# TOWARDS DEVELOPMENT OF METHODOLOGIES FOR PLASTIC DEGRADATION PATHWAY DISCOVERY FROM META-OMICS DATA

FWO DMP (Flemish Standard DMP)

### 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.**

| | | | | Only for digital data | Only for digital data | Only for digital data | Only for physical data |
|---|---|---|---|---|---|---|---|
| Dataset Name | Description | New or reused | Digital or Physical | Digital Data Type | Digital Data format | Digital data volume (MB/GB/TB) | Physical volume |
| Metagenomic data (RAW) | Output from the shotgun sequencing | *Please choose from the following options:* | *Please choose from the following options:* | *Please choose from the following options:* | *Please choose from the following options:* | *Please choose from the following options:* | |
| | | • Generate new data | • Digital | • Experimental (gene sequences) | • .fasta | • <1 TB | |
| Scripts + packages | Scripts and packages | • Generate new data | • Digital | • compiled / aggregated data (script | • .py, .r, .csv,. text | • <100 MB | |

s,
de
riv
ed
va
ria
bl
es
)

**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:**

NA

**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

NA

**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

NA

**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.**

- No

NA

**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

NA

**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

NA

**2. DOCUMENTATION AND METADATA**

**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).**

Scripts + packages : README.txt, GitHub

**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

NA

**3. DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT**

**Where will the data be stored?**

The overall data volume is small enough that data can be stored on the OneDrive accounts of the project leader (Prof. Faust) and local hard drives. The pre-processing of raw metagenomic and metatranscriptomic reads requires a large amount of RAM memory (>60GB). For this purpose, we will use the University supercomputer cluster (Vlaams Supercomputer Centrum).

**How will the data be backed up?**

As a storage drive, the OneDrive is regularly backed up and we expect a total data volume of not more than 0.5TB. Since each OneDrive account offers 2TB, there is sufficient storage capacity. All staging units managed by University supercomputer cluster (Vlaams Supercomputer Centrum) are backed up and data can be retrieved at a request.

**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 unauthorized persons?**

The access to one drive is only given to the members of the lab and all data stored on the cluster are password and user-name protected.

**What are the expected costs for data storage and backup during the research project? How will these costs be covered?**

most of the costs will not exceed around 100 euro per year. They will be covered from grant bench-fees and PI's own lab budget.

## 4. DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

**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 the raw data (raw fastq reads) will be preserved on the PI's group cluster's space and/or external drive. Scripts will be kept in the online repository ie. GitHub.

**Where will these data be archived (stored and curated for the long-term)?**

Towards the end of the project, all the raw data will be backed up for lont-term storage.

**What are the expected costs for data preservation during the expected retention period? How will these costs be covered?**

less than 100 euro a year for a lab cluster space where all other raw data from other projects is stored. Costs will be covered by the PI's lab budget or other resources.

## 5. DATA SHARING AND REUSE

**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.**

- Yes, in an Open Access repository

Metagenomic data (RAW): raw fastq files
Scripts: .py and .ry scripts, .sql dump file

**If access is restricted, please specify who will be able to access the data and under what conditions.**

NA

**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.**

- No

NA

**Where will the data be made available? If already known, please provide a repository per dataset or data type.**

Metagenomic data (RAW): will be added to NCBI SRA repository.
Scripts and the database: will be accessible from the group's GitHub.

**When will the data be made available?**

After publication is accepted

**Which data usage licenses are you going to provide? If none, please explain why.**

Scripts and database: Apache lincence 2

**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.**

- Yes

none available currently

**What are the expected costs for data sharing? How will these costs be covered?**

No costs predicted.

**6. RESPONSIBILITIES**

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

Anna Krzynówek

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

Anna Krzynówek and Karoline Faust (PI)

**Who will manage data preservation and sharing?**

Anna Krzynówek and Karoline Faust (PI)

**Who will update and implement this DMP?**

Anna Krzynówek and Karoline Faust (PI)