FWO DMP Template - Flemish Standard Data Management Plan

Project: 1164825N

1. General Project Information	
Name Grant Holder & ORCID	**Caterina Travisan (ORCID - 0000-0001-7870-4321)**
Contributor name(s) (+ ORCID) & roles	
Project number [1] & title	**1164825N - Proteolytic Surveillance of signalling molecules in the membrane by SPPL2s: unraveling the mechanisms and their endogenous regulation.**
Funder(s) GrantID [2]	-
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☒ Other: : VIB-KU Leuven Center for Brain & Disease Research 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	Intra Membrane Proteases (IMPs) act as molecular switches in the membrane turning on and off a wide variety of signalling events. Recent data present IMPs as highly dynamic and temporally regulated systems involved in the conformational surveillance of a variety of signalling. Presenilin and Signal Peptide Peptidase Like 2a/b (SPPL2a/b) are aspartyl IMPs that process (by poorly understood mechanisms) different pools of substrates. Unlike the extensively studied Presenilins, SPPL2a/b lack in-depth investigation, with a substrate repertoire that is just emerging but already suggests a link with brain pathophysiology. My research proposal aims at gaining mechanistic understanding of the action and modulation of SPPL2 a/b, in the context of the brain pathophysiology. By applying protein biochemistry and enzymology, nanobody-based technology, proteomics and pathway analysis I plan to discover the endogenous SPPL2b interactome in human-derived neuronal cells and brain tissue, validate functional interactors and substrates, and define their roles in brain pathophysiology using *ad hoc* cellular models . The outcomes will enable to unravel SPPL2s endogenous regulation and proteolytic surveillance of signalling molecules and will provide substantial mechanistic and biological insights into how these proteolytic switches control fundamental processes in physiology and disease.

2. Research Data Summary

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
Human tissue samples	Frozen postmortem brain tissue samples from humans	☐ Generate new data ☒ Reuse existing data	☐ Digital ☒ Physical	Observational			60 grams
Cell lines	Bacteria strains, insect cell lines, mammalian cell lines	☒ Generate new data ☒ Reuse existing data	☐ Digital ☒ Physical	Experimental			100ml
Vectors	Vectors for protein expression in bacteria, mammalian and insect cells	☒ Generate new data ☒ Reuse existing data	☐ Digital ☒ Physical	Experimental			20ml

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
Recombinant proteins and compounds	Human proteins (SPPL2s, newly discovered	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	Experimental			6 liters

	interactors) produced in insect or human cell lines, nanobodies expressed in bacteria						
Mass spectrometry data	Mass Spectrometry datasets	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	Experimental	- Raw data (.raw) -Quantitative tabular data: comma separated value files (.csv) - MS Excel (.xls/.xlsx)	<5 TB	
Digital images	Microscopy pictures, gel scans, graphs, illustrations, figures	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	Experimental	- Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif, .lif (Leica),	<1 TB	

					.czi (Zeiss), .gel (ImageQuant) - Digital images in vector formats: scalable vector graphics (.svg), encapsulated postscript (.eps), Scalable Vector Graphics (.svg), Adobe Illustrator (.ai)		

				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
Derived and compiled dataset 1 – Research documentation	Documentation generated by the research and technical staff or	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	Compiled/ aggregated data	-Text files: Rich Text Format (.rtf), MS Word (.doc/.docx),	<1 TB	

	collected from online sources and from collaborators, including ethical approval documents, laboratory notes and protocols.				Adobe Portable Document Format (.pdf) -Quantitative tabular data: commaseparated value files (.csv), MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb)		
Derived and compiled dataset 2 - Manuscripts	Manuscripts resulting from the project	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	Compiled/ aggregated data	-Text files: Rich Text Format (.rtf), MS Word (.doc/.docx), Adobe Portable Document Format (.pdf)	<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.		We will reuse part of the physical datasets that were generated and stored in laboratory of Lucia Chavez Gutierrez in previous projects, such as human brain samples, bacteria strains, mammalian cell lines, nanobodies and vectors for expression.					

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: S60470 ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☐ No Additional information:
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: We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. Ownership of the data generated belongs to KU Leuven and VIB in accordance with the framework agreement of both institutes. VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such the IP protection does not withhold the research data from being made public. In the case a decision is taken to file a patent application it will be planned so that publications need not be delayed.

