PROGRASS: Polymer Recovery Optimization from industrial GRAnular Sludge Systems

A Data Management Plan created using DMPonline.be

Creator: Toon Coelmont

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1SHDO24N

ID: 204325

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

To bring the world closer to “clean water for all”, this project focuses on the aerobic granular sludge (AGS) technology as a more sustainable alternative for the current activated sludge systems (less energy and footprint and higher resource recovery potential). Especially industrial wastewaters, rich in, e.g., lipids and proteins, hold much potential but are underexploited in AGS applications due to their complex and variable composition. Hydrolysis is indeed most often required and rate limiting. Moreover, if the influent required hydrolysis capacity is not met by the AGS’ hydrolysis capacity, then the granulation is impaired. With as first objective to a enable robust AGS reactor operation for industrial wastewaters, this project will (i) develop an extensive set of monitoring tools, focusing on hydrolysis capacity, storage and structural polymers and other activity measurements, (ii) develop a synthetic granule community to enable clear conclusions when the impact of the influent complexity is studied and (iii) develop a control strategy to boost and tune the AGS’ hydrolysis capacity. In general, granulation is enabled by structural polymers (stEPS), induced by feast-famine conditions, but also the slow growing organisms that store internal polymers are said to contribute. Both the stEPS and the storage polymers have industrial value when recovered. The project, therefore, also investigates waste granule pretreatment techniques to maximize (storage) polymer recovery.

Last modified: 30-04-2024

PROGRASS: Polymer Recovery Optimization from industrial GRAnular Sludge Systems
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

In this research, different qualitative and quantitative data will be generated. Depending on the specific equipment used, the data will be either noted down in lab notebooks, generated in an excel format or saved in microscopical images. The former will be stored as electronical PDF-files, while the latter will be saved in the format that contains the most information (svi). Raw 16S rRNA sequencing data will be analyzed using standard, documented pipelines in R and raw sequences will be uploaded in publicly available databases (such as NCBI). Additionally, all raw data will be stored separately from modified data. Calculations and plotting of graphs will be mainly done in Excel and Python.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

During the research, the data will be stored on dedicated university servers (OneDrive, 'property' of the PhD-student) and backups are generated on a regular basis on both university servers (OneDrive, 'property' of Prof. Ilse Smets) and external hardware (external hard drive). At the end of the research, standard procedures of KU Leuven already foresee that all data is copied to dedicated project folders, while the data will also still be available on both the external hard drive and the OneDrive of Prof. Ilse Smets.
The data will be stored on KU Leuven servers for at least 5 years and even longer, as long as storage limitations do not occur.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Not applicable, data preservation is guaranteed for at least 5 years.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Not applicable.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Not applicable.

PROGRASS: Polymer Recovery Optimization from industrial GRAnular Sludge 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
HPLC/GC/IC/TOC data	Chromatogram in standard raw format, processed data	New	Digital	Experimental	Standard Agilent Instruments format, .txt, .csv, .pdf	<100GB	NA
Measurement data (OD, dry weight, hach test kit results...)	Measurements of weight, absorbance or visual observations are logged in a lab book and digitalized in excel formats.	New	Digital	Experimental	lab notes, .csv, .xlsx	<1GB	NA
Plate reader data (fluorescence measurements)	Hydrolysis activity is measured through fluorescence measurements in a plate reader assay.	New	Digital	Experimental	.xlsx	<1GB	NA
Proteomics	Proteomics data about the proteins and their corresponding activity.	New	Digital	Experimental	chromatographic 3D maps	<10GB	NA
qPCR	qPCR data, standard curves, corresponding bacterial counts.	New	Digital	Experimental	.csv	<1GB	NA
Microscopic images	Microscopic images of bacterial samples (confocal, fluorescent, regular microscopy)	New	Digital	Experimental	.tiff	<500GB	NA
Protocols, guideline documents	Description of procedures, fill in documents	New	Digital	Experimental	.txt	<1GB	NA
Bacterial samples	All bacterial samples from, e.g., reactor runs	New	Physical	NA	NA	NA	<10L (in total)

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:

Not applicable.

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.

