HANDS-ON: WRITING A DATA MANAGEMENT PLAN

# Instructions

Below you will find a fictional data management plan inspired by real research projects. The DMP is based on the Flemish Standard Data Management template.

The classroom will be divided into five different groups. Each group chooses a representative who will present the conclusions of the group for the classroom. Each section of the DMP will be addressed by a different group:

* [GROUP 1: Ethical and legal issues](#_heading=h.2et92p0)
* [GROUP 2: Documentation and Metadata](#_heading=h.3dy6vkm)
* [GROUP 3: Data storage & back-up during the research project](#_heading=h.1t3h5sf)
* [GROUP 4: Data preservation after the end of the research project](#_heading=h.2s8eyo1)
* [GROUP 5: Data sharing and reuse](#_heading=h.17dp8vu)

Exercise step by step:

1. With your group, read the first sections (project information and research data summary) to get an idea about what the objectives of the project are and what type of data will be generated during the project.
2. Then move into the section of the DMP which corresponds to your group and read the questions and answers provided. You may read other sections of the DMP, but you may only modify your group’s DMP section.
3. Discuss with the group what your opinion is about the provided answers, taking into account the characteristics of the data described in the research data summary section. Use the evaluation checklist at the end of this document to help you out.
4. Modify and/or elaborate on the answers as discussed with the group directly in this document.

After all the groups have gone through their specific section, we look at the results and the whole DMP together with the classroom.

# The Data Management Plan

## General Project Information

Concern about future fertiliser scarcity has led to numerous research in recent years about recycling nutrients from human excreta. In this project, we address some of the major current challenges for the adoption of nutrient recycling from sewage sludge: nutrient recovery and pollutant removal (WP1) and consumer perceptions (WP2).

WP1: building upon previous results, we will test the efficiency of different thermochemical treatments of sewage sludge for nutrient recovery and removal of organic pollutants and heavy metals.

WP2: we will investigate the attitudes and willingness to consume food grown with fertilisers resulting from human excreta recycling.

Hypothetical plan for training purpose based on:

<https://doi.org/10.1080/10643389.2018.1558889>

<https://doi.org/10.1016/j.wasman.2008.09.011>

<http://dx.doi.org/10.17632/kccc8m9pn9.1>

## 

## Research Data Summary

*List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.*

*If you reuse existing data, please specify the source, preferably by using a persistent identifier   
(e.g. DOI, Handle, URL etc.) per dataset or data type:*

Only new data will be generated during the project.

**WP1**

* Sewage sludge ash will undergo thermochemical treatment under a systematic variation of parameters: composition and concentration of chemical reactors, temperature and retention time. For each of the treatments, we will measure concentrations of pollutants and nutrient bio-availability.
* Data will be compiled in excel spreadsheets using templates prepared in advance in order to facilitate data entry and avoid mistakes. Data will be migrated to csv for analysis, which will be performed in R.
* The purpose of these data is to find the most efficient thermochemical treatment(s), where pollutants are removed below the permitted legal limits, while producing high nutrient availability rates.

**WP2**

* Data on attitudes towards consumption of food that has been grown using fertilisers based on recycled human excreta will be investigated by means of an online questionnaire.
  + Online questionnaires will be performed using Qualtrics. Results will be exported as csv. The analysis will be performed in SPSS (\*.sav).
  + The purpose of these data is to evaluate the willingness or unwillingness of potential consumers, and to find potential explanatory factors. The results can be used to design e.g. awareness raising campaigns or stakeholder engagement actions, in preparation of public consultation procedures for the implementation of nutrient recycling methods.

