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	**Dr. Casey Gallagher** ORCID: 0000-0002-9451-8673
Contributor name(s) (+ ORCID) & roles	**Prof. Chris Ulens – Project supervisor and lab head** ORCID: 0000-0002-8202-5281
Project number [1] & title	Structural and functional characterisation of the *Myzus persicae* nicotinic acetylcholine receptors for the development of safer agricultural pesticides (1242724N)
Funder(s) GrantID [2]	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	Nicotinic receptors are ion channels that mediate excitatory neurotransmission. They are the primary target for neonicotinoids, which are a major class of insecticides used globally. However, their lack of selectivity between insects has resulted in the decline of pollinating bee species, which greatly impacts biodiversity and agriculture. This has prompted global bans on neonicotinoids use. There is a great need to develop safer insecticides with greater selectivity for deleterious pests, such as the *Myzus persicae*. This is a polyphagous aphid species which destroys crops. However, little is known about the structure or pharmacology of their receptors, or how current insecticides interact with them. To develop insecticides with greater selectivity for these pests - we require a greater understanding of these receptors and their pharmacology. The project aims to express Myzus receptors in established cell lines using chaperone proteins. This will allow us to explore the functionality of these receptors and efficiently screen novel pesticides. Furthermore, we aim to elucidate the molecular structure of these receptors at atomic resolution, which would represent a major breakthrough in the field. Collectively, this will aid in structure-based drug design and inspire the future development of safer insecticides.

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.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA Digital Data Type	ONLY FOR DIGITAL DATA Digital Data Format	ONLY FOR DIGITAL DATA Digital Data Volume (MB, GB, TB)	ONLY FOR PHYSICAL DATA Physical Volume
Electro-physiology results	Generates electronic traces of raw data from two-electrode voltage clamp electrophysiology (HiClamp) Values are analyzed using excel and GraphPad prism	New New	Digital	Images/ numerical Numerical	.seq (DataMining software) .xlsx (Excel) .pzfx (Prism)	☒ < 5 TB	
CryoEM analysis and final structures	Generates series of different datasets, pertaining to structural determination. All in formats and scripts from programs CryoSPARC and RELION.	New	Digital	Images Numerical Scripts	.mrc (PyMOL/Chimera) .STAR (RELION) .cs (cryoSPARC)	☒ > 5 TB	
Fluorescent microscopy	Generates images from microscope	New	Digital	Images	.tiff	☒ < 1 TB	
Gel images	Generates images from SDS and agarose gels imaged on a gel imager	New	Digital	Images	.tiff	☒ < 1 TB	
FSEC and SEC	Generates curves which are saved in pictural or numerical format, and can be analyzed using software's such as GraphPad prism and excel	New	Digital	Images/ numerical	.tiff .xlsx (Excel) .pzfx (Prism)	☒ < 1 TB	

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.	N/A
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**: *Xenopus laevis* (frog) oocytes are used for electrophysiology experiments. To obtain these cells, a minor surgery is conducted on the frogs. This is conduced within the ECD-project 074/2023 ☐ 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:

---

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

[4] See Glossary Flemish Standard Data Management Plan

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: For the functional screening of compounds - we have established a collaboration with the chemical company BASF for the development of novel insecticides. The compounds generated and tested within this project could therefore have potential commercial use and benefit as agricultural products. Both parties have signed a non-disclosure agreement and we have the freedom to publish any 3D structures of insect nicotinic receptors that we obtain.
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.*	Raw data generated in the lab is recorded in written lab journals – which are stored on site in a locked office, only accessible by lab members. Information is stored chronologically. All experimental procedures and results (Both interpreted and analysed) are recorded digitally. Most numerical data in stored in excel and GraphPad Prism files which are separated for each group of experiments. Summaries of data analysis (Graphs, tables, explanations etc.) and non-numerical information is stored in word files, separated into groups of experiments. Within each file the date and location of raw data for each separate experiment is also included, such that others may be able to retrieve the raw data as required. Upon the establishment of new experimental protocols within the lab – these are stored as word documents on the KULeuven wiki site (wiki.kuleuven.be/xtal) which is only accessible to lab members. This allows those within the lab to easily access and re-create previous methods that have been conducted within the lab. Additionally, when experiments are completed and results are ready to be published in scientific journals, the methods are written in detail – such that they can be reproduced by others in the scientific community.
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 successful, this project will generate 3-dimensional structures of proteins solved through cyrogenic electron microscopy. A prerequisite for publication of this data, is that the atomic coordinates of these structures, as well as the experimental data are made publicly available to the community via the Protein Data Bank (PDB; https://www.rcsb.org/). The metadata in this databank is in the Electron Microscopy Public Image Archive (EMPIAR). At present our laboratory has 54 entries in the Protein Data Bank, from both X-ray crystallography and cryoEM structures resulting from previous projects. 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**: Original copies of raw data will be stored on computers housed within the laboratory, which contain the specific software required to retrieve them. For example, the raw files generated from electrophysiology experiments can only be opened in DataMiner, which is proprietary software only installed on the computer attached to the HiClamp electrophysiology station. Computers containing these software's are located in labs with lockable doors. These computers are automatically backed-up onto the university's storage system. One copy of the raw values and analyzed data will also be held on a personal computer before being copied onto large volume storage and a shared network drive. Images from cryo-EM experiments are stored on hard disks drives of several TB capacity. When not actively in use these disks are stored in a fire-proof safe in the supervisor's office.
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):** My personal computer is backed-up onto an external hard drive every 1-2 months. This contains the large majority of raw and analyzed data. ☐ 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: Currently, there is sufficient storage and back-up capabilities for the project however, this will be expanded for future needs. Once the project reaches a stage where structural data is being collected from electron microscopes and analysed using structural software (CryoSPARC, RELION) – this will create very large data sets in the range of multiple TB per structure. This will require additional external hard-drives to be purchased at that time. These will be purchased using the provided FWO bench-fee.
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](#)	1. All raw data that is generated is stored on site in locked offices/labs that are only accessible to lab members. For physical lab books – once they are not in common use (i.e. are full and not used regularly), they are also stored in a fire-proof file cabinet for long-term storage. 2. All digital data that is stored on-site is stored on computers containing password protection that is only known by members of the lab. 3. Data stored on personal computers requires touch access and/or password protection, of which the owner is the only person who knows these details. 4. Back-ups of data are stored on the university IT network and central servers.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Our current expenses for the university IT back-up storage, including personal Box space, are in the range of 1000-2000 euro per year. This price is likely to fluctuate over the course of the project – and is divided up amongst the budgets of different projects. Additional storage required such as external hard-drives for structural data will be purchased using the bench-free provided by FWO (Up to $4000 euro a year).

	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)*** All data will be preserved fully for 10-year according to the KU Leuven policy, except for some intermediate structural analysis information. Structural determination generates large data sets, possibly within the 10 – 20 TB range. For this reason – once a structure has been solved and is published, some of the intermediate results are often compressed and archived, such that the full data is not stored completely. However, the data is stored in such a way that it can be retrieved and similarly re-analysed if required. Note – the raw data is always stored in completion, as is the final output files.
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 (long-term for large volumes) -** External hard-drives housed on site ☒ **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 primary costs associated with the storage of data pertain to the structural data-sets and intermediate results from data processing being extremely large. This will require large storage capacities including internal and external hard-drives. The costs for these will be shared between the lab and the bench-free provided by FWO (Up to $4000 euro a year).

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: The final endpoint of our studies is primarily the publication of 3-dimensional structure of a ligand-gated ion channel, for which the coordinates and structural data are deposited in a public database, the PDB (Protein Data Bank). In same cases the scientific journal requires us to deposit the raw data files in a public depository, for recent example see https://datadryad.org/stash/dataset/doi:10.5061/dryad.pv4097s
If access is restricted, please specify who will be able to access the data and under what conditions.	Once our structural data are published and deposited in the Protein Data Bank they are accessible to the public without restrictions.
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)** The Protein Data Bank (PDB) or Dryad for large data formats. ☐ 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)** Once deposited in the Protein Data Bank our data are not restricted for usage via a license.
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?	The Protein Data Bank does not currently charge users for depositing data. Publication in open access comes at a significant cost, certainly in high impact journals. We foresee a budget of around 3000 euro or more per publication.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Myself, Dr. Casey Gallagher and my supervisor Prof. Chris Ulens
Who will manage data storage and backup during the research project?	Myself, Dr. Casey Gallagher and my supervisor Prof. Chris Ulens
Who will manage data preservation and sharing?	Myself, Dr. Casey Gallagher and my supervisor Prof. Chris Ulens
Who will update and implement this DMP?	Whilst the project is still in an active state myself, Dr. Casey Gallagher and my supervisor Prof. Chris Ulens will bear the responsibility. However, for long term storage past my employment at KU Leuven, Prof. Chris Ulens bears the end responsibility for updating and implementing this DMP in the long term.