

## 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	<b>Arne Meulemans 0000-0002-3316-842X</b>
Contributor name(s) (+ ORCID) & roles	<b>Daniel Escudero 0000-0002-1777-8578, Promotor</b>
Project number <sup>1</sup> & title	11PQU24N - Computational investigations of organic thermo-electric polymers (CIOTEP)
Funder(s) GrantID <sup>2</sup>	
Affiliation(s)	<input checked="" type="checkbox"/> KU Leuven <input type="checkbox"/> Universiteit Antwerpen <input type="checkbox"/> Universiteit Gent <input type="checkbox"/> Universiteit Hasselt <input type="checkbox"/> Vrije Universiteit Brussel <input type="checkbox"/> Other: ROR identifier KU Leuven: 05f950310

---

<sup>1</sup> "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

<sup>2</sup> 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	<p>To mitigate the energy crisis and the climate crisis, new sources of renewable energy are vital to the future of humankind. Thermoelectric (TE) materials, which convert heat to electricity, show great potential as alternative energy sources. Inorganic TE materials are currently the most effective, but organic thermoelectric materials (OTE) show great promise and are more sustainable to produce than inorganic materials. There is however a lack of understanding on the mechanism behind TE effects in semiconducting polymers and how the morphology and chosen molecular platform affects the TE efficiency.</p> <p>Which is why, in CIOTEP, I am developing computational protocols to directly determine the parameters determining the figure of merit of OTE materials. The target-end goal of my project is to provide the community a methodology to link structure to relevant OTE properties and to provide design rules for the next generation of OTE materials. This project will furthermore benefit from direct cooperation with two experimental groups, focused on polymer synthesis and OTE device fabrication and characterization, respectively.</p>
--	--

## 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 <sup>3</sup>.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
		<input type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data	<input type="checkbox"/> Digital <input type="checkbox"/> Physical	<input type="checkbox"/> Audiovisual <input type="checkbox"/> Images <input type="checkbox"/> Sound <input type="checkbox"/> Numerical <input type="checkbox"/> Textual <input type="checkbox"/> Model <input type="checkbox"/> Software <input type="checkbox"/> Other:		<input type="checkbox"/> < 1 GB <input type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> > 5 TB <input type="checkbox"/> NA	
Simulations	MD simulation data	Generate	Digital	GROMACS (and LAMMPS) files	GROMACS (and LAMMPS) files .gro, .trr, .mdp, .top, ...	5 to 50 TB	
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	Generate	Digital	software	Git directory of	<5 GB	

<sup>3</sup> Add rows for each dataset you want to describe.

	modules (in python)				mostly .py files		
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	
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	
<p>*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.</p> <p><i>GUIDANCE:</i>  <i>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 be described under documentation/metadata.</i>  <a href="#">RDM Guidance on data</a></p>							
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.	<input type="checkbox"/> Yes, human subject data; provide SMEC or EC approval number: <input type="checkbox"/> Yes, animal data; provide ECD reference number: <input type="checkbox"/> Yes, dual use; provide approval number: <input checked="" type="checkbox"/> No Additional information:
Will you process personal data <sup>4</sup> ? 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).	<input type="checkbox"/> Yes (provide PRET G-number or EC S-number below) <input checked="" type="checkbox"/> 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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> 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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> 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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please explain:

---

<sup>4</sup> See Glossary Flemish Standard Data Management Plan

### 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.](#)

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, a minimal\* set of crucial raw data will be retained and stored together by category. Appropriate metadata 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.

\*due to the enormous size of raw data per set (going from 50 GB for medium systems to 500GB for large systems, and expecting to process over 50 large systems) and the fact that most of it could be recalculated if required, only information required to be able to regenerate the data and crucial/very useful/reusable pieces of data (final geometries and .cube files for example) will be retained.

<p>Will a metadata standard be used to make it easier to <b>find and reuse the data</b>?</p> <p>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.</p> <p><i>REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.</i></p>	<p> <input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No         </p> <p>If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used:</p> <p>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.</p> <p>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.</p>
---	---

4. Data Storage & Back-up during the Research Project	
<p>Where will the data be stored?</p> <p><i>Consult the <a href="#">interactive KU Leuven storage guide</a> to find the most suitable storage solution for your data.</i></p>	<p> <input type="checkbox"/> Shared network drive (J-drive)  <input type="checkbox"/> Personal network drive (I-drive)  <input checked="" type="checkbox"/> OneDrive (KU Leuven)  <input type="checkbox"/> Sharepoint online  <input type="checkbox"/> Sharepoint on-premis  <input type="checkbox"/> Large Volume Storage  <input type="checkbox"/> Digital Vault  <input checked="" type="checkbox"/> Other: Staging directory stg_00137 on the VSC (5 TB extra backup of computational data)         </p>



