University of
Western Sydney

eResearch

MANAGING AND WORKING WITH RESEARCH DATA: A GUIDE FOR UWS HIGHER DEGREE RESEARCH STUDENTS

ACKNOWLEDGEMENTS

Much of the content in this module has been derived from the following document which is licensed under a **Creative Commons Attribution 4.0 International Licence**:

Griffith University. (2014). *Best practice guidelines for researchers: Managing research data and primary materials* (version 1.4). Retrieved from **http://www.griffith.edu.au/__data/assets/pdf_file/0009/528993/Best_Practice_Guidelines.pdf**

TABLE OF CONTENTS

INTRODUCTION

As a researcher, it is essential that you take an active role in data management during the research life cycle and beyond. Active planning involves making decisions about how you will collect, organise, manage, store, back-up, preserve and share your data. Sound data management planning and implementation should begin early in a research project and will enable you to:

- Plan for end-to-end data management needs, including computational and storage requirements.
- Increase research impact by ensuring data is both preserved and citable.
- Ensure long-term access to data through well-described and retrievable data sets.
- Retain the potential to make datasets available for reuse and/or future research projects and collaborations.

What is research data?

Broadly speaking, research data is the data that is produced in the course of your research. This data may be physical or electronic and can be raw, cleansed or processed. The **UWS Research Code of Practice** states that research data is '…*information held in any form including paper, electronic, visual (x-rays, CT scans, videos, photos and MRIs), audio records or personnel records of any kind (such as student or job records, salary payment details or health and medical details*.'

The list below further details the forms that data can take:

- Documents, spreadsheets and presentations
- Laboratory notebooks
- Field notebooks, diaries
- Questionnaires, surveys, transcripts
- Audio and video tapes
- Photographs and films
- Test responses or results
- Slides, artefacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Databases
- Models, algorithms, scripts
- Contents of an application (input, output, schemas)
- Methodologies and workflows.

    University of Sydney. (2012). *What is research data?* Retrieved from **http://sydney.edu.au/research_support/data/what-is-research-data.shtml**

Data management planning is an important aspect of the responsible conduct of research (see the **UWS Research Code of Practice**). The datasets that you create during your research projects are valuable and may need to be preserved over long periods of time. Time spent planning and documenting data management requirements at the beginning of a research project is not only good practice but can provide benefits to you as the researcher, as well as to the University and the broader community.

As part of your research data management planning and implementation, it is important to comply with relevant UWS policies. Policies that may apply to your research work include:

- UWS Research Data Management Policy (currently in draft awaiting approval)
- UWS Open Access Policy (currently in draft awaiting approval)

The following sections will provide information about data management considerations and will highlight some of the resources available to support you during your research candidature.

AT THE BEGINNING OF YOUR PROJECT

During your early candidature it is wise to consider your options and obligations in relation to the data you plan to collect. Think about what types of data (qualitative, quantitative or mixed methods; primary or secondary) will best contribute to your research findings and to the impact of your project. Consider the expectations of researchers in your discipline and from other disciplines and how these might affect how you manage your data (including sharing, if possible).

When formulating your project idea, consider:

- The audiences for your research and how they could make use of the data you will be collecting – is your work of interest to policy makers, not-for-profit agencies, the commercial sector or the general public, as well as to other researchers?
- The data management and data sharing requirements of journals you might publish in.
- The availability of data journals e.g. **Scientific Data** for your discipline for publishing data outputs.
- How you could use data to communicate your results more effectively - data in raw and visualised forms adds interest to your publications and conference presentations.

Data management planning from the outset of a research project assists researchers plan for and articulate the following:

- Data to be produced
- Data documentation and metadata (data about the data)
- Data storage and security needs
- Ethics, copyright and Intellectual Property
- Access, sharing and reuse of data
- Data retention and disposal
- Preservation and archiving of data.

It is important to remember that data management plans are living documents and can change as your data requirements change. This enables you to complete your plan at the beginning of your project knowing that you will be able to adapt it over time, thus ensuring timely consideration of all the aspects of effective end-to-end data management.

To assist you with data management planning for your candidature, UWS has produced a **data management checklist** and a **template** for producing a data management plan. It is suggested that you complete the data management checklist prior to your confirmation of candidature as it forms the basis of a comprehensive data management plan after confirmation has occurred.

