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	**Isabel Beets (0000-0002-0968-4360)**
Contributor name(s) (+ ORCID) & roles	**Jan Watteyne (0000-0002-9106-6712) – Postdoctoral researcher / co-PI** **Luca Golinelli – PhD student**
Project number [1] & title	Functional characterization of nematode peptide GPCRs using novel receptor activation sensors
Funder(s) GrantID [2]	G036524N
Affiliation(s)	KU Leuven ROR identifier KU Leuven: 05f950310
Please provide a short project description	G protein-coupled receptors (GPCRs) form the largest group of cell surface receptors that bind signaling molecules. They are central to the control of physiology in all animals and represent the most successful class of drug targets. In nematodes, peptide-binding GPCRs are rapidly gaining traction as key regulators of nematode biology and potential targets for developing novel anthelmintics, because of their crucial roles and broad conservation. However, little is known about their interacting ligands and signaling pathways, hampering further functional studies. In this project, we will develop an *in vivo* platform for the functional characterization of nematode peptide GPCRs in the model *C. elegans* and dissect signaling pathways of widely conserved nematode receptors. In pilot work, we developed a genetically encoded GPCR sensor in *C. elegans* that visualizes peptide-evoked receptor activation. We will further optimize and apply this innovative tool to characterize the peptide ligands of highly conserved promiscuous and orphan receptors in *C. elegans* and parasitic nematodes. We will also develop *in vivo* imaging methods to study their intracellular signaling cascades. We expect our findings will provide mechanistic detail into the nematode peptide-GPCR signaling network, which may catalyze fundamental insights into peptide modulation in other animals. This work will lay a foundation to further investigate the functions of peptide GPCRs in nematodes, including human parasites.

[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.

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.

				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
Images	Confocal and epifluorescence microscopy images of transgenic *C. elegans* worms	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	czi, .tif, .oib	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB ☐ > 5 TB ☐ NA	
Processed data files and data representations	Quantitative experimental data	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.csv, .xlsx, .mat, .RData, .pzfx, .ai, .svg, .jpg, .png, .pdf, .tif, .dat	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Notebooks	Experimental logbook	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model	.xlsx, .docx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Ca 10 notebooks

				☐ Software ☐ Other:			
Software code	Code written in Python, R, and MATLAB programming languages to analyze imaging and oocyte data and perform phylogenetic studies	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	.m, .r, .py	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Transgenic *C. elegans* strains, transformed *E. coli* bacteria, plasmid DNA	Non-pathogenic laboratory organisms frozen for long-term storage + pure plasmid DNA Raw DNA sequences (fi. plasmids and *C. elegans* alleles) stored as text files (MS Word)	☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☒ Other: SnapGene (.dna)	.xlsx, .docx, .dna	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	100-200 transgenic strains and plasmids, each frozen in multiple cryogenic storage vials, divided across -20°C and -80°C freezers

	and/or SnapGene .dna files, and curated in MS Excel						

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.	Transgenic *C. elegans* strains and plasmid DNA: we re-use *C. elegans* mutants and DNA plasmids either from other labs or from community stock centers, i.e. *Caenorhabditis* Genetics Center (https://cgc.umn.edu/), National BioResource Project (https://shigen.nig.ac.jp/c.elegans/faq/index.xhtml) and AddGene (https://www.addgene.org/). These reagents cannot be distributed outside of the recipient's organization but can be obtained by other researchers directly from the original lab/stock centers. For phylogenetic analyses of nematode peptide GPCRs we use genome and/or proteome datasets from the Wormbase parasite database (Howe et al., 2016 – *Mol Biochem Parasit* 10.1016/j.molbiopara.2016.11.005; Howe et al., 2017 – *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkv1217), which all have their unique Bioproject references.
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: Only common and non-pathogenic lab organisms that do not require ethical oversight are used in this project: *E. coli* and related bacteria; *C. elegans* and *A. suum* nematodes

Will you process personal data[3]? 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: This project sets out to functionally characterize parasitic nematode peptide GPCRs, which are promising targets for the development of novel anthelmintics. As such, the findings of this project have the potential for commercial valorization as they can guide the development of novel anthelmintics.
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

[3] See Glossary Flemish Standard Data Management Plan

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.*	Digital files will be organized in folders per research objective and experiment, including a text file (f.i. .txt or .docx) with a clear description of what the data represent and how they were generated. Experimental procedures are documented as digital text files (.docx or .pdf) and in hardcover notebooks. For software code, comments are used to keep all information necessary to understand and reuse the code. Reference to any additional and necessary data/functions are made directly in the code and/or README.txt files bundled together with the software code. Details on samples, including plasmid maps and strain genotypes, are archived in excel files with an overview of their location in frozen stock collections.