<p>How will the data be backed up?</p> <p><i>WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?</i></p>	<p>Crucial and processed data is stored on KULeuven personal onedrive (and backed up to 2 pc's). Central maps with ongoing calculations are in a data-drive on the VSC when possible, which also has an reliable back-up routine.</p> <p>High-throughput calculation will be too voluminous to handle in there, so they will have to be performed in a non-automatically backed-up staging directory. The crucial data will be backed-up biweekly to the KULeuven onedrive.</p>
<p>Is there currently sufficient storage &amp; 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.</p>	<p><input checked="" type="checkbox"/> Yes*</p> <p><input type="checkbox"/> No</p> <p>For crucial and processed data there will be enough capacity.</p> <p>For raw data, especially in high-throughput steps there will not be enough space to contain everything, so removal of bulky non-essential data (f.e. MD trajectories of structure generation) will have to be performed when the data is sufficiently processed.</p>
<p>How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</p> <p><i>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.</i></p> <p><a href="#">Guidance on security for research data</a></p>	<p>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 used PC's are also properly protected with username, safe password and appropriate firewall</p> <p>Each KULeuven-associated PC requires username and password, which must be changed every year.</p>
<p>What are the expected costs for data storage and backup during the research project? How will these costs be covered?</p>	<p>The extra storage in stg_00137 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.</p>

## 5. Data Preservation after the end of the Research Project

<p>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...).</p> <p><a href="#">Guidance on data preservation</a></p>	<p><input checked="" type="checkbox"/> All data will be preserved for 10 years according to KU Leuven RDM policy</p> <p><input type="checkbox"/> 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</p> <p><input type="checkbox"/> Certain data cannot be kept for 10 years (explain)</p>
<p>Where will these data be archived (stored and curated for the long-term)?</p> <p><a href="#">Dedicated data repositories</a> 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 <a href="#">interactive KU Leuven storage guide</a>.</p>	<p>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.</p> <p><input checked="" type="checkbox"/> KU Leuven RDR</p> <p><input type="checkbox"/> Large Volume Storage (longterm for large volumes)</p> <p><input type="checkbox"/> Shared network drive (J-drive)</p> <p><input type="checkbox"/> Other (specify):</p>
<p>What are the expected costs for data preservation during the expected retention period? How will these costs be covered?</p>	<p>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.</p>

## 6. Data Sharing and Reuse

<p>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.</p> <p><i>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 &amp; RESTRICTED ACCESS. FOR MORE INFORMATION: <a href="https://wiki.surfnet.nl/display/STANDARDS/INFO-EU-REPO/#INFO-EURO-ACCESSRIGHTS">HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFO-EURO-ACCESSRIGHTS</a></i></p>	<p> <input checked="" type="checkbox"/> Yes, as open data  <input type="checkbox"/> Yes, as embargoed data (temporary restriction)  <input checked="" type="checkbox"/> Yes, as restricted data (upon approval, or institutional access only)  <input type="checkbox"/> No (closed access)  <input type="checkbox"/> Other, please specify:         </p> <p>           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).         </p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>The full dataset will be transferred to my supervisor and will be stored on the university's central servers. He or the future students could reuse the data with the approval from my PhD supervisor and me. The valuable data will be written into research papers. The paper-related information could be shared upon request by mail.</p>
<p>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.</p>	<p> <input type="checkbox"/> Yes, privacy aspects  <input type="checkbox"/> Yes, intellectual property rights  <input type="checkbox"/> Yes, ethical aspects  <input type="checkbox"/> Yes, aspects of dual use  <input type="checkbox"/> Yes, other  <input checked="" type="checkbox"/> No         </p> <p>If yes, please specify:</p>

<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<p> <input type="checkbox"/> KU Leuven RDR  <input type="checkbox"/> Other data repository (specify)  <input type="checkbox"/> Other (specify)         </p> <p>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 liras and zenodo.</p>
<p>When will the data be made available?</p>	<p> <input checked="" type="checkbox"/> Upon publication of research results  <input type="checkbox"/> Specific date (specify)  <input type="checkbox"/> Other (specify)         </p>
<p>Which data usage licenses are you going to provide? If none, please explain why.</p> <p><i>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.</i></p> <p>Check the <a href="#">RDR guidance on licences</a> for data and software sources code or consult the <a href="#">License selector tool</a> to help you choose.</p>	<p>Data will be uploaded to liras and Zenodo for the KULEuven 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.</p>

<p>Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.</p> <p><i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i></p>	<p><input type="checkbox"/> Yes, a PID will be added upon deposit in a data repository</p> <p><input type="checkbox"/> My dataset already has a PID</p> <p><input checked="" type="checkbox"/> No</p> <p>Published papers will have a DOI, so the data in the SI alongside it will have it automatically.</p> <p>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)</p>
<p>What are the expected costs for data sharing? How will these costs be covered?</p>	<p>The data sharing through university server and through github are free.</p> <p>The data shared through publication will be charged a fee, which will be covered by the FWO bench fee.</p>

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Day-to-Day management: Arne Meulemans After completion management: Daniel Escudero
Who will manage data storage and backup during the research project?	Day-to-Day management: Arne Meulemans After completion management: Daniel Escudero
Who will manage data preservation and sharing?	Day-to-Day management: Arne Meulemans Sharing and after completion management: Daniel Escudero
Who will update and implement this DMP?	Arne Meulemans