1223924N_DMP

A Data Management Plan created using DMPonline.be

Creator: Katerina Konstantoulea

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1223924N

ID: 206526

Start date: 01-10-2023

End date: 30-09-2026

Project abstract:

It is still unclear why amyloid aggregation is toxic in neurodegenerative diseases. Previously, I showed that proteins with sequence segments homologous to Aβ aggregation-prone regions can affect Aβ aggregation and modify fibril morphology, a mechanism that could help explain regional vulnerability to protein aggregation. Here, I will study this co-aggregation and co-deposition mechanism in a more disease-relevant cellular model that also considers the contribution of microglia to plaque formation, with a focus on proteins that are disease-relevant because of their function, and tissue- and cell-specific expression patterns. This research will dive deeper into heterotypic aggregation and will unravel important aspects of Aβ toxicity by mapping sequence-specific interactions between disease amyloids and the brain proteome.

Last modified: 25-04-2024

1223924N_DMP
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Experimental Data: Omics data, Biophysics data, Spectroscopy data, AFM images, Other digital images (microscopy, blots, Graphs)
Text data: Manuscripts, Journal figures, Research documentation, algorithms, science communication articles
Canonical data: Nucleic Acid sequences, Protein and peptide sequences

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Designation of responsible person (If already designated, please fill in his/her name.)
2. Storage capacity/repository
    - during the research
    - after the research

Alexander Botzki is responsible for E-Notebook, Frederic Rousseau and Hannah Wilkinson are responsible for other datatypes.
EM maps will be deposited to EMDB, Atomic models will be deposited to PDB, Raw Cryo-EM images will be deposited to EMPIAR
Research documentation and processed data will be stored in E-Notebook and in DropBox and on KU Leuven servers, with regular backups.
Manuscripts will be published and archived in public repositories. All canonical data will be added as supplementary files in manuscripts as appropriate.
Nucleic acids, protein samples, and bacterial strains will be stored at -80°C. All types of data will be preserved for at least 5 years.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

No plan to deviate, all resources will be preserved for a minimum term of 5 years after the end of the project.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

n/a

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

n/a

1223924N_DMP
FWO DMP (Flemish Standard DMP)

1. 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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Human proteome	Human proteome assembly of all proteins present in humans to search for modifier candidates	*Existing data (UniProt)*	*Digital data*	*Compiled*	*Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;*	*<1gb*	

Search for homologues peptides	search the proteome for APRs similar to those in Ab42	*Existing data (Konstantoulea EMBO 2021)*	*Digital data*	*Compiled*	*Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Mark-up Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;*	*<1gb*	
List of hits 1	APR hits for Ab42 (up to 2 mutations per 6 AA stretch)	*New data*	*Digital*	*Compiled*	*Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;*	*<100MB*	
Annotation of proteins	Annotation of proteins with datasets from brain and AD (Human Brain atlas, GWAS, Disgenet)	*Existing data*	*Digital*	*Compiled*	*Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;*	*<100mb*	
THP1 cells TIB-202	cells for amyloid generation	*Existing data*	*Physical*	*Experimental data*			10 vials, over 2million cells, used as needed
Abeta biosensor cell line	reporter cell line for high content imaging cellular assays	*Existing data*	*Physical*	*Experimental data*			10 vials, over 2million cells, used as needed

SH-SY5Y cells, CRL-2266	cells to test for toxicity of aggregates	*Existing data*	*Physical*	*Experimental data*			10 vials, over 2million cells, used as needed
Sequencing data1	Sequencing data of THP1 treated with recombinant Ab42	New data	Digital	Experimental	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<1TB	
Sequencing data 2	Sequencing data of THP1 treated with synthetic Ab42	New data	Digital	Experimental	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<1TB	
Proteomics	Proteomics of isolated plaques or/and cells	New data	Digital	Experimental	text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<1TB	