Some of the questions your checklist and data management plan will address are:

- What rights, including copyright, will subsist in the data produced by the project?
- Who will be the rights holder/s for the data?
- What terms and conditions should be applied to the data for re-use? This could include your own plans to reuse the data in future.
- How long should data be retained?
- What kind of working storage will be required, considering file types and sizes, software needs, access and collaboration?
- Whether an institutional repository or subject repository can disseminate your data. Your thesis will be published in the UWS Research Direct publications repository, but you should also think about describing and depositing your data in the UWS Research Data Catalogue. This will be linked to your thesis, thus increasing the discoverability of both outputs and the possibility of citation by other scholars.

If you have questions about this process, support is available from the UWS Library's **Research Data Coordinator** or your **School/Discipline Librarian**. For complex technical questions, you may be referred to the **UWS eResearch team**.

If your research requires an ethics application, this application will be informed by the decisions you make through your data management plan. Be explicit in your ethics application about any plans you have to make data available to other researchers or more broadly. Describe your strategies for protecting privacy and confidentiality, and ensuring informed consent.

You will also need to be explicit in your research consent forms about any plans to make data available, who will be able to access the data, and how the data would be accessed and potentially re-used.

Data does not have to be openly accessible to be shared for the benefit of future researchers and other interested groups. In many cases, mediated or restricted access could be appropriate.

Watch this five minute video on **The what, why and how of data management planning**.

DURING YOUR PROJECT

Once your project is underway, take advantage of all the resources available to ensure that your research process is efficient and your results are reproducible.

Computing, research tools, and storage requirements

UWS Researchers have access to a wide-range of computing platforms. These include:

- **Desktop and Laptop Computing:** These resources can be obtained through your School or Institute.
- **Virtual Servers:** The next step up from desktop computing is to use a virtual desktop or server. If you require Microsoft Windows, there are options available through UWS. If you are comfortable with a commandline Linux environment, the self-serve NeCTAR Research Cloud is a good place to start.
- **High Performance Computing (HPC):** For intensive computation, you'll need access to an HPC cluster or super-computing environment.

You can request help with virtual servers or HPC through the **UWS MyIT** service portal.

The UWS eResearch team can help connect researchers with appropriate analytical or visualisation tools, software instances or collaboration environments. A useful external site worth browsing is **The DiRT Directory** which brings together information about digital research tools for scholarly use.

During your project you will need access to working storage. This is where you put the data and files you're working with/on. For working storage, you can make use of UWS's Research Shared Drive. Alternatively AARNET, a federally funded academic and research company, offers working storage in the form of Cloudstor+ (aka Cloudstorplus). Cloudstor+ is just like Dropbox in many ways, but hosted in Australia, and with higher quotas.

More information and assistance, including training opportunities, is available via the **UWS eResearch website**.

Metadata standards and controlled vocabularies

Whilst undertaking your research project you should create and maintain sufficient documentation or **metadata** (i.e. structured information about the data) to enable research data to be identified, discovered, associated with its owners and creators, linked to other related data or publications, contextualised in time and space, and to have the quality of the data assessed and research results validated.

Metadata often uses a controlled vocabulary which specifies the preferred use of terms. By using an appropriate controlled vocabulary, you make it easier for others to discover your work through related works. Your **School/ Discipline Librarian** can assist in this area.

File naming conventions and identifiers

Digital file names can be important for identifying and finding digital files. Ideally, you should develop unique, persistent and consistently applied file naming conventions as part of your data management plan, and agree on these with colleagues and collaborators before data is created.

An identifier is a unique reference number or name for a data object. The emerging identifier standard for publicly available datasets is the Digital Object Identifiers (DOIs). Although DOIs have been traditionally used for electronically published journal articles, they can now be assigned to datasets so that these can be cited in reference lists. If appropriate, UWS Library can assign a DOI to a collection that you make available through the UWS Research Data Repository.

AT THE END OF YOUR PROJECT

It is very important that all research publications, including theses, are supported by datasets that can be used to prove the integrity of the research. Whatever the field and level of privacy needed for the data, managing your data for the long term is an essential part of the research process.

Archiving research data

By depositing data in a repository (or archive), you ensure that your data can be accessed and cited in the long term. Data should preferably be archived in formats which are non-proprietary to ensure access is not compromised by format or software obsolescence. Further information is available through the Australian National Data Service guide to file formats.

The UWS Research Data Catalogue is managed by UWS Library and is suitable for a wide range of data. You can submit research data and make it openly accessible, or choose to publish only a description of your data with mediated access.

