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	**Rayan Dhib 0000-0003-3670-4678**
Contributor name(s) (+ ORCID) & roles	**Supervisor: Stefaan Poedts (0000-0002-1743-0651)** **Co-supervisor: Andrea Lani (0000-0003-4017-215X)**
Project number[1] & title	
Funder(s) GrantID[2]	1S66823N
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:

---

[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.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	Solar activity is at the origin of time and space varying conditions governing the environment of the solar system and resulting in what is referred to as Space Weather (SW). The eruptive events in the solar atmosphere are generally associated with the ejection of large amounts of magnetized plasma into interplanetary space at velocities up to several thousand kilometres per second. The latter phenomena, called Coronal Mass Ejections, are considered the main cause of most severe SW events, able to damage both space-borne and ground-based infrastructures when interacting with Earth's magnetosphere. In order to allow for mitigating such threatening phenomena, numerical prediction tools require advanced physical models and computational methods able to properly characterize all relevant phenomena providing more accurate and considerably faster results matching observational data. The proposed research project aims to deliver a new-generation data-driven solver relying upon the most advanced models (including the multi-component nature of solar plasmas) using modern high-order methods and novel adaptive mesh refinement algorithms to enable accurate characterization of the solar corona. The newly developed code within the open source COOLFluiD platform will be also compatible with high-performance computing architectures (e.g. GPUs) for a further boost in run-time performance and will be used to run realistic solar corona simulations providing more reliable input to the heliospheric SW forecasting tool EUHFORIA.

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[4].

Dataset Name	Description	New or Reused	Digital or Physical	*ONLY FOR DIGITAL DATA* Digital Data Type	*ONLY FOR DIGITAL DATA* Digital Data Format	*ONLY FOR DIGITAL DATA* Digital Data Volume (MB, GB, TB)	*ONLY FOR PHYSICAL DATA* Physical Volume
(1) Source Code	new/modified C++ files corresponding to the foreseen extensions of the open source COOLFluiD software (https://github.com/andrealani/COOLFluiD) under license LGPL v3	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☒ Software ☐ Other ☐ NA	☒ other: ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
(2) Simulation files	input unstructured mesh files in Gmsh ASCII format (http://gmsh.info/doc/texinfo/gmsh.html#MSH-file-format) and COOLFluiD simulation files in CFcase format	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☐ Experimental ☐ Compiled/ aggregated data ☒ Simulation data ☐ Software ☐ Other	☒ other: .msh ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB	

---

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

(3) Observational input files	input magnetogram files for initializing solar simulations in GONG (from https://gong.nso.edu/data/magmap/) or other formats to be investigated/decided later in the project	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☒ Simulation data ☐ Software ☐ Other ☐ NA	☒ other: ☐ NA	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA
(4) output solution files	output solution files including the electromagnetic and plasma fields resulting from our numerical simulations in CFmesh (COOLFluiD ASCII/binary formats which are described in https://github.com/andrealani/COOLFluiD/wiki/CFmesh-format- for-parallel-I-O), point or block unstructured TECPLOT ASCII format (details in http://home.ustc.edu.cn/~cbq/360_data _format_guide.pdf), ParaView VTK ASCII (https://lorensen.github.io/VTKExamples/site/VTKFileFormats/)	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☐ Experimental ☐ Compiled/ aggregated data ☒ Simulation data ☐ Software ☐ Other ☐ NA	☒ other: ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA
(5) Media files	images (JPEG, PNG, PDF formats) and videos files of numerical solutions in AVI or MP4 formats	☒ Generate new data	☒ Digital ☐ Physical	☐ Observational ☐ Experimental	☐ .por ☐ .xml ☐ .tab	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB

| | | ☐ Reuse existing data | | ☐ Compiled/ aggregated data
☒ Simulation data
☐ Software
☐ Other
☐ NA | ☐ .csv
☐ .pdf
☐ .txt
☐ .rtf
☐ .dwg
☐ .tab
☐ .gml
☒ other: JPEG, PNG, PDF, AVI, MP4
☐ NA | ☐ < 1 TB
☐ < 5 TB
☐ < 10 TB
☐ < 50 TB
☐ > 50 TB
☐ NA | |

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[5] (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).

---

[5] These data are generated by combining multiple existing datasets.

