

FWO DMP Template - Flemish Standard Data Management Plan

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	Vince Maes (https://orcid.org/0000-0002-8373-8865)
Contributor name(s) (+ ORCID) & roles	Giovanni Samaey (https://orcid.org/0000-0001-8433-4523) (Promotor) Martine Baelmans (https://orcid.org/0000-0001-9905-4583) (Co-Promotor) Wouter Dekeyser (https://orcid.org/0000-0003-1954-3729) (Co-Promotor) Julian Koellermeier (https://orcid.org/0000-0002-8822-461X) (Co-Promotor)
Project number ¹ & title	3E200588 “Kinetisch-hybride methodes voor onzekerheidskwantificatie en robuust divertorontwerp voor fusie-energiereactoren”
Funder(s) GrantID ²	1S64723N
Affiliation(s)	<input type="checkbox"/> KU Leuven
Please provide a short project description	Modeling imperfection and parameter calibration lead to uncertainty in plasma edge models. To increase the reliability of plasma edge simulations, the uncertainty has to be quantified. The objective is to develop hybrid-kinetic computational methods for forward and backward uncertainty quantification, which can be used for robust divertor design.

¹ “Project number” refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. 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 data³.

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"/> Observational <input type="checkbox"/> Experimental <input type="checkbox"/> Compiled/aggregated data <input type="checkbox"/> Simulation data <input type="checkbox"/> Software <input type="checkbox"/> Other <input type="checkbox"/> NA	<input type="checkbox"/> .por <input type="checkbox"/> .xml <input type="checkbox"/> .tab <input type="checkbox"/> .csv <input type="checkbox"/> .pdf <input type="checkbox"/> .txt <input type="checkbox"/> .rtf <input type="checkbox"/> .dwg <input type="checkbox"/> .tab <input type="checkbox"/> .gml <input type="checkbox"/> other: <input type="checkbox"/> NA	<input type="checkbox"/> < 100 MB <input type="checkbox"/> < 1 GB <input type="checkbox"/> < 100 GB <input type="checkbox"/> < 1 TB <input type="checkbox"/> < 5 TB <input type="checkbox"/> < 10 TB <input type="checkbox"/> < 50 TB <input type="checkbox"/> > 50 TB <input type="checkbox"/> NA	
Simulation code	Python and Matlab scripts	New data	Digital	Software	.py and .m	<1 GB	/
Numerical experiments	Results produced by running Python and Matlab	New data	Digital	Experimental data	.h5 for Python generated results; .mat for Matlab generated	< 100GB	/

³ Add rows for each dataset you want to describe.

	scripts				results		
Manuscripts	Paper and thesis text files	New data	Digital	Text	.tex	<1GB	/
Papers	Literature study	Reused data	Digital	Text	.pdf	<100GB	/

GUIDANCE:

DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, ...). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.

EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA⁴ (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.

EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR, .SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG, .GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.

DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.

PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).

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.	The only reused data is the collection of papers that make up my literature study. The sources are the different journals in which these papers where published.
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, please describe these issues further and refer to specific datasets or data types when appropriate.	<input checked="" type="checkbox"/> No If yes, please describe: /

⁴ These data are generated by combining multiple existing datasets.

<p>Will you process personal data⁵? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.</p>	<p><input checked="" type="checkbox"/> No If yes:</p> <ul style="list-style-type: none"> - Short description of the kind of personal data that will be used: / - Privacy Registry Reference: /
<p>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.</p>	<p><input checked="" type="checkbox"/> Yes If yes, please comment: People can take the open source code (and corresponding papers) that we will make publicly available as an example to implement the newly developed methods into their code suites, the codes themselves, however, are not directly usable commercially as they are designed with academic purposes in mind.</p>
<p>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.</p>	<p><input checked="" type="checkbox"/> No If yes, please explain: /</p>
<p>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.</p>	<p><input checked="" type="checkbox"/> No If yes, please explain: /</p>

