My Honours Thesis - Student Outline DMP

1. General guidelines

PURPOSE OF THIS TEMPLATE
The purpose of the Outline DMP is to indicate your initial plans for how your data will be collected, shared and stored, and to give you a chance to think about these data-focused aspects of the research process. As your begin doing your research, your data process may change, and it is perfectly acceptable to change your data management plan to accommodate the changes in your research process.
Indicate below that you understand the purpose of completing this Outline DMP template.

- I understand the Outline DMP template is a projection of my anticipated data management planning requirements and should be updated as my project develops.

2. Authors and supervisors

PROJECT NAME
Replicate the title of your project, dissertation or thesis exactly as it appears in your proposal document.

From Broth to Beings; Investigating The Theory of Spontaneous Generation.

PERSONAL DETAILS
Indicate the name(s) and student number(s) of the student(s) who will be involved in this project, dissertation or thesis.

Carys de Klerk, DKLCAR005

SUPERVISOR(S) DETAILS
Indicate who will supervise this project, dissertation or thesis. If you do not yet have a supervisor, leave this section blank.

Professor Laura Nonpasteur

3. Data Collection/Generation

ORIGINAL DATA
Indicate whether you will collect or produce original data for your study. If yes, briefly describe the type of data and how you plan to manage it.
If you are unsure at this time, indicate what you think you are most likely to collect. If you are not intending to gather or collect your own data, declare that here.

- I intend to collect original data (described below).

I will create experimental data by observing microbial growth in sterilised nutrient broths and organic substrates under laboratory conditions. The data will include the presence of microbes, as well as the rate and quantity of their growth across various environmental conditions (air exposure, temperature range, humidity range). Data will be recorded in an Excel spreadsheet, with 50 observations per experiment.

DATA RE-USE
Indicate if you intend to re-use existing data, either from online searches or from datasets provided by your supervisor, lab, or funder.
If you are not intending to re-use existing data, declare that here. Also note any restrictions that apply to the re-use of data.

Question not answered.

DATA SENSITIVITY & SECURITY

Indicate whether your research data may contain sensitive, personal, disclosive, or otherwise at-risk information.

If yes, briefly describe the type of sensitivity involved and the steps you will take to secure and control access to your data.
If you are unsure at this stage, indicate what you think is most likely. If your data is not sensitive, declare that here and state how you will still ensure responsible storage.

- My data is not sensitive or at-risk.

All data will be stored on UCT OneDrive, and a private GitHub repository, with access limited to the project supervisor and the student.

4. Data Storage

DATA SIZE ESTIMATE

Indicate the estimated size of your completed dataset, and indicate whether or not you will need to access additional data storage facilities. If such storage is not provided by your unit or department, you may need to factor in the cost of purchasing additional storage space.

- 20GB or less

DATA BACKUPS

Indicate how you plan to ensure your data is secure and retrievable in case of errors or hardware failure. Describe what procedures you will put in place to back-up copies of your data and where they will be stored.

- I intend to backup my data using a service provided by UCT (UCT GoogleDrive, UCT OneDrive, Netstorage, ZivaHub etc.).

All data will be secured in the UCT OneDrive and frequently backed-up to prevent loss from hardware failure. A private GitHub repository will be used to maintain all xlsx and text files, allowing easy tracking of changes and another backup.

5. Data Sharing/Publication

DATA SHARING

According to UCT's Research Data Management policy, research data should be made open by default, with provisions for making it closed in specific cases (such as ethical considerations or potential commercialisation).
Indicate whether you plan to publish your research data.
• If yes: say where you will publish it and what licence (e.g., Creative Commons) you will use.
• If no: explain why, and refer to any ethical issues, commercial or patent intentions, or data-use agreements that prevent publication.

- I intend to share my data (details below).

All microbial growth observations and results will be publicly available on SANBI, as well as GitHub, after publication. There are no legal, ethical or commercial restrictions on this data. Metadata will be included to ensure the data is usable.

