DMP title

Project Name My plan (FWO DMP) - DMP title

Project Identifier 3E200706

Grant Title 1194222N

Principal Investigator / Researcher Michiel Vanslambrouck

Project Data Contact +32479102706 ; michiel.vanslambrouck@kuleuven.be

Description Tightly controlled cell shape changes underlie cellular motion and self-organization in processes as diverse as wound healing and embryogenesis. Cell shapes are molded for an important part by differences in tension and adhesion between the cells, as well as cytoskeletal protrusions. Much remains unknown about the role and regulation of force generation in embryogenesis. This project investigates a new approach to infer these cellular forces based on microscopy images and then correlate force generation with gene expression and protein localization. After carefully segmenting cell shapes from the images, force inference will be done based on a biophysical model of cell shape and numerical optimization. The use of an explicit cell shape model enables a clearer distinction between the contributions of passive mechanical forces versus local directed force generation. In the project embryos of the nematode C. elegans will be used as a model. The algorithm will yield a detailed sub-cellular force map of the developing embryo. Laser ablation experiments will be performed to validate the inferences. Finally, the generated force maps will be linked to scRNA-Seq datasets and protein expression time-lapses to explore how force generation is controlled in the C. elegans embryo.

Institution KU Leuven

1. General Information

Name applicant

Michiel Vanslambrouck

FWO Project Number & Title

1194222N

Simulation-based Cell Shape Analysis and Force Inference: Development, Validation and Application on C. elegans Embryogenesis

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. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Volume	How created
Microscopy images (raw)	.czi	2TB	confocal fluorecense microscopy of C. elegans embryos, either 3D for time-lapses or 2D for laser ablations
Microscopy images (processed)	.tiff	500GB	adjusted and selected parts of the raw time-lapse images
source code	.py, .ipynb, .java	50MB	source code, either newly created or adapted from research group members
cell meshes	.stl, .vtp	2GB	cell segmentations and simulation output
analysis output	.txt, .png	100MB	analysis logfiles containing metadata and condensed results from force inference
PIV tracing	.lineage	10MB	tracking files for measuring laser ablations, associated with .czi ablation files
Ablation settings	.xslx	10MB	settings used for each ablation measurement
Existing scRNA datasets	count matrices, RPKM	2GB	single cell rna counts measured in several stages of development through scRNA-Seq

Two main flow paths:
3D microscopy -> Cell segmentations -> Simulation-based optimization -> Physical parameters
2D laser ablations -> PIV tracing -> Physical parameters (experimental)

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?

- Yes

The source code of the Mpacts physics engine (https://mpacts.com/) cannot be distributed freely.

Instead, the developers allow a docker image to be made public, so the software can still be run.

4. Documentation and metadata

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

1. The microscopy files automatically include metadata about the (hundreds of) captured properties and settings. Most important for this project are pixel size, frametime and channel information.
2. For laser ablations, the specific settings for every measurement are kept and summarized into a configuration file to perform the analysis.
3. All source code is documented for usage and internal logic. A README overview of how all elements work together **is yet to be made** as I am still developing the pipeline.
4. The image segmentation and simulation-based optimization procedures are all integrated in a consistent folder structure that allows one to see the intermediate data from any step they wish.

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

Microscopy metadata follows the OME-XML structure.

The segmentation pipeline uses a structured XML file to keep track of properties for every cell segmentation.

5. Data storage and backup during the FWO project

Where will the data be stored?

1. The source code is continuously uploaded to GitHub, in a currently private project.
2. Raw microscopy data is kept on both the microscopy PC and shared KUL drives.
3. Processed microscopy data and all analysis files are kept on my computer, backed up to my personal harddrive and uploaded to the pCloud account of the research group.

How is backup of the data provided?

The data on KUL drives and pCloud are automatically backed up.

Source code synchronization to GitHub is done regularly.

Personal external backup system runs daily.

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

pCloud: 4TB limit, can be upgraded

KUL drives: 2TB, can be upgraded

GitHub: 10GB (plenty for source code)

external hdd: 1TB

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

The yearly cost of pCloud is below 100 euro and is covered by the research group.

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

All cloud services are password protected. No sensitive data is handled.

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 can be kept for a peariod of 5 years. The largest files are the raw microscopy and they are not strictly necessary to keep given that the processed microscopy images hold the relevant parts.

Where will the data be archived (= stored for the longer term)?
After this project, all data can be stored on the KUL drives for at least 10 years, conform the KUL RDM policy.
The source code can remain accessible on GitHub (or BitBucket, if needed).

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
The KUL storage options are free of charge.

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)?

- Yes. Specify:

The source code of the Mpacts physics engine is proprietary and will not be shared.
All data is either public to be used already, or generated ourselves.

Which data will be made available after the end of the project?
All results will all be available through Zenodo and publications.

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

- In an Open Access repository
- Upon request by mail

Open access

- Source code (GitHub)
- Segmentations (.vpt files or .stl, Zenodo)
- Intermediate and final results (.txt, Zenodo)

The complete microscopy dataset will be kept private and may be shared on request, as this is a very large dataset that would be inconvenient to host online.

When will the data be made available?

- Immediately after the end of the project
- Upon publication of the research results

Who will be able to access the data and under what conditions?
All results are made available on Zenodo as open access dataset under a CC-BY license.
Als this project does not focus on the segmentation procedure itself, the microscopy time-lapses will be shared on request.

What are the expected costs for data sharing? How will the costs be covered?
GitHub and Zenodo are free of charge

8. Responsibilities
Who will be responsible for data documentation & metadata?
PhD researcher (Michiel Vanslambrouck)

Who will be responsible for data storage & back up during the project?
PhD researcher (Michiel Vanslambrouck)

Who will be responsible for ensuring data preservation and reuse ?
PhD researcher (Michiel Vanslambrouck)

Promotor (Rob Jelier)

Who bears the end responsibility for updating & implementing this DMP?
The PI bears the end responsibility of updating & implementing this DMP.