⁵ See Glossary Flemish Standard Data Management Plan

3. Documentation and Metadata

<p>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).</p>	<p>Code is documented and has an accompanying readme. The code is in version control, where the git history allows to trace back changes. Computational experiments are explained carefully in papers and are tested for reproducibility. Additionally, published code is accompanied with a Docker file such that a virtual machine can be set up with the right environment to run the published codes. Using these codes and Docker file based virtual machines, each of the published computational experiments can be reproduced.</p>
<p>Will a metadata standard be used to make it easier to find and reuse the data?</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 checked="" 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: For computational experiments we encode metadata about the running environment for the code in Dockerfiles.</p> <p>If no, please specify (where appropriate per dataset or data type) which metadata will be created: Metadata about how to reproduce results will be documented in papers and readme files.</p> <p>Note: in the future, the plan is to automate metadata generation in our research group using iRODS. The pilot for this change in data management flow still has to be set up.</p>

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	<p>The data is stored on my PC and, depending on the data type, a back-up is stored.</p> <p>Back-up during research:</p> <ul style="list-style-type: none"> - For version-controlled files, we use the KU Leuven gitlab service. - Literature study is in a reference manager (Zotero) - For other files, we use KU Leuven OneDrive or the NextCloud server of the Department of Computer Science. <p>Back-up after research:</p> <ul style="list-style-type: none"> - When possible, all relevant data/code will be made publicly available with the paper. - Manuscripts and all corresponding data are stored with read-only access at the Department of Computer Science.
<p>How will the data be backed up?</p> <p><i>WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.⁶</i></p> <p><i>REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.</i></p>	<p>All files are synced with departmental file servers to ensure daily backup.</p> <p>(NUMA DMP POLICY)</p> <p>The sole exception is the papers of the literature study which are in Zotero.</p>
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.	<p><input checked="" type="checkbox"/> Yes</p> <p>If yes, please specify concisely: All data is digital and has a back-up at the date of writing this file and we are nowhere close to the data limits. If I would get close to data limits, I can ask for more storage space, so this is not an issue.</p>

⁶ Source: Ghent University Generic DMP Evaluation Rubric: <https://osf.io/2z5g3/>

<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. ⁷</i></p>	<p>This is covered by the groupwide data management plan (DMP) of the NUMA research unit.</p>
<p>What are the expected costs for data storage and backup during the research project? How will these costs be covered?</p>	<p>This is covered by the NUMA research unit.</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>	<ul style="list-style-type: none"> - When possible, all relevant data/code will be made publicly available with the paper. - Manuscripts and all corresponding data are stored with read-only access at the Department of Computer Science
<p>Where will these data be archived (stored and curated for the long-term)?</p>	<p>At the Department of Computer Science</p>

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	This is covered by the NUMA research unit.
--	---

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 & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEU-REPO-ACCESSRIGHTS</i></p>	<p><input checked="" type="checkbox"/> Yes, in an Open Access repository</p> <p><input checked="" type="checkbox"/> Yes, in a restricted access repository (after approval, institutional access only, ...)</p> <p><input type="checkbox"/> No (closed access)</p> <p><input type="checkbox"/> Other, please specify:</p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>The promotor of the project, Giovanni Samaey, has access to all repositories and can regulate access to the data.</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</p> <p><input type="checkbox"/> Yes, intellectual property rights</p> <p><input type="checkbox"/> Yes, ethical aspects</p> <p><input type="checkbox"/> Yes, aspects of dual use</p> <p><input type="checkbox"/> Yes, other</p> <p><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>Open Access repositories: zenodo</p> <p>Restricted Access repositories: KU Leuven gitlab servers</p>

<p>When will the data be made available?</p> <p><i>THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.</i></p>	<p>Open Access repositories will be created upon publication of the corresponding papers. Restricted Access repositories can be made available on request.</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><i>EXAMPLE ANSWER: E.G. "DATA FROM THE PROJECT THAT CAN BE SHARED WILL BE MADE AVAILABLE UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE (CC-BY 4.0), SO THAT USERS HAVE TO GIVE CREDIT TO THE ORIGINAL DATA CREATORS." ⁷</i></p>	<p>This is still under consideration.</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 checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes: We are going to use a KU Leuven system that is still under development.</p>

⁷ Source: Ghent University Generic DMP Evaluation Rubric: <https://osf.io/2z5g3/>

What are the expected costs for data sharing? How will these costs be covered?	This is not known yet, but the costs are expected to be modest.
---	--

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Vince Maes
Who will manage data storage and backup during the research project?	Vince Maes
Who will manage data preservation and sharing?	Giovanni Samaey
Who will update and implement this DMP?	Vince Maes