Initial DMP

Project Name Building robust flocs and granules on demand for wastewater treatment and resource recovery (FWO DMP) - Initial DMP

Project Identifier 1191022N

Grant Title 1191022N

Principal Investigator / Researcher Laurens Parret

Description Flocs or granules are the cornerstones of process intensification in biotechnological processes. They enable (i) an easy biomass-liquid separation, (ii) a better retention of slow growing organisms and their key metabolic functions, and (iii) a stratification of the microorganisms, allowing for synergistic processes or shielding from unfavorable environmental conditions. Evidently, the building blocks and strategies that lead to strong aggregates, need to be well known. Despite the many years of research in bioflocculation, the fundamental insights in the interaction between the different building blocks of the aggregates are still missing. Furthermore, the knowledge that has been obtained for flocs has not been transferred to granules. This project will, therefore, first focus on deepening that knowledge by zooming in on the separate building blocks: microcolonies and their polymers (such as adhesins), and filamentous organisms or fibers. Secondly, on the basis of the knowledge gained, we will propose generic guidelines to build new functionally and structurally robust flocs or granules. The aim is to bring together micro-organisms with complementary functionality but conflicting environmental needs, while retaining the strength of such an aggregate by enclosing structuring micro-organisms or elements. Novel strategies for cultivating or bioaugmenting slow growing organisms will be developed, exploiting a vast array of microscopic and molecular monitoring techniques.

Institution KU Leuven

1. General Information

Name applicant

Laurens Parret

FWO Project Number & Title

1191022N - Building robust flocs and granules on demand for wastewater treatment and resource recovery

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

The project aims to investigate natural bioflocculation in pure cultures experiments. Following a bottom-up approach, in the final stage of the study, the gathered knowledge will be applied to a validation experiment with sludge or granules. Throughout the course of the project, a set of guidelines will be constructed that could help further research (and industrial parties) towards developing more sustainable water treatment technology. The content of the guiding document will depend on the specific results of the experiments, though should contain clear .pdf diagrams, flowcharts, and an extensive protocol discription and bibliography.

In any case, the organisms of interest (which contribute to the nitrogen cycle) will be monitored in pure and co-culture conditions. Their extracellular metabolites are followed up using high pressure liquid chromatography (HPLC), while the nitrogen content in all of its oxidation states (ammonium, nitrite, nitrate) is determined with the respective Hach Testkits. Growth is indirectly measured with optical density (OD600), and linked to dry weight (DW) when needed. Amino acids can be measured with gas chromatography (GC), and proteomics is conducted with a partner university.

Visualisation of the developing structure of the floc (and in the first stage of cells entrapped in alginate beads), is achieved through both confocal microscopy, and cryosectioning combined with fluorescence microscopy. Below an overview and size estimation of the relevant data is provided.

Type of data	Format	Volume	Method
HPLC and GC data	chromatogram in standard (raw) Agilent Instruments format	>100 GB	UV transmitted through the sample, and refractive index (RI) measurements compared to a reference cell. For each measurement series, a quality control (QC) sample with known composition is included as well.
Processed HPLC data	.txt, .csv	<10 GB	Automated peak integration of the chromatogram using calibration curves.
OD measurement	lab notes, .csv	<1GB	Optical density measured at a wavelength of 600 nm
Dry weight	lab notes, .csv	<1GB	Dried (cellulose acetate) filters before and after biomass filtration
Hach testkit results	lab notes, .csv	<1GB	Sample loading in a ready-to-use vial as sold by Hach
Proteomics	chromatographic 3D maps	>10GB	Mass spectrometer data of digested proteins (into peptides)
Cryosectioning	.tiff	<100 GB	Fluorescent images of slices of the sample
Microscopic images	.tiff	<500 GB	Confocal and fluorescent images and stacks are constructed by the instrument using laser light. Its intensity is captured by the PMT detectors in the confocal (Fluoview) microscope, or simply by a camera in the case of a traditional setup
Encapsulator operating conditions	lab notebook, .csv	<100MB	Operating conditions of encapsulator (frequency and pressure) during bead formation.
Protocols	.pdf, .md, .docx	<1GB	Description of optimised protocols of the conducted experiments and setups
Guidelines document	.pdf	<100MB	Based on all previous experiments

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

The data will not directly translate to a possible tech transfer. This is more relevant for follow-up projects that can be guided based on the results of this fundamental research.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

A folder containing "README" text files in Markdown format will be maintained. Here, each experiment will get a dedicated and structured description, including (but not limited to): 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.

For encapsulator and microscope settings, the relevant manual parameters are included in the YYYYMMDD_README.md files. Detailed acquisition documentation for the confocal microscope is contained in the .oif files of the stacks themselves.

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? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/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.

5. Data storage and backup during the FWO project

Where will the data be stored?

While a working copy of all relevant 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. Instrument data from the HPLC 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 is backup of the data provided?

All files relevant to this project (data, manuscripts, scripts...) will be backed up weekly to Onedrive as provided by KU Leuven. 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.

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

What are the expected costs for data storage and back up during the 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.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The laptop with a working version of the data is protected using a password/touch-ID, while the external hard drive will be encrypted. The cloud storage is also only accessible through the staff KU Leuven login of the researcher.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

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

Where will the data be archived (= stored for the longer 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 retention period of 5 years? How will the costs be covered?

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

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

Raw experimental data and protocols will be available for reuse within the research group for follow-up projects. Relevant data in publications will be made available in the respective international journals.

Where/how will the data be made available for reuse?

- Other (specify):

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

Who will be able to access the data and under what conditions?

Published results will be available to anyone, while all other data will be shared with specific staff members who can grant others access if necessary.

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

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

8. Responsibilities

Who will be responsible for data documentation & metadata?

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

Who will be responsible for data storage & back up during the project?

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

Who will be responsible for ensuring data preservation and reuse ?

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

Who bears the end responsibility for updating & implementing this DMP?

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