[3] 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: 3rd party agreements may restrict dissemination or exploitation of data in the following cases: Data generated with cell lines obtained from repositories must adhere to restrictions on data dissemination specified in the material transfer agreement. Data generated with samples obtained from humans are subjected to patient consent forms and, if applicable, material transfer agreements, and the therein specified restrictions on data dissemination and/or exploitation.
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: If the restriction of data sharing is a consequence of securing Intellectual Property (IP), the researcher involved and the IP team of the TechTransfer office shall make the necessary arrangements in order to maintain the embargo on the public access (dissemination) of research data, at least until the essential steps in securing intellectual property (e.g. the filing of a patent application) have been taken. However, reasonable efforts will be made to avoid delays in publication. If the restriction of data sharing is a consequence of contracts with third parties or about classified data, the researcher involved and the Legal Team of the TechTransfer Office will specify the restrictions in this DMP. If the restriction of data sharing is a consequence of ethical issues, see also the Ethical and legal session of this DMP. Personal data will only be published after de-identification and identifiers will not be published. If despite all efforts it is not possible to protect the identities of subjects even after removing all identifiers, personal data will not be made public.

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*.	Data will be generated following standardised protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database, and published along with the results. Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (eLab Journal) and/or in hard copy lab notebooks that refer to specific datasets. Cryotubes of biological samples (bacterial strains, mammalian cell lines) stored at -80°C will be labelled with a reference number that links to an entry or database. All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below).
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: While specific data types might require particular metadata, as a general rule the metadata are based on a generalized metadata schema such as Dublin Core or DataCite, including the following elements: • Title: free text • Creator: Last name, first name, organization • Date and time reference • Subject: Choice of keywords and classifications • Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc. • Format: Details of the file format, • Resource Type: data set, image, audio, etc. • Identifier: DOI (when applicable) •Access rights: closed access, embargoed access, restricted access, open access.

	For specific datasets, additional metadata will be associated with the data file as appropriate. Specific naming and search procedures will be applied to: - Omics and sequencing data: Digital files, raw sequencing data files mainly (.fastq, .fastq.gz), will be named following an in-house procedure, so that all the name of all files in a given dataset will be in the same format. All names will start with a project 3 letter code (Project Code) and a random 6-alphanumeric character code (unique to each sample given the Project Code), followed by a specific descriptive name of the sample and the technology used (e.g.: 10x). Names will only contain letters, numbers and underscores. - Manuscripts: Metadata information will be submitted alongside the final version of the manuscript, including the names, titles, email addresses, ORCIDs and affiliations of all authors. Upon publication, this metadata information will also be submitted to bibliographic databases such as Medline. All manuscripts will be assigned a unique Digital Object Identifer (DOI) by the publisher. Manuscripts will be given a descriptive title, and will be accompanied by keywords provided by the authors in order to maximize their findability. - Mass spectrometry metadata will be generated at the VIB Proteomics Core facility. A copy of the results (data) will be stored at the VIB-KU Leuven Proteolytic Mechanisms Mediating Neurodegeneration laboratory.

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.*	Digital files are stored on KU Leuven servers Shared network drive (J-drive) and Large Volume Storage (L-drive). Other types of data are stored in different forms as follows; - Tissue samples: tissues are processed and stored locally in the laboratory. All human tissue samples are registered in VIB-UA biobank, in compliance with the Belgian law on human body material (dd 19-12-2008). - Vectors: Vectors will be preserved as purified DNA (in a -20°C freezer). - Cell lines: Cryo preserved cell lines will be stored locally in the laboratory in liquid nitrogen tank. At least two vials per cell line derived from independent freezings will be stored. - Other biological and chemical samples: storage at 4°C and/or as frozen samples as appropriate. - Nucleic acid and protein sequences: All nucleic acid and protein sequences generated during the project will be stored on KU Leuven servers. These servers have RAID 6 or equivalent disk setup for protection of the data. - Proteomics data: Proteomics data generated during the project will be stored on KU Leuven servers. - All other digital data (research documentation and manuscripts) will be stored on KUL servers. - The ELN database is managed by VIB Headquarters. VIB uses eLabJournal. All ELN data are stored, and remain available for (re)use. Sanger sequencing data is stored on the centralized LIMS system., initially in the staging area and later in the archive area.
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 There is sufficient storage and back-up capacity on all KU Leuven servers: - The “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end. - The “J-drive” is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system.

