# Working Title:

**Data Management: Make sense of the complexity around data storage, description, re-use and archiving.**

The research data landscape is evolving.  International publishers and funders are now mandating for best practice in data planning, description, storage, and sharing. However data sensitivities still need to be understood and managed.  This discussion will cover general best-practice principles of management, storage and sharing of research data. Using the DMPt as a guide, it will include practical tips for improving data management practices that can be implemented immediately regardless of the type of data. By attending students should feel better prepared to respond to university, employer, funder and/or publisher data requirements.

# Learning Objectives:

Students will be able to:

1. Think critically about best practice in the management, storage and sharing of research data, relating it to their discipline and research practices.
2. Share and discuss personal data management experiences.
3. Examine their current practices within conversations around the Otago DMPt.
4. Understand and use the University of Otago DMP tool.

# Resources Required:

PowerPoint Presentation

Internet and projector connected Laptop (if a tutor machine is not available) – go live

Research Data Management: Practical Tips and Tricks for All Researchers – handout

DMPt handout

# Lesson Outline:

Below is an order of activities assuming the discussion is structured. However, be prepared to deviate to address specific interests or concerns that arise.

|  |  |
| --- | --- |
| Activity | Time Guideline |
| **Introduction:**  Greet students and introduce presenters | 2m |
| **First Activity:**  What is the Research Data Management Plan? Why do you need one? Who has one? | 5m |
| **Second Activity: Introduce the Otago DMPt**  -Paper version – they are welcome to fill as we go  -Online version in two parts  -No mandatory fields – living document you complete gradually over time  -Captures bibliographic metadata more so than Subject Specific metadata but there is scope for Subject-specific  -Digital version allows you to add other researchers and has information hover overs to provide further explanations  -Influences: DCC and DMPOnline | 8m |
| Third Activity: Go through the DMPt and consider what information it is capturing and why:  \*\*Some questions we can feed to the attendees if they don’t themselves  **Project description:**  Can be used later (copy and paste) to a repository  \*\*Consent - If someone wants to access the data in the future who should be contacted?  **Dataset Information:**  \*\*Owner/Creator - Researchers who generate data as part of a wider project – do they own the data they collect? Will they be able to get access to the data if they move on?  Creator ID – I use ORCiD \*\*who has an ORCiD? [will come back to ORCiD]  Description – again useful for copy and pasting.  Start and end dates of data collection are not always the same as the project  \*\*Do you systematically record the tools used to generate data? – how detailed do you go? - Implications for reproducibility/transparency. Also note movement towards providing instrument PIDs and encouraging citation culture.  \*\*How proprietary is/are the formats your data are in? Could you access them in 10 years’ time? Format – start considering long term preservation. Plus tidy data principles – Have a copy of the raw data that you never touch. Consider a preservation copy that will be in a non-proprietary format (or at least in a format that is familiar to your research community). Issue of format shifting e.g. what if your external hard drive doesn’t plug into a modern computer in 10 years’ time?  Keywords – again useful for copy and pasting later.  Citations – those publications that cite the data – why this is a living document  Host department – responsible for managing and storing the data  \*\*How easy would you be able to find that dataset a few years on? Location may be physical or digital – Use a stable/persistent URL so a repository with DOI is better than your personal website etc.  Number of files – tidy data principles – keep like with like and be consistent file naming  Space – At start this might be an estimate but it is good to think about this early on because it will help you decide the equipment and storage you will need to manage, work with, and back up the data.  \*\*Are you planning on destroying your data after X number of years? If not have you ensured your ethics and consent reflect this [where applicable].  \*\*Is it possible you might use the data for a future project?  \*\*Are you planning on making your data available to others? Do you have a funder or publisher mandate? Have you accounted for this in ethics and consent [where applicable].  \*\*How will you share the data? - Preservation copies and shared copies may not be the same format? [Fiona to speak about OM experience here] | 30m |
| Four Activity: Data Repositories – look at Figshare as example  Subject specific VS general/institutional (<https://www.ncbi.nlm.nih.gov/genbank/samplerecord/> vs <https://figshare.com/articles/Postgraduate_Support_at_Otago_Student_Questionnaire_Answers/7464938>)  -Metadata differences  -Advantages – PID, Discoverability, Version control, Metrics, Metadata only options | 10m |
| Conclusion:  Hand out and evaluation | 1m |

# Evaluation:

Post-it notes using carpentry style

# Key Readings:

Managing and sharing research data: A guide to good practice by Corti, Van den Eynden, Bishop, and Woollard – BOOK <https://otago.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=OTAGO_ALMA21174009620001891&context=L&vid=DUNEDIN&search_scope=All&tab=default_tab&lang=en_US>