DATA DESCRIPTION

What information will you include to help others understand and use your dataset?
(For example: a short description of your study, explanations of variables, survey questions, or keywords. These are known as metadata.)

Metadata will include the experiment's methods, instruments used, and a data processing guide. Each dataset will specify that the data was collected in-house, and provide information for data cleaning or transformations.
Our dataset will include metadata following the DataCite standard. Each dataset will have structured information such as:
Dataset_Title: Name of the dataset
Creator_Name: Researcher who collected the data
Instrument_Details: Tools or equipment used
Date_Collected: When the data was generated
Location: Where the data was collected
Data_Cleaning_Notes: Steps taken to clean the data
Provenance: Source of any reused datasets, with references to the originals

6. Budget

BUDGET

Indicate any costs specifically relating to the management and curation of your data, such as purchasing additional storage space, digitisation of physical media, data storage or curation charges, and data audits. Most student research will be able to make use of free options provided by UCT and will not have to budget for data costs.

- I do not anticipate any data costs as my data is less than 10GB, and I will be using a storage system provided by UCT (UCT GoogleDrive, UCT OneDrive, Netstorage, ZivaHub, etc.) to curate my data.

My Honours Thesis - Student Full DMP

1. Project Details

PROJECT NAME - Replicate the title of your project, dissertation or thesis exactly as it appears in your proposal document.

From Broth to Beings; Investigating The Theory of Spontaneous Generation.

PERSONAL DETAILS - Indicate the name(s) and student number(s) of the student(s) who will be involved in this project, dissertation or thesis.

Carys de Klerk, DKLCAR005

SUPERVISOR(S) DETAILS - Indicate who will supervise this project, dissertation or thesis. If you do not yet have a supervisor, leave this section blank.

Professor Laura Nonpasteur

2. Project Summary

RESEARCH SUMMARY
Briefly summarise your study. Include the study's objectives, design, and methods.

This study tests the hypothesis of spontaneous generation- whether life can be created in non-living matter. Sterilised nutrient broths and organic substrates will be exposed to conditions varying in air exposure, temperature, and humidity. Microbial growth will be monitored with strict contamination controls to determine whether organisms can spontaneously form or require pre-existing organisms.

3. Data Collection/Generation

ORIGINAL DATA
If you are collecting your own data, describe the data you are gathering for your study. Briefly describe the type, scope and amount of the data you are producing.

- I am collecting my own data.

I will create experimental data by observing microbial growth in sterilised nutrient broths and organic substrates under laboratory conditions. The data will include the presence of microbes, as well as the rate and quantity of their growth across various environmental conditions (air exposure, temperature range, humidity range). Data will be recorded in an Excel spreadsheet, with 50 observations per experiment.

DATA REUSE
If you are re-using data from third-party sources in your study, record pertinent details here such as the source of the data, the extent of the data, usage rights or restrictions pertaining to the data, and how it incorporates into your study.

- I am not using existing data in my study (skip question).

4. Data Quality Control and Formats

QUALITY CONTROL
Describe what measures you are taking to ensure the data you collect are of high-quality.

All nutrient broths and substrates will be sterilised before experimentation, and strict non-contact techniques will be followed to prevent contamination as well as cross-contamination. Each treatment will be replicated at least three times to improve the precision of results.
Microbial growth observations will be cross-checked between culture counts and microscopic examination by using triangulation. All data entries will be recorded in a spreadsheet, and a copy will be kept up-to-date on GitHub. Consistency and anomalies will be checked for before statistical analysis.

FILE FORMATS
Indicate the formats in which your data will be collected and processed.
Clarify whether you will use specialised, proprietary software to produce and access your data and whether you will convert to open, accessible formats for long term access and preservation. In the case of physical objects (such as artworks or models) indicate whether these will be digitised or otherwise preserved for accessibility.

