Archive

Backup A copy of data kept for recovery in case of accidental deletion or loss of data

Bazaar A distributed version control system for collaborative tracking of project history

BibTeX A software tool and a file format for generating bibliographies in LaTeX

Binary file A file that cannot be read without appropriate software, contrast with a text file which can be read with any text editor

BlueNet Australian data archive for the marine sciences

CAI Computer Assisted Interview - a computer program which helps direct an interview based n the responses given

CD Compact Disc – the oldest form of optical disc, can store 700Mb

CiteSeer Online scientific literature digital library

Client A program used to interact with a server. For example, an FTP client is needed to download and upload files to an FTP server.

Cloud Storage is a model of data storage in which the digital data is stored in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company[77]

CloudStor AARNet's file sync, share and storage service

Creative Commons An organization that provides a generic license framework for freely distributed data

Data Digital research data

Dataset A collection of related data such as tables of numeric data or a group of related images

Data administration Anything done to protect data or enhance the quality of data

Data management Anything outside of using the data, such as organizing, protecting, sharing and archiving data

Data organization Tools and techniques for working more efficiently with data

Data sharing Actively making research data available for use by other researchers

Desktop The most common type of PC (personal computer)

DocBook An XML file format for documents

DVD Digital Versatile Disc – an optical disc used for storing data, can store 4.7Gb or roughly 7 CDs

eNotebooks a cloud-based electronic filing system for managing files and datasets, and designed to protect data and comply with regulatory funding requirements

[77] https://en.wikipedia.org/wiki/Cloud_storage

Encryption Making data unreadable to anyone without the correct password or encryption key

EndNote Software tool for managing bibliography databases and generating bibliographies in MS Word documents

Excel Microsoft Excel, software tool for working with spreadsheets

External hard-drive A device for storing data that can be easily connected to computers via a USB cable

Fileserver A centrally-managed computer which stores data and allows authorized users to access their data. Fileservers are usually backed up every night and it is recommended to save all data on a fileserver rather than a computer's hard disk

FTP File Transfer Protocol. Simple method for transferring files over a network (such as the internet). An FTP server accepts connections from a software tool called an FTP client and allows data to be downloaded and uploaded.

Homedrive A personal network storage space where students and staff can store University related documents and files. Home drives are for individual use only. All students and staff have a Home drive.

IMAP Internet Message Access Protocol. Method of checking email from several different computers.

LaTeX A markup language for writing scientific and technical publications. Widely used in academia.

LITSS Local IT Support Services. ANU Departmental IT Staff that maintain staff computers and IT infrastructure.

Lossless compression Method of compressing a file without losing quality, such as zip, gif, png

Lossy compression Method of compressing a file that results in a loss of quality, such as jpeg, mp3

Mac Apple Macintosh computer

Markup language A plain text file that can be processed into graphical format, such as LaTeX, html, docbook.

Metadata A data file that describes attributes of a data item, such as title, creators, size, date, format.

Mounted drive Remote data that appears as a directory on a computer

MS Microsoft

NCI The National Computational Infrastructure

NESSTAR Software tool for statistical analysis of datasets

Network share A device that can be accessed over a network. Similar to mounted drive but also includes printers.

OneDrive OneDrive is a free online personal storage available to all ANU staff, students and affiliates via your Office 365 account.

Open Access Providing free, electronic copies of academic articles online

Open Content Data that is free for anyone to download and use

OpenOffice A free software tool that is mostly compatible with Microsoft Office

Optical disc A data storage medium such as a CD, DVD, and BluRay

PDF Portable Document Format – common format for distributing documents

Server See also: Client

SharePoint A collaborative solution, part of Microsoft 365 Collaboration Toolset.

SPSS The IBM Statistical Package for the Social Sciences (SPSS) is a statistical application which can perform a variety of functions, including statistical analysis and graphical presentation of data

Text file See also: Binary file. A file that can be read with any text editor.

Ubuntu An open-source Linux-based operating system for computers

Unix An operating system used by many servers and high performance computers

USB Universal serial bus – method for easily connecting devices (external hard-drives, digital cameras, printers, mice, keyboards) to a computer

USB key A small device for storing data. Data is transferred to the device via USB and it can then be removed and connected to another computer.

Version Control Software Software tool for methodically tracking changes to a file or files

Web Application A software tool that can be used with any internet browser

Web Browser A software tool for accessing web pages

Webserver A server that stores web pages and other data that people connect to with a web browser

Wiki A website that can be modified by anyone, such as Wikipedia. Some wikis only allow modification by registered users.

Windows Microsoft Windows – the operating system used by most computers

Word Microsoft Word – a software tool for writing documents (word processor)

XML Extensible Markup Language – general purpose markup language that can be used for documents, metadata, databases

Zotero is a free, easy-to-use tool to help you collect, organise, cite and share research. Can be used with both LaTeX and Word/OpenOffice. Supports BibTex and EndNote databases.

