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	**Ben Sanchez** (0000-0002-6080-3476)
Contributor name(s) (+ ORCID) & roles	**Johan van Lint** (0000-0002-0275-571X) *Supervisor* **Arnout Voet** (0000-0002-3329-2703) *Co-supervisor* **Hideaki Mizuno** (0000-0002-6983-5255) *Co-supervisor*
Project number [1] & title	Development of isoform specific PKD2-14-3-3 molecular glues to treat angiolipoma, multiple myeloma, glioblastoma and pancreatic cancer.
Funder(s) GrantID [2]	1S37225N
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	In this project, we will develop a biosensor toolbox capable to examine the activity and translocation of protein kinase D2 (PKD2) in live cells and the role of 14-3-3 proteins therein. Using this toolbox, we will be able to investigate the effect of PKD2 mutations on 14-3-3 interaction, enabling us to pin-point the precise interaction site of PKD2 with 14-3-3. Later on, we will solve the crystal structure of this PKD2-14-3-3 interaction site in complex with 14-3-3, which can then be used as a basis for in silico design of molecules that glue these regions together and inhibit PKD2 in an isoform specific way.

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

2. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data [3].

				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
Protocols & SOPs (Standard Operating Procedures)		☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.pdf .docx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	< 50 pages
DNA	Purchased DNA oligos (IDT), primers (IDT), addgene constructs and own made constructs	☒ Generate new data ☒ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☒ NA	+-200 1.5ml tubes
Purified proteins		☒ Generate new data ☐ Reuse existing	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB	+-200 1.5ml tubes

[3]

		data		☐ Numerical ☐ Textual ☐ Model		☐ < 5 TB ☐ > 5 TB ☒ NA	
Glycerol stocks	Glycerol stocks of successfully made constructs in DH5alpha and corresponding BL21 cells for protein production	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☒ NA	+-200 1.5ml tubes
DNA & AA sequences	DNA and AA sequences will be kept in benchling	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.fasta .txt .pdf	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
DNA sequencing results		☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☒ Other:	.fasta .txt .abi	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Protein purifications	This dataset will contain pictures of annotated SDS-PAGE	☒ Generate new data ☐ Reuse existing	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound	.png .jpg .txt	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB	

	gels (containing samples of each purification step), western blots, AEX/SEC chromatograms (screenshots and CSV/txt-files) and absorbance spectra of the corresponding purified proteins (260nm/280nmratio, pdf files)	data		☒ Numerical ☒ Textual ☐ Model ☐ Software ☒ Other:	.csv .pdf	☐ < 5 TB ☐ > 5 TB ☐ NA	
Bio-layer interferometry (BLI) interaction experiments	This dataset will contain CSV files and screenshots (from the sartorius BLI-analyzing program) of the performed protein-protein interaction experiments	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.png .jpg .txt .csv	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Anisotropy polarisation	This dataset will contain CSV files and screenshots (from the TECAN-spark analyzing program) of the performed protein-protein interaction experiments using anisotropy	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.png .jpg .txt .csv	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

				☐ Model ☐ Software ☐ Other:		☐ NA	

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	**Protocols & SOPs :** Protocols delivered in frequently used kits (GeneJET plasmid Miniprep Kit, etc.) or existing standard lab protocols. **DNA, DNA & AA sequences:** Plasmids containing the sequences of the PKDs, already generated in house will be used as templates. Some of these constructs will be directly used for protein purifications.
Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information: Not applicable
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: Not applicable

[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: **BLI/anisotropy data** of compounds stabilizing our protein-protein interaction may be used for further valorization. In addition, **co-crystals** of our protein complex with compounds may also be used for further valorization.
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: Not applicable
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: Not applicable