| **Dataset name** | **Description** | **New or reused** | **Data type** | **Data format** | **Volume** |
| --- | --- | --- | --- | --- | --- |
| WP1 Pollutant concentration before and after treatment | Comparison of the concentration of heavy metals and organic pollutants before and after different thermochemical treatment | New | Experimental, quantitative | xlsx, csv | <5MB |
| WP1 Nutrient bio-availability | Comparison of nutrient bio-availability after different thermochemical treatments | New | Experimental, quantitative | xlsx, csv | <5MB |
| WP1 Data from statistical analysis | Results from statistical analysis to determine the influence of the dependent variables (chemical treatment/reactors, temperature, retention time) in pollutant removal and nutrient bio-availability. | New | Analysis script & results, quantitative | R, csv, png | <20MB |
| WP2 Consumers attitudes - questionnaire responses | Data on attitudes towards consumption of food that has been grown using fertilisers based on recycled human excreta. Online questionnaires will be performed using Qualtrics and will collect:   * demographic data from participants: place of residence, age, gender, education, occupation, annual income. * respondent's opinions or views about issues in relation to the research topic: environmental awareness questions, attitudes or acceptance of the proposed product, etc. | New | Observational, quantitative & qualitative | csv | <5MB |
| WP2 Consumer attitudes - analysis | Statistical analysis of survey responses | New | Analysis results, quantitative | sav | <10MB |

## GROUP 1: Ethical and legal issues

| **Question** | **Answer** | **Group answer** |
| --- | --- | --- |
| *Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.* | No |  |
| *Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.* | No  No personal data will be processed. All data from the questionnaires in WP2 is pseudonymous data (no direct identifiers are collected). |  |
| *Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.* | Yes  The thermochemical treatment(s) developed in WP1 of the project could have commercialization potential. |  |
| *Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use  (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.* | No |  |
| *Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.* | No |  |

## GROUP 2: Documentation and Metadata

| **Question** | **Answer** | **Group answer** |
| --- | --- | --- |
| *Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, “Electronic’ Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).* | * WP1.   + Documentation will be kept on the experimental design and the protocols used in a laboratory notebook.   + A data entry template will be created to capture data of the experiments.   + The statistical analysis will be performed in R. A copy of the annotated analysis script will be kept. * WP2   + A copy of the questionnaire will be kept.   + A codebook will be created in SPSS to document variables. This will indicate the variable labels, meaning and possible values.   + The statistical method to identify explanatory variables as well as the weighing method will be documented and made available upon publication of results. |  |
| *Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used.*  *If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.* | No  A readme template will be prepared and kept together with the data files of each WP. |  |

## GROUP 3: Data storage & back-up during the research project

| **Question** | **Answer** | **Group answer** |
| --- | --- | --- |
| *Where will the data be stored?* | Data will be stored in the researcher’s laptop, and will be organised according to WP. |  |
| *How will the data be backed up?* | The data will be backed up weekly to an external hard drive. |  |
| *Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.*  *If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.* | Yes |  |
| *How will you ensure that the data are securely stored and not accessed or modified by unauthorised persons?* | The researcher’s laptop is password protected and only the researcher has access to the data.  Qualtrics account is also password protected. |  |
| *What are the expected costs for data storage and backup during the research project? How will these costs be covered?* | Due to the low volume of the data, no additional storage costs are foreseen. |  |

## GROUP 4: Data preservation after the end of the research project

| **Question** | **Answer** | **Group answer** |
| --- | --- | --- |
| *Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).* | All data will be preserved for at least five years. |  |
| *Where will these data be archived (stored and curated for the long-term)?* | The final data will be archived in an external hard drive. |  |
| *What are the expected costs for data preservation during the expected retention period? How will these costs be covered?* | Due to the low volume of the data, no extra preservation costs are foreseen. |  |

## GROUP 5: Data sharing and reuse

| **Question** | **Answer** | **Group answer** |
| --- | --- | --- |
| *Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.* | Other, please specify:  All data will be made available upon request after publication of results. |  |
| *If access is restricted, please specify who will be able to access the data and under what conditions.* | Data will be made available for verification purposes and for additional research. |  |
| *Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.* | Yes, Intellectual Property Rights  Depending on the results of the thermochemical treatment, the project might result in a patentable process. In such case, some data might be restricted until the patent has been filed. |  |
| *Where will the data be made available? If already known, please provide a repository per dataset or data type.* | Data will be made available upon request. Data will be sent using a secure transfer mechanism (e.g. Belnet). |  |
| *When will the data be made available?* | Upon publication of results and, eventually, after patent application. |  |
| *Which data usage licenses are you going to provide? If none, please explain why.* | No licenses will be used. |  |
| *Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.* | No |  |
| *What are the expected costs for data sharing? How will these costs be covered?* | No expected costs due to the low volume of data. |  |