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: 1) Metadata about *C. elegans* strains and DNA samples are created and curated manually (in MS Excel and Word) following the community guidelines as published on the Nomenclature section of the community resource (www.wormbase.org). DNA sequences are kept either as text (MS Word, *.txt) and/or SnapGene (*.dna), which are widely used formats that can be opened using various software. 2) Biological imaging data (confocal microscopy and calcium imaging) are stored following the OME (Open Microscopy Environment) standard to encode metadata on microscopy experiments in image files metadata (*.tif, *tiff, *.obi, *.czi). This proprietary microscopy image data and metadata can be read using the Bio-Formats/Open Microscopy Environment (OME) standard, for which plugins for most commonly used image analysis software packages exists, including ImageJ/Fiji and MATLAB. 3) For software code (Python, R, or MATLAB) core calculation functionality is programmed and commented in the source code. Comments are used to keep all information necessary to understand and reuse the code. Reference to any additional and necessary data/functions are made directly in the code and/or README.txt files bundled together with the software code. 4) Digital data are organized in folders per research objective and experiment. Metadata records will be kept for each experimental folder in text/MS Word format, documenting for instance date, time, unique identifier and version of the protocol used to generate the data, deviations from the protocol, unique identifier(s) of physical reagents, treatment(s), and person code. Individual file headers refer to the key experimental factors (e.g. reagent and treatment) 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: in-house NAS system All digital data will be stored locally on an internally backed-up NAS system and on cloud/network storage, such as KU Leuven's OneDrive and Large-Volume Storage. The ICT team of the Biology Department will facilitate the technical infrastructure and authentication to access stored files through KU Leuven's Active Directory (Luna). Hardcover notebooks will be kept personally by all researchers involved during the project, and by the PI after the end of the project. Physical samples will be stored in central stock collections of the research group.

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) Network storages (fee-based) for digital data are hosted in the KU Leuven ICTS data center, with incremental backups on at least a daily basis. All data is mirrored to a second ICTS data center. Physical samples are stored at least in duplicate at two different storage facilities (freezers) in the research group, equipped with emergency power.
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: Network storages for digital data hosted by the KU Leuven ICTS data center provide storage capacity up to 2 TB per active staff member. If this storage capacity does not suffice for our imaging datasets, our in-house NAS system and KU Leuven Large Volume Network storage services will be used to store large-volume datasets, which costs 565 EUR per 5 TB per year. We expect no major changes in pricing of data storage and back-up during the project.

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*	We are not working with personal, confidential or sensitive data but will ensure data security by storing data at secured KU Leuven Network storages and buildings.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Expected costs for data storage and back-up during the project are estimated 3000 EUR, which will be covered by the project's budget.

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 (specify): All digital data are stored centrally on a backed-up NAS-system and on KU Leuven-based and external cloud services (i.e. Large-Volume Network Storage and KU Leuven OneDrive for Business). Physical samples are kept for long-term storage in frozen stock collections (-20°C, -80°C freezers or liquid nitrogen).
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Expected costs for data storage and back-up after the project are estimated at 5000 EUR, which will be covered by research grant budgets.

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: Digital data and samples resulting from this project will be made available after publication of results. Publications resulting from this project will be archived in the KU Leuven Lirias 2.0 repository. Software code will be placed on GitHub.com in an open accessible manner to the public upon completion of the project. Samples, raw and analyzed data, and smaller software scripts can be requested from the principal investigator by mail.
If access is restricted, please specify who will be able to access the data and under what conditions.	NA
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.	☐ KU Leuven RDR ☐ Other data repository (specify) ☒ Other (specify) Digital data and samples resulting from this project will be made available after publication of results. Software code will be placed on GitHub.com in an open accessible manner to the public upon completion of the project. Publications resulting from this project will be archived in the KU Leuven Lirias 2.0 repository. Samples can be requested from the principal investigator by mail.
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) The data will be released under a CC-BY 4.0 license, meaning the data is free to share and adapt given appropriate credit is given and whether changes were made to the data. Software code: GNU General Public License 2 or later (GPL-2.0) You may copy, distribute and modify the software as long as you track changes/dates of in source files and keep all modifications under GPL. You can distribute your application using a GPL library commercially, but you must also disclose the source code.

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?	Expected costs for data sharing encompass publication fees (estimated 6000 EUR), which will be covered by the research project funds.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	All researchers involved in the project are primarily responsible for data documentation and metadata management during the project.
Who will manage data storage and backup during the research project?	All researchers involved in the project are responsible for data storage and back-ups on KU Leuven servers during the project.
Who will manage data preservation and sharing?	The principal investigator and co-PI are responsible for ensuring data preservation and reuse.
Who will update and implement this DMP?	All researchers involved in the project, PI and co-PI share the end responsibility of updating & implementing this DMP.