- Yes

Specific technological inventions (e.g., separation mechanism for flocs and granules) have the potential for valorization (patent). The data (movies, pictures, design and measurements on the design) are stored under a separate folder and a READ Me-file is added with comments on how to use this data (or not use the data) in research communication before application for a patent. The Leuven Research and Development team is contacted and will provide specific instructions on handling this information/data appropriately and potentially file for a patent.

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).

README.txt files will accompany each data set to explain experimental procedures, data-generation, data-analysis. These files will have the following content: the
experiment date, reference name, description, goal, used protocols, generated type(s) of data, comments about the data, general conclusions, remaining questions and thoughts/ideas for future experiments. The corresponding (raw) data will be available in the parent folder of this README directory, in the folder "Data", with all corresponding filenames and directories using the same reference name as described in the documentation. All relevant files will be named YYYYMMDD_shorthand_description.extension for easy lookup purposes. The referenced protocols (SOP) are also contained in a separate folder, and structured in .docx and .md formats. Shortcuts are included in the filesystem if needed.
In terms of processing data, all manipulations will be done in Jupyter Notebooks, containing relevant documentation in Markdown format where needed. Documentation for the applied Python modules (e.g. Pandas) will not be included, as this is readily available in the module docs themselves.

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

To our understanding, there is no formal metadata standard for our discipline.
However, for image acquisition, the Bio-formats standard is applied by saving all raw confocal image data in the supported .oif format. This directory containing raw data is separated from the processed versions for transparency and reproducibility.

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

Where will the data be stored?

While a working copy of all relevant data to this project (data, manuscripts, scripts...) are kept on the researcher's laptop, both a local hard drive and remote (Onedrive and Bitbucket repository) services will contain the same information (automatic sync on Onedrive is used if network is available). Instrument data from the HPLC/GC/IC/TOC and Olympus microscope are stored on the machine itself for one month, and archived locally by our technical staff. Confocal images are stored on a network drive.

How will the data be backed up?

All files relevant to this project (data, manuscripts, scripts...) will be backed up automatically on Onedrive as provided by KU Leuven (if network is available). In addition, a separate hard drive is used for the same purpose to provide a local backup as well. These files will be retracable for at least 8 weeks. Regular backups will also be made to the Onedrive of the PIs.
Concerning code (mostly Jupyter notebooks), LaTeX, and text documents such as protocols, READMEs, sources and bibliography will be pushed to a remote (private) Bitbucket repository on the researcher's (KU Leuven email) account. Large files such as microscopic images and raw chromatographic data are omitted from this form of data storage to avoid bloating the Git repository.

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 Onedrive storage space (both of the researcher and the PIs), together with the local drives provide over 2 TB of storage space. This suffices for this project.

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

KU Leuven security measures (2-step authenticator, confidentiality level security...) are in place and will be used.

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

The storage service (Onedrive) is provided by KU Leuven, and the associated costs are therefore indirectly covered by project overhead cost.

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, processing scripts, structured output and papers will be retained for the expected 5 years, no restrictions apply.

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

Backups as stored on Onedrive are stored for unlimited time, and will be shared with the group for accessibility.

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

The costs of preservation are again indirectly covered by overhead project costs.

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

Open Access or restricted access is decided by the projects PIs. Restricted access would be necessary for patent/confidential/company/non-published data.

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.

Since the research group will have access to the full data via Onedrive, interested third parties can obtain raw data upon request. Published data, however, will be available through a public repository (e.g. BioRxiv) or through the journal itself.

When will the data be made available?

Immediately after the end of the project.

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

CC-BY-NC-ND-4.0 or another creative common license, if more applicable for the specific data set. The data should be protected and may only be used upon agreement with the projects PIs.

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

DOI numbers.

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

The costs of data sharing are again indirectly covered by overhead project costs.

6. Responsibilities

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

The PIs (Prof. Smets and Prof. Dries) will be the main responsible. However, the daily management of data and its documentation is the responsibility of the doctoral researcher.

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

The doctoral researcher. For backups, on the instruments themselves, the responsibility lies with the technical staff.

Who will manage data preservation and sharing?

Onedrive backup preservation is a responsibility of the KU Leuven ICTS. Possible data sharing is monitored by the PIs.

Who will update and implement this DMP?

The PIs bear the end responsibility of updating & implementing this DMP.