Microbiome insights into pollution exposure and impact in harbour seals
Application DMP

Questionnaire

The questions in this section should only be answered if you are currently applying for FWO funding.
Are you preparing an application for funding?

- No

Microbiome insights into pollution exposure and impact in harbour seals
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Microbiome insights into pollution exposure and impact in harbour seals
GDPR

GDPR

Have you registered personal data processing activities for this project?

- No

Microbiome insights into pollution exposure and impact in harbour seals
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
WP1_T1.1_T1.2_sampling	rectal swabs samples collected from free-ranging harbour seals and stranded harbour seals (at admission, during rehabilitation, before release)	New	Physical	Observational			400 samples
WP1_T1.1_T1.2_metadata	animal sex, weight, age class, health status, diet composition and pollutant and metabolite concentrations in the blood	Reused	Digital	Observational	.csv .tab	< 100MB	
WP2_T2.1_database	custom database of genes encoding pollutant-degrading enzymes, focused on PCBs, PBDEs, DDTs, HCHs and HCB. It will include DNA and RNA sequences, catalysed reactions and bacterial taxonomy, determining pathways and enzymes involved in anaerobic processes as well as aerobic processes.	Reused	Digital	Experimental	.csv .tab .fastq	< 1TB	
WP2_T2.2_sequencing	metagenomics sequencing data from rectal microbial communities sampled from free-ranging and stranded harbour seals - shotgun sequences (paired-end 150bp) from Illumina NovaSeq.	New	Digital	Observational	.fastq	< 1TB	
WP2_T2.3_sequencing	metatranscriptomics sequencing data from rectal microbial communities sampled from free-ranging and stranded harbour seals - shotgun sequences (paired-end 150bp) from Illumina NovaSeq.	New	Digital	Observational	.fastq	< 1TB	
WP2_T2.4_POP_concentrations	POP concentrations throughout the biodegradation experiment (measured using GC-MS)	New	Digital	Experimental	.csv	< 100MB	
WP2_T2.4_MG_sequencing	metagenomics sequencing data from the end of the biodegradation experiment - shotgun sequences (paired-end 150bp) from Illumina NovaSeq.	New	Digital	Experimental	.fastq	< 1TB	
WP2_T2.4_MT_sequencing	metatranscriptomics sequencing data from the end of the biodegradation experiment - shotgun sequences (paired-end 150bp) from Illumina NovaSeq.	New	Digital	Experimental	.fastq	< 1TB	
Materials and methods	Protocols and experimental designs	New	Digital	Experimental	.pdf	< 100MB	
Data analysis scripts	R scripts for the data analysis of both sampled seals as well as degradation experiment	New	Digital	Experimental	.R	< 100MB	
Manuscripts	Written manuscripts, revisions and papers	New	Digital	Experimental	.pdf	< 100MB	

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:

WP2_T2.1_database - for this custom database a comprehensive literature search and existing database search will be performed. For that reason I cannot specify specific DOIs or URLs yet.
WP1_T1.1_T1.2_metadata - this metadata will be created by collaborators, for that reason I cannot specify an identifier yet.

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

I have confirmed with KU Leuven that this project does not require ethical approval from the Ethical Committee for Animal Experimentation. However, I am required to inform KU Leuven about the collaboration with ITAW, Sealcentre Pieterburen, and Sea Life Blankenberge, providing details about the research plan and its ethical considerations, as detailed below:
The Institute for Terrestrial and Aquatic Wildlife Research (ITAW), led by co-supervisor Dr. Kristina Lehnert, catches free-ranging seals twice yearly in the frame of a health monitoring of the Schleswig-Holstein government. Harbour seals are examined clinically, samples of blood, saliva, hair and rectal swabs are taken. ITAW is experienced in this work and has the necessary permits to carry out the seal catch, and sampling described in the project. ITAW is in possession of national permits from the Ministry for Energy Transition, Climate Protection, Environment and Nature (MEKUN) of Schleswig-Holstein, as well as a University's ethical approval (V312-72241.121-19 (70-6/07)).
The seal catch for rectal swabs has been designed with the 3R principle of replacement, reduction and refinement. The lowest manageable number of seals needed to produce meaningful results and solid data from a statistical perspective are being sampled (reduction). The catching action and subsequent sampling are carried out by experienced veterinarians by working silently and swiftly to reduce stress for the seals (refinement). During sampling animals are retained by trained staff and cooled with seawater poured over their backs to prevent overheating (refinement).
Sealcentre Pieterburen and Sea Life Blankenberge operate under Dutch and Belgian government authorization for harbour seal admission and rehabilitation, without the need for invasive sampling permits (as rectal swabs are seen as non-invasive), in compliance with EU directive 2010/63/EU.

