FWO DMP Template - Flemish Standard Data Management Plan

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	Marika Caruso 0000-0003-0425-1934
Contributor name(s) (+ ORCID) & roles	/
Project number[1] & title	Elucidating the molecular dynamics of mammary tissue remodeling and tumor initiation using multiplexed biosensors in a novel branched mammary organoid system.
Funder(s) GrantID[2]	1O4423N
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. 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.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	Organoid models are a powerful tool to gain insights into tissue morphogenesis. However, current available mammary organoid culture protocols result in morphologically simple structures that do not resemble the branched morphology of the in vivo mammary gland. To study dynamic mammary gland remodeling at the molecular level, I developed a protocol to obtain complex branched mammary organoids that enable ex vivo optical monitoring of branching and remodeling events. To unravel the molecular signaling pathways driving mammary gland remodeling, I will combine this novel organoid model with multiplexed signaling reporters and timelapse microscopy. I will challenge the system by introducing sporadic loss of BRCA1, an oncogenic driver in the mammary gland, to elucidate the molecular dynamics that drive mammary tumor initiation. Using the dynamic readouts from the multiplexed signaling biosensors in the complex mammary organoids I will build a 4D map of dynamic signaling during mammary gland remodeling and tumor initiation.

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[4].

				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	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data ☒ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☒ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: tiff, lif ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☒ < 50 TB ☐ > 50 TB ☐ NA	

Work package 1 (WP1): develop a 4D map of biosensor signaling dynamics in primary spheric organoids and in branched organoids mimicking estrus cycle, pregnancy and lactation. To develop an in vitro model of estrus cycle, pregnancy and lactation, organoids will be cultured with combinations of growth factors and stained for markers of the different stages by whole mount staining.
The data regarding *in vitro* organoid culture being important timepoints as seeding, induction dates, age and estrous cycle of the mice used to obtain mammary glands, intermediate media replacement/replating time points and staining information will be stored in **Excel file (.xlsx)** organized per each

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

experiment. The raw data regarding imaging of organoids at static timepoints to characterize the different remodeling stages will be stored in the **Leica microscopy original format (.lif)** on the Z Drive and representative images will be stored in **.tiff** format on the LDrive.
The presentation of data in the graph form will be stored in **GraphPad Prism file (.pzfx)** and final versions of graphs will be exported to **.tiff** format. To develop a 4D map of signaling dynamics, in vitro organoids labelled with biosensors will be followed by confocal or multiphoton timelapse imaging (4D). The raw data from this timelapses will be stored in **Leica microscopy original format (.lif)**. The representative images will be stored in **.tiff** format and representative timelapse movies in **.avi** format. All data analysis will be performed in Cell Profiler and ImageJ and stored in **Excel (.xlsx)**. All this data will be digital. For what concerns physical data, whole organ staining and multiplexed staining as well as eventual RNA and DNA samples for organoid stage characterization cannot be stored for 5 years due to sample instability, but the data (.lif and .fastq files and processed analysis) obtained from these samples will be stored on the data server.
Work package 2 (WP2): Develop a 4D map of dynamic signaling in the presence of Brca1-deleted oncogenic cells in branched mammary organoids using multiplexed biosensor labeling. In order to assess the impact of mammary gland remodeling on the spreading of Brca-mutant clones, I will derive organoids from a mouse model in which I can induce Brca deletion together with expression of EGFP to mark mutant cells. I will then apply the protocol developed in WP1 to mimic in these Brca-deleted organoids the estrus cycle, pregnancy and lactation remodeling. In these organoids I will follow Brca deleted cells at different timepoints of the organoid remodeling protocol by confocal imaging. At the endpoint, organoids will be characterized by whole mount staining to assess mutant cell identity. The data regarding *in vitro* organoid culture being important timepoints as seeding, induction dates, age and estrous cycle of the mice used to obtain mammary glands, intermediate media replacement/replating time points and staining information will be stored in **Excel file (.xlsx)** organized per each experiment. The raw data regarding imaging of organoids at static timepoints to count how mutant cell number changes over time in the different remodeling stages will be stored in the **Leica microscopy original format (.lif)** on the Z Drive and representative images will be stored in **.tiff** format on the LDrive. Mutant clone analysis will be performed using **LasX** and **ImageJ,** quantified data will be stored in **Excel (.xlsx) file** divided by timepoints and treatment condition.
The presentation of data in the graph form will be stored in **GraphPad Prism file (.pzfx)** and final versions of graphs will be exported to **.tiff** format. Organoids carrying Brca-mutant cells will be transduced with selected biosensors and biosensor signaling dynamics will be followed over time by confocal or multiphoton timelapse imaging (4D). The raw data from this timelapses will be stored in **Leica microscopy original format (.lif)** on the Z Drive. The representative images will be stored in **.tiff** format and representative timelapse movies in **.avi** format. All data analysis will be performed in Cell Profiler and ImageJ and stored in **Excel (.xlsx)**. Processed data and images will be stored on LDrive.
All this data will be digital. For what concerns physical data, whole organ staining and multiplexed staining as well as eventual RNA and DNA samples for organoid stage characterization cannot be stored for 5 years due to sample instability, but the data (.lif and .fastq files and processed analysis) obtained from these samples will be stored on the data server.
WP3: Perform comparative analyses between the healthy organoids and the organoids with pre-oncogenic cells.