All data will be recorded in non-proprietary formats to ensure longterm use. Quantitative measurements, such as colony counts or growth rates, will be stored in xlsx spreadsheets. Qualitative notes/observations will be recorded in a Word document.

5. Data Management and Documentation

STORAGE AND BACKUP

Describe how your data is being stored and backed-up. If you are using a data service provider, provide details on how long they will retain the data.

All data will be secured in the UCT OneDrive and frequently backed-up to prevent loss from hardware failure. A private GitHub repository will be used to maintain all xlsx and text files, allowing easy tracking of changes and another backup.

DATA MANAGEMENT

Describe how you organise and manage your data.
Specify folder structures and any file-naming conventions or community data standards you have adopted.

All data will be organised in a folder system by experiment and treatment type. Each experiment will have a main folder, with subfolders for raw data, notes, and processed data. Files will be named 'date, experiment name, treatment, replicate'.

DATA DOCUMENTATION

Describe what supporting notes or files (documentation) you will provide. Documentation helps others interpret your dataset, with details such as variable definitions, coding explanations, and README files.

All data will include a README file that explains the folder's structure, naming conventions, and treatments. Any coding used will be included within the files, with in-text explanations on what the code is doing. Qualitative notes will include the context, methods, and annotations if needed to fully explain their function.

METADATA STANDARDS

Explain what structured information (metadata) you will provide about your data. Metadata helps others understand the context in which the data was collected, including methods, instruments, and provenance.

Metadata will include the experiment's methods, instruments used, and a data processing guide. Each dataset will specify that the data was collected in-house, and provide information for data cleaning or transformations.
Our dataset will include metadata following the DataCite standard. Each dataset will have structured information such as:
Dataset_Title: Name of the dataset
Creator_Name: Researcher who collected the data
Instrument_Details: Tools or equipment used
Date_Collected: When the data was generated
Location: Where the data was collected
Data_Cleaning_Notes: Steps taken to clean the data
Provenance: Source of any reused datasets, with references to the originals

6. Data Security and Confidentiality

SENSITIVE DATA

Indicate to what extent your data are sensitive.
Describe how you will control access to your data. Indicate whether you anticipate a need for encryption or password-controlled access, and if so, how you will enforce that access.

- My data is not sensitive or at-risk

All data will be stored on UCT OneDrive, and a private GitHub repository, with access limited to the project supervisor and the student.

ETHICS AND PRIVACY

Describe, as per your Ethics Clearance form or other similar documentation, any ethical or privacy issues that your data are subject to (if any). Summarise the main risks to the confidentiality and security of information related to human participants, the level of risk, and how this risk will be managed. If your project did not require ethical clearance, you may ignore this section.

- I do not require ethical clearance for my research.

Ethical clearance is not required, as the study involves only microbes with no sensitive data.

7. Data Sharing and Open Access

DATA PUBLICATION

Do you intend to publish your research data?
If you are working with sensitive data or are re-using third-party data, are there any restrictions or conditions attached to your datasets that affect data sharing, reuse, or attribution? If so, please describe them.

If you are not sharing your data, provide the appropriate justification as per the UCT Research Data Management guidelines.

- I do intend to publish my data.

All microbial growth observations and results will be publicly available on SANBI, as well as GitHub, after publication. There are no legal, ethical or commercial restrictions on this data. Metadata will be included to ensure the data is usable.

DATA REPOSITORY
Where will the research data be deposited or made available at the end of the project?

All data will be stored on a GitHub repository during the experiment. After publication, the processed data, documents, and metadata will be made publicly available to ensure long-term accessibility.

DATA LICENCE
Indicate under which licence you intend to share your research data.

- CC BY

The research data will be shared under a CC BY license. No sensitive data occurs in the study, thus the raw data will be published.

8. Relevant Institutional or Study Policies

Indicate the relevant departmental, unit, or institutional policies that influence your data management activities.

As a UCT student, the following policies guide my data management activities: the UCT Intellectual Property Policy, the UCT Open Access Policy, and the UCT Research Data Management Policy.