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	**Igor Steve BECKERS** (ORCID 0000-0002-4363-4893)
Contributor name(s) (+ ORCID) & roles	**Dirk De Vos** (supervisor) (ORCID 0000-0003-0490-9652)
Project number [1] & title	1291724N
Funder(s) GrantID [2]	CROSS-COUPLING OF DIVERSIFIED C(SP3) FRAGMENTS WITH (HETERO)ARYL HALIDES BY UPGRADING PD CATALYSIS WITH THE ORGANOCATALYTIC ACTIVATION OF AMINO ACIDS
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	The success of Pd-catalyzed cross-coupling reactions in drug discovery has led to a large share of one- and two-dimensional drug scaffolds among pharmaceutical compounds. This leaves the three-dimensional medicinal compounds underexplored. In this project, we will tackle this issue by enabling the Pd-catalyzed cross-coupling of pharmaceutically relevant (hetero)aryl halides with amino acid-derived C(sp3) fragments. An innovative approach will be investigated: by combining the Pd catalysis with the organocatalytic decarboxylation or deprotonation of amino acids, potent nucleophiles will be generated for C(sp2)-C(sp3) coupling in a convergent manner. The optimal combination of organocatalyst and Pd catalyst will be discovered using a research methodology that goes beyond the conventional screening of catalysts. Aided by fundamental understanding of the reaction mechanism (via kinetic experiments, computational modelling, advanced XAS on the Pd K-edge and NMR spectroscopy with 13C and 15N labelled amino acids), structure-activity relationships of the organocatalyst will be obtained via the thoughtful variation of its molecular structure.

[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
GC/HPLC/ NMR data	Data acquired upon analysis of chemical samples using GC-FID, HPLC and NMR instrumentation respectively	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.gcd .txt .csv	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Output of DFT calculations	Geometry optimizations and single-point energy calculations of intermediates and transition states	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.xyz .txt .out	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Data analysis and publication drafts	Quantification of results and interpretation thereof in text	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical	.docx .xlsx .pptx	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB	

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

	documents, spreadsheets and presentations			☒ Textual ☐ Model ☐ Software ☐ Other:		☐ > 5 TB ☐ 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.	NA
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:
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:

[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: The project is dedicated to the fundamental investigation of chemical reactivity in homogeneous catalysis (GC/HPLC/NMR data, output of DFT calculations and data analysis/publication drafts) in the first years. The fundamental understanding could enable a reliable assessment of the future applicability in an industrial context at a final stage (*i.e.* last 3-6 months of the project).
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

3. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded). *RDM guidance on documentation and metadata*.	Details to reproduce reaction procedures are described in physical lab notes. Intermediate results are regularly interpreted and transferred to written reports or presentations for internal dissemination. Upon publication of data, procedures are added to the supporting information, that are described in sufficient detail to enable reproduction of the generated 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: If no, please specify (where appropriate per dataset or data type) which metadata will be created: The data will be stored and named in a consistent manner with unambiguous identifiers. Descriptions of experiments and materials are kept in physical lab notes for each identifier. The data will be processed and summarized in a comprehensive excel worksheet with references to the identifiers. The resulting summaries will culminate in presentations and manuscripts for publication.

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 ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ Digital Vault ☒ Other: Laptop hard-drive with regular back-up
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) Research data will be stored automatically on a OneDrive cloud storage for back-up, and will regularly be saved on an external hard drive
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 cloud storage provides a storage space up to 100 GB for each user, while the project is expected to generate less than 100 GB.
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*	Data storage costs are included in an internal service contract with the KU Leuven IT support service (SET-IT).

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	No additional costs are expected for the storage of data. In case additional costs do arise, they will be covered by the project budget or reserve funds.

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)
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): Upon termination of the contract, the data will be transferred and stored on an external hard drive (Samsung Portable SSD T5 1 TB), managed by Annelies Van Vlasselaer
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The high capacity of the available external hard drive (1 TB) enables the preservation of data from multiple terminated or finished projects. Currently, an average of 15 GB is used for the finished projects of each user, which allows to divide its cost over approximately 60 users. Given the cost of the available hard drive of 120 EUR, the expected costs are negligible. The involved IT-expenses are included in the project's consumable expenses or covered by reserve funds.

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.	Upon publication of the research results, the associated datasets will be made available upon reasonable request.
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: Unpublished data (GC/HPLC/NMR data, output of DFT calculations, data analysis and publications drafts) with prospect on publication will only be shared internally to maximize chances on impactful publication

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): Upon reasonable request via e-mail
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) Only uses for research purposes will be allowed and commercial reuse will be excluded.
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 All relevant data will be provided upon publication, either in the main text or in supporting information
What are the expected costs for data sharing? How will these costs be covered?	No additional costs are expected as the data can be shared via the already budgeted storage cloud from Box. If additional costs would occur they will be covered from reserve funds.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Igor Beckers (grant holder)
Who will manage data storage and backup during the research project?	Igor Beckers (grant holder) and Annelies Vanvlasselaer
Who will manage data preservation and sharing?	Annelies Vanvlasselaer and Dirk De Vos (supervisor)
Who will update and implement this DMP?	Igor Beckers (grant holder)