Data collected on signaling biosensor activity in healthy (WP1) and Brca-mutant organoids (WP2) will be used for comparison of signaling activity. All data analysis will be performed in Cell Profiler and ImageJ and stored in **Excel (.xlsx)**. The presentation of data in the graph form will be stored in **GraphPad Prism file (.pzfx)** and final versions of graphs will be exported to **.tiff** format.

WP4***: In vivo* validation of branching morphogenesis and tumor initiation signaling dynamics by intravital microscopy of multiplexed biosensors.**

Based on the dynamic signaling data from the branched organoid models, I will identify the key signaling pathways involved in branching morphogenesis and adult mammary gland remodeling and I will validate these signaling dynamics in vivo, by intraductally injecting the biosensors and performing 4D timelapse intravital imaging by multiphoton microscopy. The raw data from this timelapses will be stored in **Leica microscopy original format (.lif)**. The representative images will be stored in **.tiff** format and representative timelapse movies in **.avi** format. All data analysis will be performed in Cell Profiler and ImageJ and stored in **Excel (.xlsx)**. The presentation of data in the graph form will be stored in **GraphPad Prism file (.pzfx)** and final versions of graphs will be exported to **.tiff** format.

All this data will be digital. After imaging, glands will be derived and used for whole mount staining and imaging. For what concerns physical data, whole organ staining and multiplexed staining cannot be stored for 5 years due to sample instability, but the data (.lif and .fastq files and processed analysis) obtained from these samples will be stored on the data server.

- Estimated upper volume limit of WP1: 100 GB for fixed timepoint imaging of organoids for characterization + 12 TB for biosensor timelapse (3 TB for biosensor, 4-5 biosensors)
- Estimated upper volume limit of WP2: 100 GB for fixed timepoint imaging of organoids for characterization + 12 TB for biosensor timelapse (3 TB for biosensor, 4-5 biosensors)
- Estimated upper volume limit of WP3: 10GB of analysis data
- Estimated upper volume limit of WP4: 18 TB for in vivo biosensor timelapse (6 TB for biosensor, 2-3 biosensors)

GUIDANCE: *DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, …). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.* *EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[5] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.* *EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR,. SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG,. GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.* *DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.* *PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).*	
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.	/
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, please describe these issues further and refer to specific datasets or data types when appropriate.	☐ Yes, human subject data ☒ Yes, animal data ☐ Yes, dual use ☐ No If yes, please describe: Reference to ethical committee approval: Animal experiments will be performed as part of this project. All animal experiments will be performed in the laboratory of Prof. Colinda Scheele and have been approved by the Ethical committee for Animal Experimentation (ECD) at KU Leuven, and are outlined in ECD project P003-2023.

[5] These data are generated by combining multiple existing datasets.

Will you process personal data[6]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☐ Yes ☒ No If yes: - Short description of the kind of personal data that will be used: - Privacy Registry Reference:
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:
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:
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:

[6] See Glossary Flemish Standard Data Management Plan

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).	In order to preserve my data, I will use a licensed E-lab journal to record all my results and protocols in a findable way, as well as the backup files regarding the results will be stored in the L-drive. The raw files acquired through imaging, sequencing and analysis will be saved on the institutional server being Ku Leuven Large Volume Storage drive (backed up every 12h), being at the same time accessible and reusable by staff members granted server access.
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: If no, please specify (where appropriate per dataset or data type) which metadata will be created: Each folder containing a separate experiment will also contain an information in the form of Word (.docx) and Excel (.xlsx) file explaining data methods and all relevant metadata being experimental conditions, genetic models used, all sample identification numbers and computational analysis pipelines. The files with detailed explanation stored at Large Volume Storage drivee will ensure the reusability of the data and the reproducibility of any further data generation.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	The data will be temporarily stored on the expansion drives and a copy of the data will be immediately uploaded to the KU Leuven Large Volume Storage space (L-Drive) and Z Drive for long-term preservation and backup.
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.*[7] *REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.*	Data stored on the KU Leuven L-Drive and Z Drive is managed, maintained, and backed up by KU Leuven IT services. Specifically, mirror copies of the stored data are made immediately upon upload, for safety backup purposes.
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 yes, please specify concisely: If no, please specify:

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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.* [7]	KU Leuven is responsible for the security of the used drives.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The raw data during preservation period will be stored on Ku Leuven Z Drive and already processed images on LDrive. The costs will be covered by the budget of the project lead Prof. Scheele. We expect a volume of raw data of 45 TB. The cost of the Z Drive (sinology) is 16.000 euros for 192TB and lasts 5 years. We expect the cost of the storage of raw data (45TB) to be 3750 euros for 5 years. We expect to have up to 2TB of processed data that will be stored in the Archive drive. The cost of the Archive drive is 5.69 euro per 100GB The cost of storing 2TB is 113.80 euros per year so 569euros over the 5 years. The costs will be covered by the budget of the project lead Prof. Scheele.

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...).	The raw imaging data (.lif files) and analysis (.tiff and .xlsx) will be preserved for a least 5 years on the Ku Leuven sinology (Z Drive). The processed images (.tiff) will be stored on one external hard drive and on KU Leuven L Drive. Generated organoids will be preserved in the -80 °C. Whole organ staining and multiplexed staining as well as RNA and DNA samples cannot be stored for 5 years due to sample instability, but the data (.lif and .fastq files and processed analysis) obtained from these samples will be stored on the data server.
Where will these data be archived (stored and curated for the long-term)?	The raw data will be stored on Z drive (sinology) while processed data will be stored on the KU Leuven L-drive and K-drive (Archive drive)
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The raw data during preservation period will be stored on Ku Leuven Z Drive and already processed images on LDrive. The costs will be covered by the budget of the project lead Prof. Scheele. We expect a volume of raw data of 45 TB. The cost of the Z Drive (sinology) is 16.000 euros for 192TB and lasts 5 years. We expect the cost of the storage of raw data (45TB) to be 3750 euros for 5 years. We expect to have up to 2TB of processed data that will be stored in the Archive drive. The cost of the Archive drive is 5.69 euro per 100GB The cost of storing 2TB is 113.80 euros per year so 569euros over the 5 years. Similar costs apply to preserve the data after the end of the PhD.

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, in an Open Access repository ☒ Yes, in a restricted access repository (after approval, 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.	The raw data as well as unpublished protocols, additional experiments and imaging data will be in principle only accessible to members of Prof. Scheele lab members. The staff and students within VIB–KU Leuven Center for Cancer Biology, as well as within the KU Leuven Department of Oncology will be able to access these data upon reasonable request. Any other user can also request data upon reasonable request .
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:

Where will the data be made available? If already known, please provide a repository per dataset or data type.	/
When will the data be made available? *THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.*	Published data will be made available at the time of publication in case of open access or upon request for other publications. Additional, not-published data will be made available for external users upon request during the post-project trajectory.
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.* *EXAMPLE ANSWER: E.G. "DATA FROM THE PROJECT THAT CAN BE SHARED WILL BE MADE AVAILABLE UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE (CC-BY 4.0), SO THAT USERS HAVE TO GIVE CREDIT TO THE ORIGINAL DATA CREATORS."* [8]	Data from the project that can be shared will be made available under a creative commons attribution license (CC-BY 4.0) so that users have to give credit to the original data creators.

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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 ☐ No If yes:
What are the expected costs for data sharing? How will these costs be covered?	There is no expected costs related to data sharing.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The applicant (Marika Caruso) and the project lead (Prof. Colinda Scheele) will share the responsibility for data documentation and metadata generation/preservation.
Who will manage data storage and backup during the research project?	The applicant (Marika Caruso) will be primarily responsible for data collection, generation and storage. The applicant will also take responsibility for documentation and uploading the data onto the L-Drive storage space. The KU Leuven IT department will be responsible for maintenance and back up of the L-Drive data storage space.
Who will manage data preservation and sharing?	The applicant (Marika Caruso) and the project lead (Prof. Colinda Scheele) will share the responsibility for ensuring data preservation and reuse
Who will update and implement this DMP?	The PI bears the overall responsibility for updating and implementing this DMP.