Metabolomics data	Metabolomics data of THP1 with ab42 (synthetic or recombinant)	New data	Digital	Experimental	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<1GB	
Cytoscape networks	Building of cytoscape networks for identifying the most interesting candidates	New data	Digital	Compiled	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<1GB	
List of hits 2	Protein list after filtering	New data	Digital	Compiled	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<100MB	

pFTAA staining THP1	aggregation deposition of abeta in THP1 cultures	New data	Digital	Experimental	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<1gb	
CellTiter Blue	measure metabolic activity of the cells	New data	Digital	Experimental	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<100MB	
Plasmids of candidate proteins	TWIST plasmids for candidate proteins	New data	Physical	Experimental			Glycerol stocks and DNA store frozen
Viral particles	Viral particles for candidate proteins	New data	Physical	Experimental			Stored frozen
Microscope images from high-content screening	determine the effect of expressed modifiers on induction of abeta aggregation	New data	Digital	Experimental	Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif;	<100gb	

Transmission electron microscopy images	morphological analysis of Ab42 aggregates formed in cells	New data	Digital	Experimental	Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif;	<100gb	
AFM images	morphological analysis of Ab42 aggregates formed in cells	New data	Digital	Experimental	Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif;	<100gb	
Screening data	Screening data of THP1 cells	New data	Digital	Experimental	Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif;	<1tb	
Screening data	Screening data of HEK biosensor	New data	Digital	Experimental	Digital images in raster formats: uncompressed TIFF (.tif/.tiff), JPEG (.jpg), JPEG 2000 (.jp2), Adobe Portable Document Format (.pdf), bitmap (.bmp), .gif;	<1TB	

Manuscript	Manuscript	New data	Derived_and_compiled_data	Manuscripts	Text files: Rich Text Format (.rtf), plain text data (Unicode, .txt), MS Word (.doc/.docx), eXtensible Markup Language (.xml), Adobe Portable Document Format (.pdf), LaTex (.tex) format;	<100GB	

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:

Existing data from Uniprot, Human protein atlas, GWAS catalog, Disgenet and previous publications will be used.
https://www.uniprot.org
https://www.disgenet.org
https://www.proteinatlas.org
https://www.ebi.ac.uk/gwas/
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)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

n/a

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

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

The goal of the project is to find proteins that interfere with amyloid formation Ab42. A protein relevant for Alzheimer disease and modifying proteins that promote amyloid formation and those that slow down amyloid formation would be of great interest as potential therapeutic targets. 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 in the comment section to what data they relate and what restrictions are in place.

- No

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 in the comment section to what data they relate and which restrictions will be asserted.

- No

2. 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 research and technical staff 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 (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The following metadata standards will be used for certain datasets

- Nucleotide sequence files (vectors and sequencing) : GenBank Sequence Format (https://fairsharing.org/FAIRsharing.rg2vmt)
- Protein structures will be saved in Protein Data Bank Format (PDB) (https://fairsharing.org/FAIRsharing.9y4cqw)
- For sharing computer code, we use the Zenodo format (https://zenodo.org/)
- 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.

3. 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 researchers 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.

Algorithms, scripts and software: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd).

The screening core has a database system in place to handle the data stream from the high content imaging screen, including archiving facilities and will store the data during the project. Representative images and the quantitation of the images will be transferred to the Switch laboratory storage for long term storage.

Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (-80°C). All published vectors and the associated sequences will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request.

Cell lines: Newly created cell lines will be stored locally in the laboratory in liquid nitrogen storage and will be deposited in the UZ Leuven-KU Leuven Biobank.

Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.

How will the data be backed up?

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

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.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

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.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All digitally generated data would be retained for at least five years.

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. -Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacterial glycerol stock (-80°C). -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 cryostorage 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.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in a restricted access repository (after approval, institutional access only, …)

Data will be available for reuse after publication of findings.

If access is restricted, please specify who will be able to access the data and under what conditions.

Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. All research outputs supporting publications will be made openly accessible. Depending on their nature, some data may be made available prior to publication, either on an individual basis to interested researchers and/or potential new collaborators, or publicly via repositories (e.g. negative data). 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. Physical data (e.g. cell lines) will be distributed to other parties if requested.

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 in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Upon publication of the 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.

When will the data be made available?

Only after publication

Which data usage licenses are you going to provide? If none, please explain why.

License will be provided upon publishing of the data

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

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 (e.g. cell lines).

6. Responsibilities

Who will manage data documentation and metadata during the research project?

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 (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?

Konstantoulea Aikaterini