## Responsibilities

(We will not tackle these questions in the exercise)

*Who will manage data documentation and metadata during the research project?*

*Who will manage data storage and backup during the research project?*

*Who will manage data preservation and sharing?*

*Who will update and implement this DMP?*

# 

# DMP Revision checklist

## GROUP 1: Ethical and legal issues

* Indicates whether collecting/generating or using the data involves ethical issues that need a formal ethical approval.
  + If so, clearly specifies which issues and which datasets.
* Indicates whether personal data will be used,
  + If so, refers to the specific datasets and provides details about what kind of personal data.
* Indicates whether the research has valorisation/tech transfer potential that will require restrictions on (some of) the data to protect intellectual property.
  + If so, clearly specifies which restrictions will be put in place, and for which data.
* Indicates whether third-party agreements restrict dissemination or exploitation of the data.
  + If there are third-party agreements imposing restrictions, clearly specifies what restrictions, and what data they apply to.

## GROUP 2: Documentation and Metadata

* For each dataset, describes what approach will be followed to capture the accompanying descriptive and contextual information necessary to keep data understandable and usable.
* Clearly indicates whether or not a standard (discipline-specific) metadata schema will be used to describe the data.
  + If (a) metadata standard(s) is/are used, clearly specifies which one(s).
  + If no metadata standard is used, specifies the approach taken to create metadata.

## GROUP 3: Data storage & back-up during the research project

* Clearly describes the location and storage medium that will be used for storing data during research.
* Clearly describes the locations, storage media and procedures that will be used for backing-up data
* Clearly indicates whether or not there currently is sufficient capacity for data storage & backup during the project, and justifies the answer.
  + If available storage and backup capacity is not sufficient, explains how this issue will be resolved.
* Clearly specifies what (if any) costs are expected to be incurred for storing and backing up data during the project (e.g. based on similar research/datasets and/or university policy).
  + States how these costs will be covered if applicable (e.g. by using part of the allocated grant/project budget).
* Clearly describes the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe.

## GROUP 4: Data preservation after the end of the research project

* Clearly describes which (subsets/versions of the) data and accompanying documentation will be retained for preservation and which will be destroyed, and explains rationale (e.g. legal or contractual reasons, practical issues…).
* Clearly describes what means, facilities etc. will be used to effectively preserve digital and non-digital data and accompanying documentation.
* Clearly specifies what (if any) costs are expected to be incurred for preserving data for a minimum of 5 years after the project’s end.
  + States how these costs will be covered,if applicable (e.g. by using part of the allocated grant/project budget).

## GROUP 5: Data sharing and reuse

* Clearly describes which (subsets/versions of the) data and accompanying documentation will be made available to others for reuse.
* Clearly describes who will be able to access (subsets of) the data, and under what (if any) conditions.
* Clearly indicates whether the external sharing of (some) data and accompanying documentation should be restricted or delayed.
  + If so, clearly specifies why.
* Clearly indicates where/how data will be made available to others for reuse.
* Clearly indicates when data will be made available (e.g. immediately after the end of the project, upon publication of the research results, or after an embargo period).
  + If an embargo period is indicated, specifies why and for how long.
* Clearly specifies which licence will be chosen for (subsets of) the data that will be made available.
  + If no licence is provided, clearly explains why.
* Clearly specifies what (if any) costs are expected to be incurred for sharing data.
  + EStates how these costs will be covered, if applicable (e.g. by using part of the allocated grant/project budget).