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).

Experimental protocols and designs will be documented in .docx files, including sample handling, sequencing, and analytical procedures.
Metadata (e.g., abbreviations, variable definitions, data structure) will be described in dedicated .docx and .tsv codebook files.
Raw and processed data will be stored in standard formats (.csv, .xls, .fastq, .qza, .txt) with clear, versioned filenames.
Scripts for analysis will be written in R and saved as .R files, accompanied by .docx files that describe script function, input/output formats, and parameter settings.
README.txt files will be created for each WP and updated regularly to describe file contents, locations (e.g., OneDrive, institutional servers), and naming conventions.
Version control will be implemented via Git for code and key documentation files.
Where applicable, Electronic Lab Notebooks (ELNs) will be used to track experimental decisions, deviations, and observations in real-time.

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.

- Yes

Where appropriate, metadata standards aligned with the target repositories will be applied to ensure findability and reusability. For example, sequencing datasets will follow the MIxS (Minimum Information about any (x) Sequence) standard; ecological and environmental data will comply with Ecological Metadata Language (EML); and chemical data (e.g., POP concentrations) will follow ISA-Tab or equivalent. For datasets deposited in Dryad or institutional RDRs, repository-specific metadata schemas will also be completed to ensure full interoperability with FAIR data principles.

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

Where will the data be stored?

The data will be stored on my personal computer and on my OneDrive account.

How will the data be backed up?

The data will be backed up on a local server (K drive) as additional long-term back-up at the KU Leuven Campus Kortrijk.

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

The available space on the K drive is 10TB which is enough to store all data.

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

The data on my personal computer, OneDrive and the K drive server are all protected by a password.

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

There are no expected costs of the data storage, since the available storage is enough to store all our data.

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 retained for 10 years after the end of the project, according to the KU Leuven guidelines.

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

All data will be archived on the K drive local server of our research group. Once published, R scripts and bio-informatics pipelines will be deposited on GitHub, sequencing data will be submitted to the NCBI Sequence Read Archive (SRA) and other data (POP concentrations, metadata) will be available online on data repositories such as Mendeley Data.

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

There are not costs for the data preservation as our local server is already in place and GitHub, Mendeley Data and the SRA are free to use.

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

Each dataset will be deposited in an online data repository after publication.

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

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

R scripts and bio-informatics pipelines will be deposited on GitHub
Sequencing data will be submitted to the NCBI Sequence Read Archive (SRA)
Other data (POP concentrations, metadata) will be submitted to online data repositories such as Mendeley Data.

When will the data be made available?

Upon publication of the results.

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

We will use the CC BY 4.0 license, which allows others to share and adapt the data with appropriate attribution. This license aligns with FAIR data principles and is recommended by Mendeley Data and most major journals

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?

There are no expected costs for public data repositories, such as NCBI, Mendeley Data or GitHub.

6. Responsibilities

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

Broos Van de Moortel

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

Broos Van de Moortel

Who will manage data preservation and sharing?

Broos Van de Moortel

Who will update and implement this DMP?

Both Broos Van de Moortel and the PI Ellen Decaestecker bear this responsibility.