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	**Daniel Escudero 0000-0002-1777-8578**
Contributor name(s) (+ ORCID) & roles	
Project number [1] & title	G022324N - NEW TOOLS TO MODEL PHOTOCHEMISTRY IN THE LONG TIMESCALE (PHOTOLONG)
Funder(s) GrantID [2]	
Affiliation(s)	☑ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: 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	PhotoLong is a fundamental research project at the frontier between theoretical/computational chemistry and photochemistry. It aims to develop, implement and apply theoretical tools expanding the state-of-the-art modelling toolset for photochemistry, with an emphasis on tools enabling the prediction of photochemical properties and photoreactivity at longer timescales, i.e., beyond the picosecond scale, but also to push the modelling of larger and more complex systems, including multichromophoric systems and multiphoton processes. Specifically, we will develop and implement novel excited state reaction dynamics methods that will enable to treat (i) on the same footing all possible competing deactivation channels (e.g., internal conversion; radiative processes, EET process, two-photon absorption and/or photoreactions); and (ii) to attain quantum yields and lifetimes of these complex photochemical scenarios; so that a complete "on-the-fly" excited state dynamics up to the ns regime will be obtained for the first time. By expanding the state-of-the-art modelling toolset for photochemistry it will permit us to do unprecedented predictions of photochemistry. By predicting and controlling complex photochemical events we will facilitate the next breakthrough in light-to-energy conversions, such as e.g., in artificial photosynthesis.

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
		☐ 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	
Geometries	Final and crucial geometry data (+ parameter data)	Generate	Digital	Geometry data (structures) and info on force field settings	.gro / .pdb / .xyz also .top and other param. files	250GB to 1 TB	
Cubefiles	Surfaces, potentials and other grid based data over structures	Generate	Digital	Data defined over Grids of atomistic structures	.cube/.cub (Gaussian .cube format)	50 GB to 200 GB	
Software*	Developed software modules (in python)	Generate	Digital	software	Git directory of mostly .py files	<5 GB	
Electronic_properties	Data concerning the calculation and determination of electronic properties	Generate	Digital	Computational data, approach and format to be decided/developed	To be decided/ developed	<20 GB	

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

Results_Visual	Visual data records of structures and properties (also graphs)	Generate	Digital	Images (and movies) of structures and their analysis (+graphs)	.png (images) and .mp4 (videos)	~100 GB	
Results_Data	Summarised and/or tabulated data of results	Generate	Digital	Data tables (and graphs)	.xlsx, .json, .pdf	<2 GB	

*Written software will be categorised per software module and will be labelled accordingly. For example: Local Mass Density module is called LoMaDe and has the structure of a python module, using default git format for easy portability.

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.	Not applicable
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:
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:
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

[4] See Glossary Flemish Standard Data Management Plan

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*.	While in use/being processed, computational directories will be labelled and structured with subdirectories in as systematic a way as possible, separating different types of calculations/approaches or including specifically labelled subdirectories with routinely structured processing for additional steps. Readme files are and will be included, listing and clarifying the purpose of each subdirectory. High throughput calculations will be done in an automatised manner, ensuring (and requiring) a systematic organisation of the corresponding data. When calculations are fully processed, crucial raw data will be retained and stored together by category. Appropriate metadate will be included and where the type or nature of the data might be unclear or is nonstandard, extra clarification will be included. Resulting data (tables, graphs, figures and the like) will be fully retained, but otherwise stored in a similar way as the above. Newly developed codes will be made available through github.
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: Many of the used types of data don't have a general metadata standard. Readme's and classification based on software will be used to help clarify the context and the nature of data. Many datafiles also contain small metadata blocks, which will be used appropriately. For coding, metadata will be made available in the form of readme's, manuals and documenting of the code, adhering to python PEP-8 standards.

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: Staging directory on the VSC (5 TB extra backup of computational data)
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	Full output from all calculations is very voluminous, but the field has standard protocols for collecting the key data. Specifically, data will be stored in our computers belonging to the Quantum Chemistry and Physical Chemistry Division (QCPC, KU Leuven) and on the data storage services of the Flemish Supercomputing Centre (VSC). Automatic daily back-up procedures are ensured at both our QCPC cluster as well as at the VSC facilities. In addition, crucial and processed data is stored on KULeuven personal's onedrive.
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 Our data storage capabilities are currently on the TB scale, and therefore sufficiently large for the project needs.

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*	Both on the KULeuven personal drive and on the VSC there are strict authorizations in place so no external/unauthorized user can access the data. The data will be stored in the university's secure environment. Only authorized persons 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?	The extra storage in VSC is €20/TB /year , calculated on how much is used, so a maximum cost of €100/year. The cost is overall minor and will be covered by the bench fee.

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.*	The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy. ☒ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The data will be stored on the university's central servers for at least 10 years for free. Extra generated costs will be afforded by the research group.

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: We prefer to make our data available under a CC-BY license. Developed codes (software) will be made available (through github). Some results will be available through papers and their SI (supporting information). The other data will remain available within our research group and can be provided to third parties if requested (and approved).
If access is restricted, please specify who will be able to access the data and under what conditions.	No restriction. The responsible person of the data will be Prof. Escudero, the PI of this project. The data will remain available within our research group and can be provided to third parties if requested (and approved).
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.	☒ KU Leuven RDR ☐ Other data repository (specify) ☐ Other (specify) Besides using KU Leuven RDR. Data will be included in the supporting information (SI) of published papers to the extent that is possible. Software will be available through github (links to its location will be included in relevant published papers). The remainder of the data (plus a summary of the other data) will be uploaded through lirias and zenodo.
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.*	Data will be uploaded to KU Leuven RDR, Zenodo for the KU Leuven community with a CC-BY-NC license. As much as possible data is also published in the SI of papers, where they will follow the standard licensing in that journal.

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 Published papers will have a DOI, so the data in the SI alongside it will have it automatically. For the data published in Lirias and zenodo we will follow standard KU Leuven protocol. (if this includes or recommends adding a specific type of PID then we will comply)
What are the expected costs for data sharing? How will these costs be covered?	The data sharing through university server and through github are free. The data shared through publication will be charged a fee, which will be covered by the FWO bench fee.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Prof. Daniel Escudero
Who will manage data storage and backup during the research project?	Prof. Daniel Escudero
Who will manage data preservation and sharing?	Prof. Daniel Escudero
Who will update and implement this DMP?	Prof. Daniel Escudero