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	**Brajabandhu Pradhan , 0000-0001-5732-6675**
Contributor name(s) (+ ORCID) & roles	**Brajabandhu Pradhan, FWO Junior post-doctoral fellow**
Project number[1] & title	12AWB24N, Structural and morphological studies of heterotypic amyloid fibrils
Funder(s) GrantID[2]	FWO
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	Amyloid aggregates (insoluble fibrils rich in cross-beta-sheet structures) that spread to selective neuronal populations are a hallmark of neurodegenerative diseases and eventually cause neuronal death. We do not know the reasons for the selective vulnerability of neuronal cells, and I will address this problem at the molecular level with this project. Amyloidogenic proteins contain aggregation-prone regions (APR) that drive their assembly into fibrils. Interestingly, the morphology of amyloid fibrils made of the same protein can differ considerably: fibrils of different morphology ("strains") spread to different areas in the brains of patients and cause different diseases. Recent findings of the host laboratory suggest that different fibril morphologies are the result of interaction with other components present in the cellular milieu where the amyloid aggregates form. They demonstrated in two key articles that peptides similar to the APRs in the amyloid-forming protein could interact with disease-causing amyloids and alter their morphology. In short, the interaction of unrelated proteins with amyloidogenic proteins could explain how strains arise. In this project, I will study the mechanism of how polymorphs form at a molecular level. Additionally, I plan to contribute to our understanding of the biological relevance of these interactions by examining the structure of fibrils isolated from more complex environments, namely mammalian cell models and mouse models.

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
Cryo-EM Raw images	Image datasets containing Cryo-EM movies in tif format and their averaged images in mrc format	☒ Generate new data	☒ Digital	☒ Experimental	☒ other: .mrc ☒ other: .tiff	☒ < 50 TB	
Cryo-EM maps	Processed Cryo-EM 3D Electron potential maps in MRC format	☒ Generate new data	☒ Digital	☒ Compiled/ aggregated data.	☒ other : .mrc	☒ < 1 TB	
Protein Structure	3D atomic model constructed based on the 3D EM maps.	☒ Generate new data	☒ Digital	☒ Compiled/ aggregated data	☒ other : .pdb	☒ < 1 GB	
Aggregation kinetics	Monitor amyloid formation using THT flouroscence assay	☒ Generate new data	☒ Digital	☒ Experimental	☒ .csv ☒ .pdf ☒ .xml ☒ .tab ☒ .csv ☒ .pdf ☒ .txt	☒ < 1GB	
LC-MS, SEC, and HPLC data	For quality control of peptides and protein sample to be used for	☒ Generate new data	☒ Digital	☒ Experimental	☒ .csv ☒ .pdf ☒ .xml ☒ .tab ☒ .csv ☒ .pdf ☒ .txt	☒ < 1GB	

	various experiments							
AFM images	Morphological analysis of Abeta40/42 fibrils form by heterotypic interaction	☒ Generate new data	☒ Digital	☒ Experimental	☒ other : .tif ☒ other : .png	☒ < 100 GB		
Transmission electron microscopy images	Morphological analysis of Abeta40/42 fibrils form by heterotypic interaction	☒ Generate new data	☒ Digital	☒ Experimental	☒ .tif ☒.png	☒ < 1 TB		
Synthesized APR peptides	Synthetic peptides used for coaggregation with Abeta40/42	☒ Generate new data ☒ Reuse existing data	☒ Physical				15-ml vials	

[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.	https://doi.org/10.15252/embj.2021108591
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:

[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 aims at determining high-resolution structures of amyloid polymorphs formed by disease-associated amyloidogenic proteins. Such structures are of high value as they may be the basis for design and development of specific binders that could modulate activity or toxicity of such amyloids. Therefore, we do hope that the proposed work will lead to tech transfer and valorisation of the research data. VIB and KU Leuven 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. Further research beyond the scope of this project may be necessary for developing a strong IP portfolio.
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:

[6] See Glossary Flemish Standard Data Management Plan

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).	Metadata will be documented by the researcher at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) and/or in hard copy lab notebooks that refer to specific datasets. All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below). The data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database and published along with the results.