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.	For dataset (1) some source code files are reused in the code extension process. Both reused and new files are available on the COOLFluiD GitHub repository (https://github.com/andrealani/COOLFluiD). For dataset (3), the main source of used magnetogram files will be taken from GONG (https://gong.nso.edu/data/magmap/)
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.	☐ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☒ No If yes, please describe:

Will you process personal data[6]? 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.	☐ Yes
	☒ No
	If yes:
	
	   - Short description of the kind of personal data that will be used:
	   - Privacy Registry Reference:

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:

---

[6] See Glossary Flemish Standard Data Management Plan

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).	The resulting data (datasets (2), (4) and (5)) will be organized in folders with meaningful names, one per simulation run. Each folder will contain a README.txt explaining how to run the specific case and explaining what the related input/output files are. In addition, the header of the case file (.CFcase) will summarise the conditions and purpose of the testcase. All the output files will be put inside an internal folder RESULTS, also providing its own README.txt with additional information concerning the content of those files and visualization tips if any (for instance guidance on how to use additional scripts/macros for etracting some relevant quantities). In addition, newly generated code files (dataset (1)) will be uploaded to the GitHub repository and well-documented (for example through meaningful naming and use of explanatory comments in the code)

| 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 no, please specify (where appropriate per dataset or data type) which metadata will be created:

The CFcase file which drives COOLFluiD simulations is in fact a quite powerful metadata file which summarize all the simulation parameters, the paths and names to I/O files etc.
Current efforts are being put into improving that format, describing in all details on the COOLFluiD wiki website and, in particular, enforce stricter rules on what to write on the file header in order to provide a quick glimpse of:
- the main parameters characterizing the simulation;
- which variables are saved on each output file;
- which scripts (if any) should be used for visualization or producing media (images or videos) out of the data;
- paths to all I/O files and description of each of them.
This way, just having access to the CFcase file alone will give full details about the numerical simulation and all the corresponding data. |

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

| Where will the data be stored? | The larger volume data, typically solution files, will be temporarily stored on the VSC cluster, then periodically selected and copied to our research unit central storage facility and, partly, to Box when needing to share data more efficiently within the team. Other data (e.g. source code, input simulation files, smaller mesh files, magnetograms) will be directly stored on the COOLFluiD GitHub repository. |

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.[7]* *Refer to institution-specific policies regarding backup procedures when appropriate.*	The data will be stored on the university's central servers with automatic daily back-up procedures. In addition, all data uploaded to the COOLFluiD github repository will be automatically backed up under versioning control.
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 If yes, please specify concisely: The storage and backup capacity is sufficient. The larger data in our case will be the mesh files and especially solution files in TECPLOT and/or CFmesh and/or Paraview formats. Those will be initially stored in the VSC cluster and then selected and backed up in our research unit central storage facility and/or Box.
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]*	During this project, there is no use of sensitive data, so this issue is not of concern. All the developments will be uploaded to the open source platform COOLFluiD with no particular worries about who can use or access the produced data.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	No expected expenses for data storage and backup. Most of the generated data will be uploaded to the GitHub repository while larger (solution) files will be stored and backed up using our research unit central storage facility.

---

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

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...).	With reference to the Data description at point 2: (1) all new source code will be preserved and automatically backed up; (2) a selection of representative CFcase files and meshes will be stored; (3) a selection of input magnetograms will be stored; (4) a selection of the numerical output (~30% of the total, due to high storage requirements) will be stored; (5) a selection of images and videos will be stored. No particular contractual or legal restrictions shall apply on the usage of (1) (2) (3). However, data of type (4) and (6) resulting from this project may be reused by third party only after negotiating, case by case, the conditions for their exploitation with the researcher and supervisors of this project.
Where will these data be archived (stored and curated for the long-term)?	Part of the data will be archived directly on the COOLFluiD Github repository (source code, input files, some images and videos) while the larger part (big meshes, solution files, all images and videos) will be archived on the Large Volume Storage (LVS) system, after having deleted data corresponding to unsuccessful or useless simulation runs.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The expected cost for storing data on the LVS system is 156,60 EUR/TB/Year. This expense will be covered by the FWO bench fee if needed.

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, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify: With reference to the Data description at point 2: (1) all new source code will be uploaded to the GitHub repository and thus will be available for reuse during and after the project; (2) a selection of representative CFcase files and meshes will be also uploaded to GitHub; (3) input magnetograms used in representative cases will be uploaded along the mesh and CFcase files; (4) and (5) a selection of solution files (Tecplot, CFmesh, images and/or videos) will be made available on the repository.
If access is restricted, please specify who will be able to access the data and under what conditions.	

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.	As aforementioned, apart from large solution files most of the data will be uploaded on the GitHub repository (https://github.com/andrealani/COOLFluiD).
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'.*	Upon 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."* [8]	New/modified C++ files corresponding to the extensions of the open source COOLFluiD software (https://github.com/andrealani/COOLFluiD) under license LGPL v3.
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 ☒ No If yes:
What are the expected costs for data sharing? How will these costs be covered?	No costs are expected since most of the developments will be made available on an open access repository.

7. Responsibilities

Who will manage data documentation and metadata during the research project?	**Rayan Dhib**

---

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

Who will manage data storage and backup during the research project?	**Rayan Dhib**
Who will manage data preservation and sharing?	**Rayan Dhib**
Who will update and implement this DMP?	**Rayan Dhib**