

Assisted bioaugmentation for improving biological organic micropollutant degradation in drinking water treatment systems
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
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Generate new data • Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Digital • Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ... • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA 	
degradation data	Degradation data in biofilm settings under different conditions.	generate new data	digital	experimental	.xls	< 200 GB	/
biofilm images	biofilm image stacks (with confocal laser scanning microscopy)	generate new data	digital	observational	.csz, .tiff	10 GB per image, up to 1000 images	/
sequencing data	16S rRNA gene amplicon high throughput and RNAseq Illumina sequencing data	generate new data	digital	other	FastQ format	100 files, size unknown	/
OTU table	OTU table from 16S amplicon sequences	generate new data	digital	other	.csv	< 200 GB	/
manuscripts	manuscripts for publications	generate new data	digital	compiled	.pdf	< 100 GB	/

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:

/

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

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

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

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).

Overall, (e-)labbooks will contain information on experimental design, protocols, sampling location, abbreviations used, structure of the data (including link with physical storage of data), and steps involved in data analysis and relevant analysis scripts (R scripts, MOTHUR/QIIME scripts). A clear coding for all data files related to the project will be used. In the concluding stage of the project, a master index file containing the combined information for all experiments will be compiled which will be archived and also stored on the personal harddrives/PC of the PI. Altogether, this should allow any secondary analyst to use the data accurately and effectively. More specifically, the following information will be given on the items described in section 2: Degradation data in biofilm settings under different conditions will be documented with a detailed description of the used protocol in the xls file and will contain information about the sample of origin (which biofilm, which experiment, aim of experiment etc.). Biofilm microscopy images (.tif; generated by confocal laser scanning microscopy); The methodology and protocol will be described in detail in the lab book and a ReadMe file of the image collection will be written including the protocol. 16S rRNA gene amplicon high throughput Illumina sequencing data deposited at EMBL via BANKIT, will include the information/documentation required by the data base. OTU table derived from 16S amplicon sequences and containing the relative abundances of each Operational Taxonomic Unit (OTU) for each sample will contain information about the sample of origin (which biofilm, which experiment, aim of experiment etc.)

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

16S rRNA gene amplicon high throughput Illumina sequencing data deposited at EMBL via BANKIT, will include the information/documentation required by the data base.

3. Data storage & back-up during the research project

Where will the data be stored?

Data collected/generated will be stored on KU Leuven OneDrive (capacity: 250 GB) and on the KU Leuven personal I server (capacity 50 GB) and shared J server (capacity: 1 TB). Nucleic acid sequence data will be submitted and stored in official nucleic acid data bases like EMBL upon publication or at the end of the project. Manuscripts will be made publicly available through the LIRIAS repository from KU Leuven.

How will the data be backed up?

The KU Leuven OneDrive automatically makes a back up of all data. Additional back-ups are made on the personal computer of the researcher and on the KU Leuven servers.

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

Data collected/generated will be stored on KU Leuven OneDrive (capacity: 250 GB) and on the KU Leuven personal I server (capacity 50 GB) and shared J server (capacity: 1 TB).

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The KU Leuven OneDrive is secured, however, there is no need for special security since no sensitive data is generated in this project.

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

The storage is provided by KU Leuven servers, no extra costs are required. If there are any additional costs (not expected), this will be covered by the FWO project.

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 data will be preserved for at least 5 years after completion of the project.

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

The data will be archived on the KU Leuven central server (:K) archive drive, conform the KU Leuven RDM policy.

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

Not expected to be high. If costs, they will be covered by on-going related projects.

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

The full dataset (except the nucleic acid sequence data) will be deposited in a .cvs format in KULEuven RDR under a CC-BY license. The 16S nucleic acid sequence data will be available through official nucleic acid databases like EMBL.

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

Not applicable

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

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

The full dataset will be made available through KU Leuven RDR.

When will the data be made available?

The full dataset will be uploaded and made available in a cvs format in RDR immediately after the end of the project in case published. Others will be added upon publication. If not published within 1,5 years of project completion, all datasets will be made available.

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

The full dataset will be deposited in cvs format in KU Leuven RDR under a CC-BY license.

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

The KULeuven uses a DIO (not yet available).

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

No direct idea but expected to be not high. In case, costs will be covered by related projects.

6. Responsibilities

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

The researcher

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

The researcher

Who will manage data preservation and sharing?

The promotor

Who will update and implement this DMP?

The PI bears the end responsibility of updating and implementing this DMP.