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*	Security of digital datasets in KU Leuven drives: Both the "L-drive" and "J-drive" servers are accessible only by laboratory members and can only be accessed after login (username + password), and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour. Security of the physical datasets that are stored in the laboratory unit are ensured by allowing physical access only to the laboratory members through KU Leuven ID card access.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The total cost of data storage during the project is projected to be max. 1000 € per year. This includes the following costs: The costs of digital data storage are as follows: 174 € /TB per year for the "L-drive" and 519 € /TB per year for the "J-drive". These costs will be covered by the VIB dotation budget assigned to the laboratory of Lucía Chávez-Gutiérrez.

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 at least 5 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): Vectors: vectors are preserved in the form of purified DNA (in -20°C freezer). Cell lines: cell lines are stored locally in the laboratory (liquid nitrogen). Other biological and chemical samples: storage either at 4°C and/or as frozen, as appropriate, locally in the laboratory. The proteomics data will be stored by the VIB proteomics core and as a copy by the VIB-KU Leuven Proteolytic Mechanisms Mediating Neurodegeneration laboratory in the KUL servers. In case data would be archived, we would duplicate our tape and store this in the optimal environment, which has an estimated lifespan of 15-30 years.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The total estimated cost of data storage during 5 years after the end of the project is max. 10'000 euro. This estimation is based on the following costs: Electricity costs for the -80° freezers present in the labs are included in general lab costs. The costs of digital data storage are as follows: 569,2€/5TB/Year for the "L-drive", 519€/TB/Year for the "J-drive. These costs at VIB CBD will be covered by the VIB dotation budget assigned to the laboratory of Lucía Chávez-Gutiérrez.

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.	We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply. Any scientist will be able to access the physical data, upon request, after publication, or before publication upon agreement. In order to respect the patient's privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes.

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: Protein sequences, eg. Nanobody - IP in case of a possible patent application will not be shared before patenting. In order to respect the patient's privacy, clinical samples will only be available to the research and technical staff involved in the project, not to other groups, studies or purposes.
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 will be stored on the E-notebook (eLab Journal). The minimum preservation term of 5 years together with the principle of preservation of data will be applied without restriction to raw data as well as processed data. -Vectors: Relevant published vectors and associated sequences will be sent to the non-profit plasmid repository Addgene (which will take care of vector storage and shipping upon request) or be stored in the host institute and shared upon MTA. -Tissue samples: human tissues will be stored in the laboratory and registered with a Belgian biobank in compliance with the Belgian law on human body material (dd 19/12/2008) -Cell lines: Newly created human cell lines will be stored in liquid nitrogen storage and deposited in the UZ Leven-KU Leuven Biobank. Other cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when used for experiments.

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) Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. For restricted-access repositories we will work with a Data Access Committee (DAC) from VIB, which provides guidance and oversight of controlled-access data and systematically reviews data access requests to the data repository. Only after positive evaluation by the DAC will a user be able to access the deposited data (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and License, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PIs, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.
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, upon publication DOI or other accession number will be assigned to the data.

What are the expected costs for data sharing? How will these costs be covered?	It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible. Data management costs will be covered by the laboratory budget. Additionally, we aim for open access upon publication of the results and the research groups will cover the cost.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Metadata will be documented by the fellowship holder Caterina Travisan at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (eLab Journal) that refer to specific datasets.
Who will manage data storage and backup during the research project?	The fellowship holder – Caterina Travisan will ensure data storage and back up supervised by the PIs, with support from René Custers and Alexander Botzki for the electronic laboratory notebook and from Raf De Coster for the KU Leuven drives.
Who will manage data preservation and sharing?	The PI Lucia Chavez Gutierrez will be responsible for data preservation and sharing, with support from René Custers and Alexander Botzki for the electronic laboratory notebook and from Raf De Coster for the KU Leuven drives.
Who will update and implement this DMP?	The fellowship holder Caterina Travisan will be responsible to update and implement the DMP.