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	**Arne Bouillon** (ORCID ID 0000-0002-2745-1982)
Contributor name(s) (+ ORCID) & roles	**Supervisors:** Giovanni Samaey (ORCID ID 0000-0001-8433-4523) and Dirk Nuyens (ORCID ID 0000-0002-4555-2314)
Project number[1] & title	Design and analysis of multilevel interacting-particle methods with surrogate models for efficient Bayesian inversion
Funder(s) GrantID[2]	Funded by a Fundamental Research predoctoral mandate (**1169725N)** by FWO
Affiliation(s)	KU Leuven
Please provide a short project description	We propose three main contributions. First, we extend our recent multilevel methods, which combine models at different fidelity levels. We will add a resampling step and rigorously analyse the non-asymptotic behaviour of these algorithms. Second, we will carefully analyse and extend methods that build surrogate models with inaccurate samplers, and then sample accurately with the surrogate. Third, we will design entirely new methods that combine all these aspects to perform reliable, efficient parameter estimation in large-scale problems, with turbulent flows as our main application.

[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 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
Own code	Research code, and potentially software libraries created during the project	New	Digital	Software	Various software formats	< 100 MB	N/A
Own manuscripts	The source code for the LaTeX-typeset papers written during the project	New	Digital	Software Other: Manuscript	.tex .bib Various image formats	< 100 MB	N/A
Computational experiments	The exact code needed to reproduce the experiments included in our papers, as well as their results, kept for reproducibility and open-science purposes	New	Digital	Software and experiment output data (text files, images…)	Various software formats .txt Various image formats	< 1 GB	N/A

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

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 data[4] (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.	N/A
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.	No
Will you process personal data[5]? 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.	No

[4] These data are generated by combining multiple existing datasets.
[5] 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.	Both the techniques in our papers and potential software libraries could be useful to commercial companies.
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.	No
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.	No

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).	Code will be documented and accompanied by a README.md file. The level of detail of the documentation varies – software libraries will be documented very extensively, research code less so. Computational experiments will always include the exact code needed to reproduce them. They may also be commented (e.g. in Jupyter Notebooks).

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.*	**No**, for the types of data we produce, we deem a metadata standard not applicable. All metadata consists of documentation.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	**Own code, Own manuscripts:** On KU Leuven's GitLab server **Computational experiments:** We document the experiments and their outcome for reproducibility (code version, scripts, data files), either on KU Leuven's GitLab server or – experimentally – within KU Leuven's ManGO platform
How will the data be backed up? *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] *Refer to institution-specific policies regarding backup procedures when appropriate.*	For version-controlled files, we use the GitLab of KU Leuven. For other files, we use KU Leuven's OneDrive or the departmental NextCloud.

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

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; this project requires a very limited amount of storage space, which KU Leuven's storage options can definitely provide
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.* [7]	These storage and backup solutions are protected by KU Leuven access control measures. We defer to them. Additionally, we currently foresee all research data being publicly accessible once the associated research is published.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	These storage and backup solutions are already widely established in our research group and do not come with additional costs

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…).	We plan for all data to be retained for at least five years after the end of the project.

Where will these data be archived (stored and curated for the long-term)?	These data will remain in their respective locations after the end of the project.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	There are no expected costs associated with this preservation.

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*	**All data:** Yes, in an Open Access repository
If access is restricted, please specify who will be able to access the data and under what conditions.	N/A
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.	No
Where will the data be made available? If already known, please provide a repository per dataset or data type.	The GitLab repositories are made public under the `gitlab.kuleuven.be/numa/public/` group. Manuscript data is made available in LaTeX form as source code on arXiv, and in PDF form on arXiv.
When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	All data is made available upon publication (or pre-publication) of research results.

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.* *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]	We plan for reusable software libraries to be released under an MIT license. Research code may or may not be accompanied by a license. If their sole goal is to reproduce our results without extensibility having been taken into consideration in their design, we may deem a license unnecessary. Our manuscripts will be released under the licenses used by the journal of publication. ArXiv preprints are released under the CC BY 4.0 license.
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.*	No
What are the expected costs for data sharing? How will these costs be covered?	We do not expect any costs associated with data sharing.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Arne Bouillon**

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

Who will manage data storage and backup during the research project?	**Arne Bouillon**
Who will manage data preservation and sharing?	**Arne Bouillon** and, after the end of his doctoral programme, **Giovanni Samaey**
Who will update and implement this DMP?	**Arne Bouillon**