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*.	Each experiment will be documented in an electronic lab notebook (which will be available in benchling) by the performing scientist following standard operating procedures (SOPs), which will be or have been written down. Each lab notebook has a table of summary containing the background/ rationale with the objective, protocols and samples used, results (referencing digital data) and conclusion. The raw data along with the processed data of each experiment will be sorted per experiment type and stored in document folders on the personal laptop of the performing scientist. The data contains a README.txt file explaining the design/protocol, analysis methods, results, labels used and references to the electronic lab notebook. Metadata will link the data files, lab samples and experimental notes (including descriptions of equipment, setting and used experimental settings), allowing other researchers to follow all the steps performed during data acquisition and processing. For the various forms of publication that require a combination of data from multiple experiments, new folders will be created for the processed data with links to the respective raw data. The folder will be created per publication.
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: When depositing data in a local or public repository, the final dataset will be accompanied with a README.txt file containing all relevant information, following the Dublin Core Metadata standard *(if no other meta-standard is available yet)*. 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.

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: NAS (Network-Attached Storage)
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 (KULeuven OneDrive) ☒ Personal back-ups I make (on separate drives, located at different locations) ☒ Other (The data from the Voet lab (crystallography) is saved on a double backup NAS (Synology), which is secure and requires authorization before access).
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 The KU Leuven ICTS provide sufficient storage and archival capacity during the PhD project. Additionally, the Voet lab has 128TB of space on the NAS system, which will cover all crystallography generated data for the next 4 years.

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*	The data is secured by the 2-factor authentication system of the KU Leuven, while the Synology NAS system is only accessible through authorization. In addition, the personal network drive is encrypted. Confidential data can and will be protected with a password that is only available for the (co-)promotors. Visitors, master thesis and internship students as well as other unauthorized persons will not have access to the data.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	KuLeuven onedrive (up to 250GB) is free for staff members. Separate back-up drives will be bought using the benchfee. The LBMD NAS system is already present and is available for all LBMD members.

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) The raw data files from the X-ray diffraction will not be kept for 10 years, if it has been confirmed that data is not relevant, as the cost of storing the large amount of raw imaging data will be unaffordable.

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): After the end of the project, the research data will be stored at the large-volume-storage system of the KULeuven. Dissemination data containing files corresponding to papers and presentations, will be stored on the PCs of involved PIs and backed-up daily on the departmental server for long term storage. Analysis code will be stored on ICTS-hosted code repositories.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	It has been estimated that +- 5 TB storage will be collected in the duration of the project. Currently the ICTS storage pricing is 95 Euro/TB/year, we estimated storage costs at around 4750 Euro (95 * 5TB * 10years) to cover the duration of the project and long-term storage of 10 years. This will be covered by the funding acquired by the project PIs in the context of other research projects.

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: All information, including negative results, will be released at the end of the project to ensure future researchers and AI algorithms can learn from it and prevent repetition. In case we identify promising active compounds, we will consult LRD and may opt to keep results secret while preparing patent applications in order to protect our intellectual property. The data will only be released when the project is finished as of the guidelines of LRD. In case the data is valorizable, it may be kept secret, however, we will strive for the least interesting compounds to be released as part of my PhD while keeping the most promising and valorizable compounds confidential.
If access is restricted, please specify who will be able to access the data and under what conditions.	The project collaborators will be authorized to have access to all obtained digital and physical data after the project. For researchers outside of the consortium, the open data can be made available upon e-mail request and under the condition that the users agree to give proper credit through co-authorship on their papers that used the data. Data usage for commercial purposes will require a license or an equivalent arrangement.
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: Structural data of proteins and compounds, as well as BLI and anisotropy data fall under IP protection.

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)
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)
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?	A restricted access repository can also be implemented on free tools, such as Dropbox & WeTransfer, up to a certain volume. If the storage volume is insufficient, time-limited storage systems will be considered to download the data. Upon publications, relevant datasets will be made public via relevant databases (e.g. PDB).

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The researcher will be responsible for data documentation & metadata.
Who will manage data storage and backup during the research project?	The researcher will be responsible for data storage & back up during the project.
Who will manage data preservation and sharing?	During the course of this project, the researcher, while after completion of this project, the PI will bear the end responsibility of ensuring data preservation and reuse.
Who will update and implement this DMP?	The PI bears the end responsibility of updating & implementing this DMP.