To deposit research data in the Research UWS Data Catalogue please complete the submission form.

Data and metadata that you choose to share publicly can be cited by others, and will be discoverable via the Research Data Australia portal and through web search engines such as Google. This exposes your research to new audiences and potential collaborators.

Licensing data for reuse

There are many options for licensing data for reuse by potential collaborators, or even your future self. An open licence lets you reserve some rights as the owner of the material, but it allows potential users more rights than they would have under copyright legislation. UWS Library staff can provide guidance on this should you choose to deposit your data.
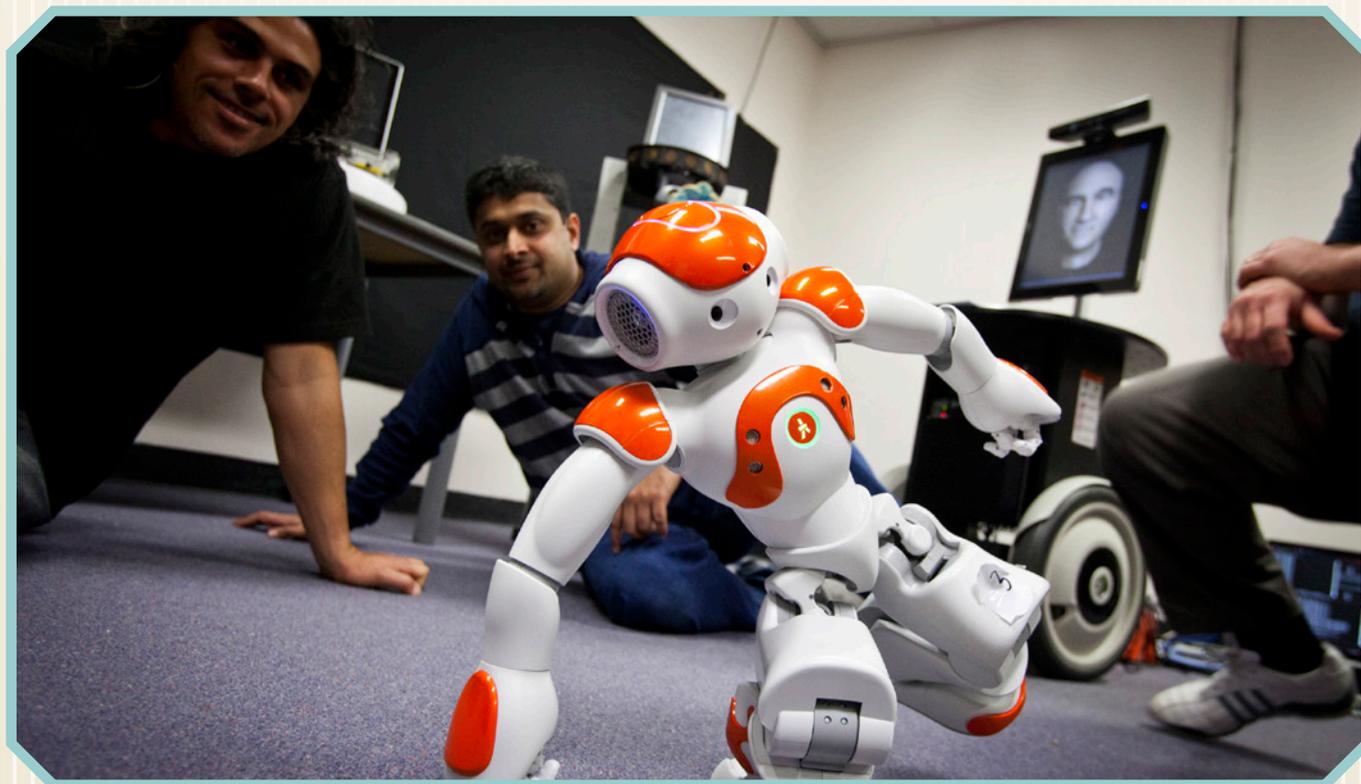
CONCLUSION

Research data is increasingly recognised as a valid and valuable research output, and should be managed and preserved accordingly. This module provides an introduction to the topic; however you may wish to explore some aspects further. Some useful resources are:

- **Research Data MANTRA** [online course] by EDINA and Data Library, University of Edinburgh.
- **Australian National Data Service** (ANDS) website.
- **Digital Curation Centre** (DCC) UK website.