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	**Hanne Peeters & 0000-0001-7419-8812**
Contributor name(s) (+ ORCID) & roles	**Shehab Ismail & 0000-0002-4150-1077** **Peter Dedecker & 0000-0002-1882-2075**
Project number[1] & title	3E221135 Developing 'molecular glues' to target KRAS in cancer
Funder(s) GrantID[2]	1SF2323N
Affiliation(s)	☒ 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	It is reported that ~25% of all cancers are related to a mutation in the RAS oncogenes. The RAS isoform KRAS has the most prevalent mutations, which are linked to more than 85% of pancreatic cancer and a significant percentage of lung and colorectal cancers. Despite many efforts and large investments by biotech and pharma, an adequate treatment for RAS dysfunction remains elusive.Using biochemical, structural, and in-cell screening approaches, and in close collaboration with the Centre for Drug Design and Discovery (CD3), I will develop a small-molecule treatment for RAS dysregulation. This strategy will be based on stabilizing the association of KRAS proteins with the prenyl binding protein PDE6D. Practically, the molecule will act as “molecular glue”, which will prevent PDE6D from shuttling KRAS back towards the plasma membrane, interfering with its signaling activity. Building upon recent results within the research group, I will realize a novel treatment approach that deviates from traditional methods, for KRAS-driven cancers.

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:	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Scans and pictures of agarose gels	Agarose gels to check DNA size for cloning	Generate new data	Digital	Experimental	.jpg	<100GB	NA
Scans and pictures of SDS-PAGE gels	Polyacrylamide gels to check protein sample molecular weights	Generate new data	Digital	Experimental	.tiff and .jpg	<100GB	NA
Platereader data	Fluorescence anisotropy titration and testing data, excitation/emission scans	Generate new data	Digital	Experimental	.xlsx	<100GB	NA
High throughput	HTS results	Generate new data	Digital	Experimental	.xlsx	<1TB	NA

screening data							
Microscopy images	Imaging of (live-cell) and staining experiments for evaluation of compounds with biosensors	Generate new data	Digital	Experimental	.tiff	<5TB	NA
Sanger sequencing data	Checking of cloning experiments and existing plasmids	Generate new data	Digital	Experimental	.ab1, .scf, .txt	<100GB	NA
Plasmid maps	Maps of used plasmid construct	Generate new data/Use existing data	Digital	Software	Files on Benchling platform	1GB	NA
Synchotron data	Data from synchotron data collection (X-ray crystallography) for protein structure determination	Generate new data	Digital	Experimental	.dat	<5TB	NA
Protein purification data	Curves obtained from purification with Akta systems	Generate new data	Digital	Experimental	.xlsx	<100GB	NA

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

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.	Plasmid maps in Benchling files of existing plasmids that are purchased (from Addgene)
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: NA

[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: the project will generate hit compounds to stabilize the interaction of KRAS-PDE6D these hit compounds will be further optimized either by commercial collaboration or spin offs. Data will be generated from cell assays and biochemical experiments e.g. microscopic imaging, plate reader screening data or specialized data from protein structure analysis. Data and metadata file formats include .txt, .tiff, .xlsx,.pdb, .fasta, .fcs,.pzfx
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: Not at the moment, depending on the continuation of the project and results obtained with collaborators, this might change. DMP will be updated then.
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: NA

[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).	A detailed electronic lab notebook will be generated in OneNote for Windows10. Protocols will be available in this ELN, and general protocols will also be available in the lab's Benchling space. For images of agarose- and SDS-PAGE gels, they will be named yyyymmdd_lane-contents. For platereader data, they will be named yyyymmdd_protein_type of experiment with specific concentrations and calculations as well as excitation/emission wavelength, bandwidths, and G factor determination type will be noted down in the .xlsx file. Microscopy data will be named yyyymmdd_protein_type of experiment and specific parameters will be attached in the documentation. All other data will be also named yyyymmdd_protein_type of experiment with summary details and stored in folders specific for the experiment that is carried out
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: Metadata will be recorded by the instrument software acquiring the data and in lab books. The following metadata will be collected: i) name of data creator, ii) date of creation, iii) description of resources used for data creation, iv) sample ID used to identify the data, v) the format of data and vi) where and how the data can be accessed by other researchers.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	Data will be stored on OneDrive for Business on the KU Leuven account of the applicant (2TB storage space), the applicants personal computer and external harddrives.
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.*	Backup procedures are in place in the OneDrive for Business programme. Multiple data storage methods will also ensure backup.
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: 2TB storage space from OneDrive for Business is sufficient currently, but it can be increased to 5TB if motivated. Also 1TB external hard drives are available in the lab. 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]	OneDrive for Business has restricted access and the other storage methods are personal. Also no sensitive personal or patient information will be generated. To access OneDrive for Business online, KUL multifactor authentication is in place.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	OneDrive for Business is free for KUL employees, external harddrive costs (50 euro per harddrive) will be covered with the bench fee of the applicant

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...).	All data is digital and will be retained for at least five years.
Where will these data be archived (stored and curated for the long-term)?	Published data will be made available, other data will be archived on the KU Leuven network drive K and external harddrives will be archived in the promotors/copromotors lab.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Data archived on the KUL network drive K will be retained for an annual fee of 150 euro per TB, the promoters budget will be used to cover these costs

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.	Not applicable
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.	In an open access repository, but specifics are not known yet.

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'.*	Upon publication of research results
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]	We strongly support open access of resulting data, codes, and publications with strict adherence to FAIR-data principles after securing valuable intellectual property (IP). Early and open access to peer-reviewed publications is ensured by dissemination under the green and the gold open access model.
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?	All data will be published with open access, so there will be a large cost to be paid. The costs will be covered using the applicants bench fee

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

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Hanne Peeters (the applicant), supervised by Shehab Ismail (promotor).
Who will manage data storage and backup during the research project?	Hanne Peeters (the applicant), supervised by Shehab Ismail (promotor).
Who will manage data preservation and sharing?	Hanne Peeters (the applicant), supervised by Shehab Ismail (promotor).
Who will update and implement this DMP?	Hanne Peeters (the applicant), supervised by Shehab Ismail (promotor).