8 References and Links

ARDC (Australian Research Data Commons). https://ardc.edu.au/
ANU Data Commons https://datacommons.anu.edu.au/DataCommons/
ANU IT Security. https://services.anu.edu.au/campus-environment/safety-security
ANU Statistical Consulting Unit. scu.anu.edu.au
Bazaar. bazaar.canonical.com/en/
Computer Assisted Interviewing. *Wikipedia*, en.wikipedia.org/wiki/Computer-assisted_telephone_interviewing
Creative Commons. creativecommons.org
Data Validation. *Wikipedia*, en.wikipedia.org/wiki/Data_validation
Digital Literacy Training. *ANU Library*, anulib.anu.edu.au/training
Dropbox. www.dropbox.com
eNotebooks https://enotebooks.anu.edu.au
eNoteBooks Training https://www.labarchives.com/training-webinars/
Endnote. www.endnote.com
Ewa Deelman et al. "Workflows and E-Science: An Overview of Workflow System Features and Capabilities." *Future Generation Computer Systems* 25, no. 5 (2008): 528–40.
File Format. *Wikipedia*, en.wikipedia.org/wiki/File_format
Firefox Web Browser. www.mozilla.com
FTP. *Wikipedia*, en.wikipedia.org/wiki/File_Transfer_Protocol
FTP Explorer. www.ftpx.com/
GitHub. https://en.wikipedia.org/wiki/GitHub
GNU Free Documentation License. www.gnu.org/copyleft/fdl.html
Harnad, Stevan, and Tim Brody. "Comparing the Impact of Open Access (OA) Vs. Non-OA Articles in the Same Journals." *D-Lib Magazine* 4, no. 6 (2004).
ICPSR. *Guide to Social Science Data Preparation and Archiving (4th Edition)*. Inter University Consortium for Political and Social Research (ICPSR), 2009. www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf
Homedrive, Remote Access. https://myfiles.anu.edu.au/
Information Services website, information.anu.edu.au
JabRef. jabref.sf.net
LITSS. *ANU*, services.anu.edu.au/business-units/information-technology-services/local-it-support-staff-litss
Metadata. *Wikipedia*, en.wikipedia.org/wiki/Metadta
NCI (National Computational Infrastructure). nci.org.au/
NCI National Facility. *ANU*, nf.nci.org.au/
Open Research@ANU. https://openresearch.anu.edu.au/
OneDrive.https://services.anu.edu.au/information-technology/software-systems/onedrive
Privacy Act (1983). *Australian Government*, https://www.oaic.gov.au/privacy-law/privacy-act/
Publish My Data. *Australian National Data Service*, https://www.ands.org.au/working-with-data/publishing-and-reusing-data/publishing
RDA (Research Data Australia). https://researchdata.ands.org.au/
Revision Control. *Wikipedia*, en.wikipedia.org/wiki/Revision_control
rsync. *Wikipedia*, en.wikipedia.org/wiki/Rsync
Sakai. sakaiproject.org/
Science Commons. sciencecommons.org
SharePoint. https://anu365.sharepoint.com/teams/collab/help
Software Information. *ANU*, information.anu.edu.au/daisy/infoservices/8/29.html
Anti-Virus Software. https://services.anu.edu.au/information-technology/software-systems/sophos-anti-virus
SSH File Transfer Protocol. *Wikipedia*, en.wikipedia.org/wiki/SSH_File_Transfer_Protocol
TortoiseSVN. tortoisesvn.net/
Vislab. *ANU Supercompting Facility*, anusf.anu.edu.au/Vizlab/
WinSCP. *Wikipedia*, en.wikipedia.org/wiki/WinSCP
Zotero. www.zotero.org

Other resources

Training notes

To access training notes, visit the Research & learn webpage anulib.anu.edu.au/research-learn and select the skill area followed by the relevant course. You can register for a workshop and find other information.

Research & learn how-to guides

Explore and learn with the ANU Library's how to guides (anulib.anu.edu.au/howto). Topics covered are:

- Citations & abstracts
- E-books
- EndNote
- Evaluating Sources
- Finding books and more
- Finding journal articles and more
- Finding theses
- Increasing your research impact
- ORCID iD (Open Researcher and Contributor ID)
- Research Data Management
- Text and Data Mining
- Topic analysis

Subject guides

Find subject-specific guides (anulib.anu.edu.au/subjectguides) and resources on broad range of disciplines. Such as:

- Asia Pacific, Southeast Asia and East Asian studies
- Business, economics, art, music and military studies
- Criminal, human rights and taxation law
- History, indigenous studies, linguistics and philosophy
- Biological, environment, physical & mathematical sciences, engineering & computer science, health & medicine

Navigating the sea of scholarly communication

An open access course designed to build the capabilities researchers need to navigate the scholarly communications and publishing world. Topics covered include finding a best-fit publisher, predatory publishing, data citations, bibliometrics, open access, and online research identity. Five self-paced modules, delivered by international and local experts/librarians (anulib.anu.edu.au/publishing).

Training Calendar

A range of workshops are offered to help with your academic research and studies (anulib.anu.edu.au/training-register).