DMP title

Project Name C1 project biofilm architecture - DMP title
Project Identifier C14/22/077
Grant Title ZKE2317
Principal Investigator / Researcher Bart Smeets
Institution KU Leuven

1. General Information

Name of the project lead (PI)
Bart Smeets

Internal Funds Project number & title
ZKE2317
The impact of biofilm architecture on the evolution of antimicrobial resistance: a single cell modelling and experimental approach

2. Data description

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

- Generate new data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

Data Type	Format	Estimate Volume	Use
Wide-field microscopy images	.czi	<200 GB	Tracking and adhesion experiments
Colony forming units (CFU) counts	.csv	<500 MB	Biofilm cell counts
Flow cytometry	.fcs	<50 GB	aggregate and cell quantification
Confocal microscopy	.czi	<1 TB	Architectural quantification of biofilms
DEM simulation scripts	.py	< 10 MB	in silico biofilm architecture generation
Simulation results	.vtp & .mpacts	<1 TB	quantification and visualization of simulations

3. Ethical and legal issues

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.
NA

3.2. 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).

NA

3.3. Does your research 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?

Yes, our project has the potential to generate a novel set of test that can aid in the risk assessment of antimicrobial resistance development in bacteria. Intellectual property generated in this project will be protected by copyright and patent applications.

3.4. 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 regarding reuse and sharing are in place?

NA

4. Documentation and metadata

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

- All generated in vitro data will be accompanied by a metadata text document specifying the key characteristics of the experiment (e.g. date, strain, substrate, etc.). In addition, the generated files will be named using a format that is easily parsed (e.g. DATE_STRAIN_CONDITION.csv). Each of the different experimental datasets will also be accompanied by a detailed protocol, ensuring reproducibility in future work.
- For the generated *in silico* algorithms, virtual machines (e.g. Docker) will be created. These virtual machines facilitate the reproduction generated algorithms and allow dissemination through version control tools such as the KULEUVEN gitlab. These virtual machines will also contain a ReadMe file on how to build and execute the packaged simulation and analysis algorithms

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

We will use metadata standards for the microscopy, both confocal and wide-field, data (OME-XML) and for the flow cytometry (FCS).

Because there is no metadata standard for the CFU counts and the DEM simulations, we will use in-house standards. For the CFU counts this consists of the key characteristics as described in 4.1. For the simulation data, a metadata file (.csv) is automatically generated consiting of simulation specific (e.g. parameter values, parameter descriptions, etc.) and system specific metadata (system specifications, date & time, software version, etc.).

5. Data storage and backup during the project

5.1. Where will the data be stored?

1. Data generated by the different partners will be saved on a collective sharepoint for duration of the project. In view of the size of the raw microscopy data (~1TB) and raw simulation data (~ 1 TB), this will be exempt from the sharepoint.
2. Additionally, each partner will store their generated data on the central storage facility of their respective research units.
3. Personal copies can be made and kept on personal devices.

5.2. How will the data be backed up?

1. The collective sharepoint is backed up three times per day.
2. The internal back-up of the specific partners is managed according the procedures of their respective research institutions.

5.3. 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, current internal storage suffice for the projected data volume of the project.

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

The total costs of the internal storage at the central servers of the KU Leuven are estimated on ~€3,000/year which will be paid using the allocated project budget.

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

1. Viewing and modification rights in sharepoint are granted based on the involvement in the specific work packages.
2. Critical documents, e.g. reports, presentations,... can be (temporarily) locked by the author(s).
3. Sharepoint provides a changelog for detecting and reverting possible unauthorized changes.
4. The internal storage of the partners provides a back-up for the sharepoint and vice versa.

6. Data preservation after the end of the project

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

All of the generated data will be stored for minimum 10 years after the end of the project.

6.2. Where will these data be archived (= stored for the long term)?

1. The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Due to the size of microscopy data (1 TB), our current storage volume of the archive drive at KU Leuven needs to be increased. The costs for an expansion of 2 TB for 10 years are estimated on € 4,000, and will be covered by the allocated project budget.

7. Data sharing and re-use

7.1. 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 or because of IP potential)?

Yes, the analysis pipeline that predicts the risk of antibiotic resistance development will not be shared, as it can decrease the valorisation potential of this, or follow-up, projects.

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

1. The virtual machines will uploaded to Gitlab and Zenodo under a CC-BY license.
2. Cytometry data will be added as Zenodo/supplementary information to potential publication output.
3. Raw microscopy and simulation data will not be shared due to its size and potential loss of valorization.

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

- In an Open Access repository
- Upon request by mail

7.4. When will the data be made available?

- Upon publication of the research results

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

1. Flow cytometry data will be uploaded in a cvs format in Zenodo as an open access dataset under a CC-BY license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.
2. For the other datasets, access will be considered after request. Only uses for research purposes will be allowed and commercial reuse will be excluded.

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

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

N.A

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

1. An internal workgroup will be responsible for creating and conveying the data documentation guidelines & metadata standards for the project.
2. The PI's carry the end responsibility for the correct documentation of their respective generated data.

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

1. Each partner is responsible for correct data storage, i.e. sharepoint and internal, for the duration of the project.
2. The above mentioned workgroup is responsible for the coordination of the data storage.

8.3. Who will be responsible for ensuring data preservation and sharing?

The coordinator of the project will collect all the data after the project and ensure correct preservation of the data.

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

The end responsibility for updating and implementing the DMP is with the supervisor (promotor).