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	Stefania Pedrotti 0000-0002-4119-7677
Contributor name(s) (+ ORCID) & roles	/
Project number [1] & title	Dissecting the role of chromosomal instability during different stages of cancer evolution
Funder(s) GrantID [2]	11P5J24N
Affiliation(s)	x 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	Chromosomal abnormalities caused by chromosomal instability (CIN) are a hallmark of cancer. Strikingly, the exact role of CIN during tumor evolution is still unclear: is CIN an initiating factor, a facilitator or simply a side effect of the tumorigenic process? Previous studies have been inconclusive and suggested a tumor-promoting, neutral or even a tumor-suppressive role of CIN. In my project I aim to dissect the effect of CIN during different stages of mammary tumor progression, starting from healthy tissue until the metastatic dissemination phase. I will employ a novel conditional CIN mouse model (CiMKi), that enables control over the intensity, timing and tissue-specificity of CIN. This will uniquely allow me to determine for which phases and at what severities CIN promotes tumor evolution, and in which conditions CIN has a less relevant or even negative effect. I will use CiMKi mice in combination with lineage tracing and longitudinal intravital microscopy to identify in real-time the effect of different degrees of CIN on tumor cell behaviour, proliferation, and progression. Image-guided micro-dissection of CIN tumor clones and their micro-environment followed by single-cell RNA sequencing will elucidate the impact of CIN on tumor cell fate. Moreover, it will enable deep phenotyping of the niches in which CIN cells reside. Ultimately, this spatiotemporal characterization of CIN will elucidate the role of different degrees of CIN during tumor initiation and progression.

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
initiation – *microscopy*	Collection of confocal imaging files of the mammary glands whole mounts	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☒ > 5 TB ☐ NA	none
initiation – *sequencing*	scRNAseq of CIN clones within the mammary glands	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext .count	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	RNA will be stored at - 20°C
initiation – *processed/analysed data*	Collection of processed and analysed finite data regarding the initiation part (codes,	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model	.tif .ome-tif .lif .liftext .count .ppt	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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

	graphs, figures, etc.)			☐ Software ☐ Other:	.doc .txt .xlxs .r .pzfx		
carcinoma – *intravital microscopy*	Collection of intravital imaging files	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☒ > 5 TB ☐ NA	
carcinoma – *sequencing*	scRNAseq of CIN clones within the carcinoma	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext .count	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	RNA will be stored at - 20°C
carcinoma – *processed/analysed data*	Collection of processed and analysed finite data regarding the carcinoma part (codes, graphs, figures, etc.)	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext .count .ppt .doc .txt .xlxs	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

					.r .pzfx		
metastasis – *microscopy*	Collection of metastasis imaging files	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
metastasis - *sequencing*	scRNAseq of metastasis at the level of the lungs	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext .count	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	RNA will be stored at - 20°C
metastasis – *processed/analysed data*	Collection of processed and analysed finite data regarding the metastasis part (codes, graphs, figures, etc.)	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.tif .ome-tif .lif .liftext .count .ppt .doc .txt .xlxs .r .pzfx	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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 analyzed 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.	/
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: ECD P170/2020 ☐ Yes, dual use; provide approval number: ☐ No Additional information:
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:
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:

[4] See Glossary Flemish Standard Data Management Plan

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:

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*.	To preserve my data I will keep track of my experiments in a physical lab journal, which will be present in the lab at all time. An accompanying .doc file will also be created to facilitate results deciphering. All the results and protocols will be stored in the L-drive. 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 a file (either word/txt/xlsx) with all data methods and all relevant metadata (experimental conditions, genetic models used, all sample identification numbers and computational analysis pipelines). The files with detailed explanation stored at Large Volume Storage drive 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? *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: Temporary storage will be performed on expansion drives, a copy of the data will always be uploaded to the 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?*	☒ 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:
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*	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?	Raw data will be stored on Ku Leuven Z Drive during preservation period and already processed images on LDrive. The costs will be covered by the budget of the project lead Prof. Scheele. 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 (estimated 100TB) to be 8333 euros for 5 years. We expect to have up to 5TB 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 5TB is 284 euros per year so 1422 euros over the 5 years.

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): 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?	Raw data will be stored on Ku Leuven Z Drive during preservation period and already processed images on LDrive. The costs will be covered by the budget of the project lead Prof. Scheele. 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 (estimated 100TB) to be 8333 euros for 5 years. We expect to have up to 5TB 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 5TB is 284 euros per year so 1422 euros over the 5 years.

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.	Raw data as well as unpublished protocols will be accessible to members of the Prof. Scheele lab. Staff and students within VIB-KU Leuven CCB center as well as the Department of Oncology will be able to access data upon reasonable request and permission from the project lead (Prof. Colinda Scheele). Others interested in the data will have access upon duly motivated request and granted permission.
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?	☐ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify) Published data will be available at time of publication in peer-reviewed journal. For non-open access request will be required. Similarly, request will be needed to access non-published data.
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) – *for the public 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?	No expenses are envision for the sharing of public data

7. Responsibilities

Who will manage data documentation and metadata during the research project?	I (Stefania Pedrotti) will be directly involved in the management of the data documentation and metadata generation/preservation, with the support and shared responsibility of my project lead, Prof. Colinda Scheele.
Who will manage data storage and backup during the research project?	I (Stefania Pedrotti) will be primarily responsible for data collection, generation and storage. Same for the uploading of the data on the appropriate storage (Z- and L- Drive) as well as documentation. 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?	Me (Stefania Pedrotti) 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.