FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	Francesca Pollet - 0000-0002-4129-7205
Contributor name(s) (+ ORCID) & roles	Jeroen Lammertyn - 0000-0001-8143-6794 - Supervisor
Project number [1] & title	1SHF424N - SURVIVAL OF THE FITTEST ON A CHIP: A MICROFLUIDIC TOOLBOX TO STUDY CELL-CELL COMPETITION.
Funder(s) GrantID [2]	Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)
Affiliation(s)	■ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

Please provide a short project description	Cell-cell competition (CCC) is a defense mechanism of the body to actively remove cells with lower cellular fitness from tissue, which has been repeatedly hypothesized to play a role in the containment or outbreak of a variety of human malignancies (e.g. tumorigenesis, ageing, implantation failure and miscarriages). Unravelling the underlying mechanisms of CCC would therefore offer unprecedented insights into disease progression and, with that, the opportunity to discover new and more effective treatments. Nevertheless, due to a lack of fundamental understanding of this process, progress is significantly hampered. The origin of this information vacuum is two-fold, as current platforms to study the mechanisms driving CCC lack: (i) control over the cells (i.e. cell types, spatial configurations and extracellular environment) and (ii) compatibility with state-of-the-art omics approaches. To address these limitations, we propose DisCCCover, i.e. a microfluidic toolbox to detect, study and influence CCC in vitro. We will combine microfluidics and multiplex cell micropatterning strategies to establish two complementary platforms: (i) DisCCCover-E to elucidate the extrinsic mechanisms driving CCC, and (ii) DisCCCover-I, compatible with the state-of-the-art transcriptomics approaches, to study the intrinsic mechanisms underlying CCC. Finally, we will leverage this generic toolbox to gain unmatched fundamental insights into the role of CCC in aneuploid mosaicism in embryogenesis

2. 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 [3].

				ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*	*ONLY FOR PHYSICAL DATA*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
		☐ Generate new data ☐ Reuse existing data	☐ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Microscopy images/movies	Raw and processed data	New data	Digital	Images	.nd2, .tiff, .png, .jpeg, .avi, .mp4	<5TB	
Microfluidic chip designs	Design files	New data	Digital	Other	.stl, .dwg, .f3D, .cif, .dxf, .cad	<1GB	
MATLAB scripts	Scripts, raw and processed data	New data	Digital	Other	.m, .mat	<100 GB	
Next-generation sequencing data	Sequencing data	New data	Digital	Numerical	.sam, .bam, .bed, .bg, .bw, .vcf(.gc), .bc, .tsv(.gz), .mtx, .loom	> 5 TB	
Observational data	Electronic lab notebook (eLABJournal) notes of	New data	Digital	Textual	.txt	< 1GB	

	observations during experiments						
Fixated-stained cells	Glass slides containing frozen cell samples used for omics analysis / protein stainings	New data	Physical	Other	/	/	An estimated amount of 100 physical samples will be stored in a box in the fridge (5 boxes of 20 glass slides = 1,000 cm^3)

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

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.	/

[3] Add rows for each dataset you want to describe.

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☒ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☐ No Additional information: Human samples will be used: commercially available induced pluripotent stem cells (IPSCs), in which chromosomal instability is induced (by reversine pulse) to obtain aneuploid IPSCS. The ethical approval for these cells will be obtained before they will be used, which is expected in the fourth year of the FWO-SB project.
Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☒ Yes (provide PRET G-number or EC S-number below) ☐ No Additional information: Sequencing data of the above described human cell lines will be generated. The privacy register number for this data will be obtained before the sequencing data is obtained, which is expected in the fourth year of the FWO-SB project.
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.	☒ Yes ☐ No If yes, please comment: If research data have potential for patent filing, scientific staff working on this project will discuss this with the IOF manager in the group and LRD to make sure that data are protected prior to publications.
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 to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:

[4] See Glossary Flemish Standard Data Management Plan

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 to what data they relate and which restrictions will be asserted.	☒ Yes ☐ No If yes, please explain: Due to the generation of IP, restrictions in data sharing might potentially be implemented. This will be followed up through regular meetings with KU Leuven LRD, who will evaluate and protect potential IP generated during the project that could lead to valorization actions. In this case, IP will (i) not be shared, (ii) put under embargo, or (iii) a suitable license will be applied to the data when published (e.g. Creative Commons License).