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: The following metadata standards will be used for certain datasets. • Protein structures will be saved in Protein Data Bank Format (PDB) (https://fairsharing.org/FAIRsharing.9y4cqw) • Cryo-EM raw images and 3D maps will be stored in MRC format. • For instrument-specific datasets, additional metadata will be associated with the data file as appropriate. • For other datasets, the metadata will include 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. The final dataset will be accompanied by a README.txt document. This file will be located in the top-level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse. 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?	Digital files will be stored either on KU Leuven servers or in shared laboratory folders of an off-site online backup service. The researcher working on the project will have copies of the data files as well as of the derived and compiled data stored on their personal computers. The Switch Lab has a professional subscription to an off-site online backup service with unlimited space, version control and roll-back capability, which will be used for storage during the project and after. There is a secondary on-campus physical backup of the online storage which synchronizes with the online content with a one-day delay. Cryo-EM raw data will be eventually uploaded to publicly accessible data storage servers such as EMPIAR. Protein structure files will be submitted to RCSB's PDB database and the EM maps will be submitted to EMDB.
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.*	The Switch Lab has a professional subscription to an off-site online backup service with unlimited space, version control and roll-back capability, which will be used for storage during the project and after. There is a secondary on-campus physical backup of the online storage which synchronizes with the online content with a one-day delay.
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: The Switch Lab has a professional subscription to an off-site online backup service with unlimited space, which will be used for storage during the project and after. 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]	All notebooks and physical data are stored in the labs. Entry to the lab requires ID-card and key. Access to the digital data is u-number and password controlled.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Data storage and backup costs are included in general lab costs. The Switch Lab has a yearly subscription to an off-site online backup service paid from the general budget of the laboratory. The yearly cost of the service is 5500 Euros. This cost includes unlimited data storage, not only the data belonging to the present project. Electricity costs for the -80° and -20° freezers and refrigerators present in the labs as well as the cost of liquid nitrogen cryostorage are included in general lab costs.

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...).	The minimum preservation term of 5 years after the end of the project will be applied to all datasets.

Where will these data be archived (stored and curated for the long-term)?	For the datasets that will be made openly accessible, we will use, whenever possible, the existing platforms that support FAIR data sharing (www.fairsharing.org), at the latest at the time of publication. For all other datasets, long term storage will be ensured as follows: -Digital datasets will be stored on storage space of an online data-backup service. Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Electricity costs for the -80° and -20° freezers and refrigerators present in the labs as well as for in liquid nitrogen cryo-storage are included in general lab costs. The cost of the laboratory's professional subscription to the online data backup service is 5500 Euros per year (27 500 Euros for 5 years). This cost includes unlimited data storage, not only the data belonging to the present project. Data storage and backup costs are included in general lab 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.	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.
	• The data will be shared upon request by mail.
	• Possible ways of sharing the generated data:
	  • Transmission Electron microscopy images: Image Data Resource (http://idr.openmicroscopy.org/about/)
	  • Protein structure https://www.rcsb.org/
	  • Cryo-EM raw images https://www.ebi.ac.uk/empiar/
	  • Cryo-EM maps https://www.ebi.ac.uk/emdb/
	  • manuscripts: bioRxiv (https://www.biorxiv.org/)
	  • other digital data: Zenodo data repository (https://zenodo.org/) or KU Leuven RDR: (www.kuleuven.be/rdm/en/rdr)

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** Generally, research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed - or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

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]	Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse and will be conform to community norms. These repositories clearly describe their conditions of use (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 Licence, 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 PI, 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 re-use 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 ☐ No If yes: Identifier not available yet
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. The receiving party will pay for sharing physical data.

7. Responsibilities

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

Who will manage data documentation and metadata during the research project?	Metadata will be documented by the researcher at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) that refer to specific datasets.
Who will manage data storage and backup during the research project?	The research and technical staff will ensure data storage and back up, with support from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives
Who will manage data preservation and sharing?	The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.
Who will update and implement this DMP?	The researcher and the PI are ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.