Synthetic Transduction as a Novel Tool for Population Programming

A Data Management Plan created using DMPonline.be

Creators: Kevin Broux, n.n. n.n.

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Data Manager: Kevin Broux

Grant number / URL: 11Q4T24N

ID: 204609

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

In this project, genetics, live cell biology and image analysis will be used to establish a novel system for population programming. Genetic transduction by phages will be engineered as a tool for unlocking future microbial biotech applications in industry and therapy.

Last modified: 13-03-2024

Synthetic Transduction as a Novel Tool for Population Programming
DPIA

DPIA

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

- Not applicable

Synthetic Transduction as a Novel Tool for Population Programming
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Synthetic Transduction as a Novel Tool for Population Programming
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)

Type of data	Format	Volume	How created
Microscopy images	.nd2/.tif	200-500GB	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.

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)

1. 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.
2. 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'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

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

Synthetic Transduction as a Novel Tool for Population Programming
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.

Type of data	Format	Volume	How created
Microscopy images	.nd2/.tif	200-500GB	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.

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.

- 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).

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 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. Digitalized protocols for individual experimental archetypes are present or will be created for potential later reproduction of results by other individuals.
2. 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. Additionally, for all primer sequences and all constructed bacterial strains, a comprehensive list will be provided featuring all the necessary information for further use (annealing temperatures, explanation of the primer's purpose and the primary template for the primer's table; For the strain's table: Species, full genotype, present plasmid(s), growing conditions, comprehensive notes on the construction of the strain, date of construction and location in the -80°C freezer). Sequence data itself is always coupled to a genetic sequence in 'Benchling'. The folder containing the raw sequence data will be dated and given the same name as its cognate genetic sequence.
3. Experimental results and other work documents are readily summarized and/or stored on our institute-based cloud "Onedrive for business", together with text files describing the respective experimental setup and conditions.

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

- For microscopy data, a metadata model format with fill in fields has been created, so that the blanks can easily be filled in with the respective information. As mentioned before all other generic metadata has a standard format in the NIS-elements software and is provided for every microscopic picture taken.
- Genetic data is not described using a general metadata format. A description of any genetic sequence/construct is provided and annotations of certain parts of the sequence are given. The list of primers, used to create these genetic sequences, and the strainslist summarising strains equipped with these genetic constructs provide additional information. A personal semi-metadata system is used to describe day-to-day research activities in the lab journal or in digital summarized experimental results/ work documents: The same type of information is always provided when performing a certain experiment.
    - An example: For any PCR that is run, the following information is provided: Type of PCR and additional information if needed for a specific PCR type, Purpose of PCR, Polymerase used, primers used, annealing temperature, template used and length of produced DNA with elongation time of the PCR. For instrumental readouts, standard metadata formats have already been established explaining the respective numerical outputs.

3. Data storage & back-up during the research 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 micoscopy computer and imeadiately 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

How will the data be backed up?

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

Two personal 4TB hard drives from different suppliers and different types (SSD/HDD) are used to create regular two-fold backups of all research data, especially regular copes of microscopy data.
Online institute-based storage space of 100GB 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.
The Mango platform of KULeuven will be used for cloud storage of active microscopy data.

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.

What are the expected costs for data storage and backup during the research 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.

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 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 these data be archived (stored and curated for the long-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. Lab Journals are physcically archived in the host institution's lab in a secure cabine with an easy access.

What are the expected costs for data preservation during the expected retention period? 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.

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.

- No (closed access)
- Other, please specify:
- Yes, in a restricted access repository (after approval, institutional access only, …)

1. The main tool to find and reuse data for the lab is the institute-based cloud service: 'Ondrive for business'. For genetic sequences/constructs and sequencing data specific, the data is/can be shared via the 'Benchling' cloud service.
2. For larger file sizes the internal network will be used.
3. For external collaborators, files can be shared via mail or Onedrive for business.

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

Experimental data will only be shared inside the lab or with collaborators upon request. In line with the FWO open access obligation, if possible publications resulting from this project will be published in open access journals and thus be publicly available.
Data related to unpublished work will not be made available and internally used for further research.

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.

- Yes, Intellectual Property Rights

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

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 shred via mail

When will the data be made available?

The dissertation 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.

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

Attribution-NonCommercial-NoDerivs 4.0 International

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

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

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

6. Responsibilities

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

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

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

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

Who will manage data preservation and sharing?

The PhD student is 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 its intended accessibility.

Who will update and implement this DMP?

The PhD student is responsible for direct implementation & updating of this DMP. The PI bears the end responsibility of this DMP.