3. 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). *RDM guidance on documentation and metadata*.	The Biosensors group uses the electronic lab notebook in which a number of predetermined topics have to be described for each experiment (objective, protocol, results, and conclusion). The electronic lab notebook facilitates searching for particular metadata through a search engine. By mimicking the folder structure of the electronic lab notebook in the server-based folder with the experimental data, linking of the metadata to the actual data will be facilitated. As a general rule, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org). When depositing data in a local or public repository, the final dataset will be accompanied by this information in a README.txt document, following the Dublin Core Metadata standard if no other meta-standard is available yet. This file will be located in the top-level directory of the dataset and will also list the contents of the other files and outline the filenaming convention used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse. For each peer-reviewed article, a separate folder will be made on the server, containing the latest Word version and all raw and processed data used in the article. In addition, a separate file will be made in the electronic lab notebook for each article, containing clickable links to all metadata files of data that were used in that particular article, to facilitate tracing back of protocols, results and conclusions.
Will a metadata standard be used to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☒ Yes ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: To guarantee reusable aspect of data, sufficient documentation and methods information will be provided, whereas CC-BY license will be attached to data through data repositories. If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project	
Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☒ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☒ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☐ Other (specify)
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 ☐ No If no, please specify: During, as well as after the project, KU Leuven can provide sufficient storage and backup capacity. A dedicated folder is made for the project to store data files.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	The network drive for the project shared folder and the large volume storage folder are secured by the ICTS service of KU Leuven with a mirror copy. Confidential data can and will be protected with a password (available only for PI Jeroen Lammertyn). Visitors, MSc thesis students and internship students in the groups as well as other unauthorized persons will not have access to the data on the shared folder. Data storage in the cloud will be avoided, unless for temporary use only, e.g., to transfer large files between the researchers involved in the project.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Type 1 server back-end storage with mirror backup for the project shared folder will be used. Using the shared network drive (J-drive) and large volume storage (K-drive) cost respectively 503.66 Euro per TB per year and 104.42 Euro per TB per year. Costs will first be covered by the project consumables budget. If additional storage would be required for sequencing data, my co-promotor Prof. Voet has acquired a large-scale VLIR infrastructure grant to procure high-volume data storage as part of the Leuven Institute of Single-Cell Omics (LISCO) which could be used for this project.

5. 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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☒ Large Volume Storage (longterm for large volumes) ☒ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	• The volume corresponding to dissemination data is expected to be relatively low (<100 GB), and therefore can be seamlessly embedded on the (K:) drive of KU Leuven. The costs (1000 EUR/year) will be covered by other on-going projects at that point in time. • For research data, at current archiving costs of 10 Euro/(TB*year), we estimate a cost of 2000 Euro/year. These costs will be covered by funding acquired by the project PIs in the context of other research projects. • If additional storage would be required for sequencing data, my co-promotor Prof. Voet has acquired a large-scale VLIR infrastructure grant to procure high-volume data storage as part of the Leuven Institute of Single-Cell Omics (LISCO) which could be used for this project.

6. Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☒ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	Only researchers participating in the project will be able to access the data for the duration of the project. As soon as the article associated with the data is ready for publication, the data will be made open through the institutional repositories mentioned in 2.1. The data will be deposited in the institutional repositories: (KU Leuven: Research Data Repository (RDR) Research Data Repository (RDR) - RDR - Research Data Repository (kuleuven.be). Data will be assigned with DOIs to create trustworthy and persistent links for online content.
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 per dataset or data type where appropriate.	☐ Yes, privacy aspects ☒ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☐ No If yes, please specify: Before making data and other research output from the project (e.g. journal articles, book chapters and conference proceedings) openly available, they will be aligned with the project IP strategy to avoid premature disclosure, which can compromise the patent filing application(s).

Where will the data be made available? If already known, please provide a repository per dataset or data type.	☒ KU Leuven RDR ☐ Other data repository (specify) ☐ Other (specify)
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☒ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No

What are the expected costs for data sharing? How will these costs be covered?	A restricted access repository can be implemented on a free tool, such as OneDrive, up to a certain volume. If this volume does not suffice, time-limited storage will be considered, thus limited to the time needed to download the data. The costs associated with data storage will be covered by the budget foreseen in the project agreement.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Francesca Pollet**
Who will manage data storage and backup during the research project?	**Francesca Pollet**
Who will manage data preservation and sharing?	**Jeroen Lammertyn**
Who will update and implement this DMP?	**Francesca Pollet, Jeroen Lammertyn**