DMP_11H0321N

Project Identifier G0C5322N

Principal Investigator / Researcher Abram Aertsen

Project Data Contact Abram.Aertsen@kuleuven.be

Description In this proposal, we will interface synthetic biology, gene regulatory modeling and population dynamics modeling to establish and standardize novel and tunable tools to genetically reprogram bacterial populations. More specifically, we will focus (i) on phage mediated delivery as highly efficient way to transmit a genetic instruction throughout a bacterial population, (ii) on the design of synthetic circuitry that imposes staged genetic instructions to the cells (facilitating growth, production and harvesting), and (iii) on the construction of a population balance model equipped with genetic circuitry and metabolic pathways that computes optimal phage and circuitry tuning and bioprocessing conditions. The elaborated transmission-tool, genetic circuitries, and modeling approaches will provide a novel platform to dynamically trigger expression of desired proteins throughout bioreactor populations, and will lay the basis for an even broader venue to reach and instruct bacterial populations in their natural and often difficultly accessible habitats.

1. General Information

Name applicant

Abram Aertsen

FWO Project Number & Title

G0C5322N – Next-generation bioprocess engineering: phage-based synthetic reprogramming of bacterial bioreactor populations

Affiliation

- KU Leuven

2. Data description

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

- Generate new data
- Reuse existing 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.

Type of data	Format	Volume	How created
Microscopy images	.nd2/.tif	1000GB	Phase contrast and fluorescence (timelapse) microscopy of bacterial samples
Genetic sequences	.gb/.fasta	<5MB	Results from sanger/whole genome sequencing and design of constructs/primers mainly using the "Benchling" portal
Notes, Reports, Presentations	.docx/.xlsx/.pptx Lab-journal	<5GB	Using the Microsoft Office suite handwritten notes i.e. plate countings will be digitalized and further expanded. Furthermore, presentations and summaries/papers based on the obtained research data will be created
Multiwell plate reader	.xlsx	<5MB	Multiwell plate readers will quantify the development of optical density and fluorescence over time and report respective numerical values in tables.
Bioreactor data	.xlsx	5GB	Bioreactor data are threefold: (i) a materials and methods overview and experimental design (ii) data recorded from the online monitoring and control of each bioreactor run

			(extracted from the datalogging software), and (iii) off-line analysis including cell and phage counts, metabolites data, flow cytometric data, etc.
Model versions	.m, .py	10GB	Models will be programmed in Matlab, Python, or alike. Model versions are carefully tracked and meaningful versions are stored.

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

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?

1. All microscopy experiments have metadata included into the resulting image file with manually written information about the type of experiment, strains, conditions. For time lapse microscopy experiments further information about temperature of incubation and specific information of each respective recorded position are gathered, if applicable. Generic metadata about i.e., dimensions, image type, bitdepth, pixel sizes and microscope settings are also included. Experimental specifications i.e., about sample preparations are reported in handwritten notes in lab journals with clear reference to the microscopy-file. Digitalized protocols for individual experimental archetypes are present or will be created for potential later reproduction of results by other individuals.
2. Experimental results are readily summarized and stored locally as well as in our institute-based cloud "OneDrive", together with text files describing the respective experimental setup and conditions.
3. Genetic data is either generated or imported into the "Benchling" portal, a cloud based online platform and respective metadata is added to the files. Access to data is shared inside the lab, making data internally accessible even after a person has left the lab. At regular intervals offline backups of the generated data are made and saved on an internal network and external hard drives.
4. Github will be used as a repository for constructed models with careful version control. Through this repository, published models versions and data can be made available for external parties.

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.

- Yes

For microscopy data a standardized protocol with fill-in fields has been created, so that the blanks can easily be filled with the corresponding information.

The protocols describing how experimental data has been obtained are stored on a shared network drive.

For instrumental readouts, protocols have already been established explaining the respective numerical outputs.

5. Data storage and backup during the FWO project

Where will the data be stored?

All research data will be digitalized or is already digital by nature.

1. Copies of the data will be/are stored on a personal computer and automated cloud storage in parallel.
2. Microscopy data is saved on the microscopy computer and immediately copied onto a PC and an external hard dive.
3. Genetic sequence data is stored in a could for easy shared access.
4. Lab journals are kept in the lab and are stored in an archive once a person has left the lab.
5. Model and analysis source code is kept on a cloud-based version control system (github / bitbucket).

How is backup of the data provided?

Data is backed up automatically via online cloud storage.

At regular intervals backups of all data will be created on large external hard drives. This is especially the case for microscopy data too large for online clouds, these data will be thus stored in threefold to avoid loss of data due to data corruption (local computer, external computer, external hard drive).

Once projects or chapters of larger projects are completed, i.e. publication of data, said data will be archived in an internal network.

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

Sufficient space is arranged for local data storage on computer hard drives.

Online institute-based storage space of 2000GB per researcher for automated backups is sufficient and may be expanded if necessary.

Lab Internal network storage can be expanded if needed but still has plenty of capacity.

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

Ca. 100 Euro/year/person for extended storage on backed-up network drives are allocated. This is covered by the budget for computer equipment.

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

None of the data are publicly available and no personal data will be generated in the scope of this project making the data less sensitive.

However, all raw data files will only exist on internal networks or hard drives protected by passwords, respectively, and thus not be accessible by unauthorized persons.

The same goes for cloud-based data i.e., sequencing data.

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 important data, namely data from conclusive experiments linked to the eventual dissertation or publications will be stored in an institute-based cloud. Furthermore, all relevant microscopy data will be archived. Lab journals will physically be stored and kept by the lab manager.

Preliminary data might not be retained in case datasets would be too small to be of statistical relevance and/or in case of easy and low-cost reproducibility.

Where will the data be archived (= stored for the longer term)?

Data and metadata will be stored on the university's central servers (with automatic back-up procedures) for at least 5 years, conform the KU Leuven RDM policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Ca. 100 Euro/year/person for extended storage on backed-up network drives are allocated. This is covered by the budget for computer equipment.

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?

Experimental data will only be shared inside the lab or with collaborators upon request. In line with the FWO open access obligation, publications resulting from this project will be published in open access journals or accepted manuscripts will become available via the KU Leuven Lirias depository.

Data related to unpublished work will not be made available and internally used for further research.

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

- In a restricted access repository
- Upon request by mail
- Other (specify):

1. The main tool for reuse of data for the lab is the institute-based cloud service
2. For larger file sizes the internal network will be used

3. For external collaborators files can be shared via mail or by sharing cloud-based OneDrive folders

When will the data be made available?

- After an embargo period. Specify the length of the embargo and why this is necessary
- Upon publication of the research results

The dissertations resulting from this project and supporting data will typically be held under an embargo for 5 years to protect follow-up projects and related ongoing work.

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

After the embargo has been lifted the dissertations and supporting data will be publicly available.
All other research data will remain only accessible to members of the lab unless specific exceptions are granted for non-profit researchers.

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

Minimal costs are expected. These can be covered by budget allocated to miscellaneous costs.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The PhD students pursuing this project are responsible for maintaining sufficient documentation about generated data and creating respective metadata.

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

This obligation is shared between the students and lab manager. The students create short term backups, as previously mentioned, and stores the data. Long term storage and backups are being overseen by the lab manager.

Who will be responsible for ensuring data preservation and reuse?

The PhD students are responsible for initial data preservation and general reusability of data. Once the data has been backed up/archived, the institute´s IT department makes sure the data is preserved and maintains it intended accessibility.

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

The PhD students are responsible for the direct implementation of this DMP. The PI bears the end responsibility of updating & supervising the lab’s data management.