WEAVE project G.0025.23N

A Data Management Plan created using DMPonline.be

Creators: Stefaan Poedts https://orcid.org/0000-0002-1743-0651, Marian Lazar

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Stefaan Poedts https://orcid.org/0000-0002-1743-0651

Data Manager: Marian Lazar

Grant number / URL: G.0025.23N

ID: 200582

Start date: 01-08-2023

End date: 31-07-2026

Project abstract:

The main goal of the proposed project is the advancement of the non-equilibrium plasma theory and specific key applications by employing the Regularized Kappa Distribution (RKD) introduced recently for a realistic interpretation of suprathermal particle populations and their macroscopic implications in various astrophysical systems. The new RKD concept motivates systematic explorations of its consequences in kinetic as well as fluid theory and corresponding studies of astrophysical plasma systems on small and large scales. With the proposed research we will intensify the successful cooperation between the two applicants that resulted in a number of joint publications during the last 13 years. It will enable a significant strengthening of the combination of the complementary expertise of the two groups in the field of the fluid modelling (primarily Leuven) and the kinetic modelling (primarily Bochum) of the solar wind. The proposed project aims to develop mutually consistent theoretical kinetic and fluid approaches based on the RKD, and to validate them by applying corresponding results to non-equilibrium
plasmas in the heliosphere. This validation of the theory and the
modeling will be made on the basis of various spacecraft data,
particularly from the recently launched missions Parker Solar Probe (NASA) and Solar Orbiter (ESA).

Last modified: 23-06-2023

WEAVE project G.0025.23N
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Question not answered.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Question not answered.

What’s the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Question not answered.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Question not answered.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Question not answered.

WEAVE project G.0025.23N
FWO DMP (Flemish Standard DMP)

1. 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.

				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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Observational analysis	input data: NASA, ESA databases; various spacecraft modeling, plotting, fitting output data, plots	Reuse existing data: input data files, numerical software for analysis Generate new data: output files	Digital	• Observational • Experimental • Compiled/aggregated data	.xml, .tab, .cvs,.pdf, .txt, .rtf,	• <100GB	
Linear/QL solvers	numerical solvers	Reuse existing data: DIS-K (solver for Maxwellian/Kappa plasmas) Generate new data: adapted ALPS - new solver (imported) adapted for RKD plasmas	Digital	• Software	.cvs,.pdf, .txt, .rtf others	• <1TB	
Source codes	Vlasov, Hybrid, PIC, MHD	Reuse existing data Generate new data	Digital	• Software	• Fortran, C++, IDL, Ptython, etc.	• <1TB	
manuscripts	technical reports, documentation, peer-reviewed papers	generate new data	Digital	other	.tex, .docx, .pdf	<100GB	

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:

DIS-K: DIspersion Solver for Kappa plasmas available at https://github.com/ralopezh/dis-k, and protected by the publication doi 10.1017/S0022377821000593
ALPS is provided by our collaborator D. Verscharen (UCL)

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

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.

- No

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 in the comment section 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 in the comment section to what data they relate and which restrictions will be asserted.

- No

2. 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).

Each data type will be accompanied by some form of documentation.
Observational data will be accompanied by various forms of documentation (specific Readme explanatory files), and will be organized in raw data (input) and processed data (output) files (formats: dat, txt, etc.), in folders and directories with meaningful names associated with each type of observational data and/or instrument:

- particle input data: fluxes, pitch angle and velocity distributions,
- particle output data (moments of the velocity distributions, density, temperature components, drift (bulk) speeds, heat flux, kappa and alpha parameters of RKD distributions);
- waves and fluctuations input data (frequencies, wave numbers, wave lengths, wave density energies).

Source files/scripts will be accompanied by documentation (specific Readme explanatory files), and will be organized in folders and directories with meaningful names associated with each type of numerical operation:

- scripts and input data for numerical solvers (DIS-K, wave dispersion and stability analysis),
- scripts and input data for the simulation codes, macroscopic fluid models (EUHFORIA, ICARUS), and kinetic models (PIC and hybrid simulation codes of the waves and fluctuations, turbulence, etc.), organized in folders and explained by specific Readme files. Some of these codes are already described on the EUHFORIA website (...???...), but the explanation will be updated if needed.
- output data from the wave dispersion and stability analysis (wave number dispersion of the wave frequency and growth/damping rates)
- output data from the macroscopic simulation of RKD plasmas.

Reports (PDF) and plots (PDF/JPG/PNG) describing the data/methodology for presentations at internal meetings
Publication files, journal, conference papers (tex, PDF, PNG), describing details of models, run setups and analysis corresponding to the generated data to be preserved

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

We will use the local (BOX, VC) and public repositories (Github, Zenodo) for all data that are meant to be released with open access.

- programing and simulation code(s), I/O files names and paths, main input parameters
- URL, format, name, description of input observational data (e.g., tabulated distributions, plasma parameters, magnetic field)
- quantities saved/shown on each output file (e.g., tabulated, wave frequency, wave number, growth rate, time steps)
- detailed description of media content (pics, video)
- how to use given scripts for visualization or producing media (readme.txt files)

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on the local repositories (BOX, VSC) and Department servers.
Alternatively, for the results published as papers and accompanied data sets, we opt
to archive and share them on public repositories, locally (Lirias) and external (Github, Zenodo, NASA-ADS, Harvard's Dataverse, EUDAT, Open Science Framework).

How will the data be backed up?

The data will be backed up on local disks and external hard drives.

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

There is more than sufficient storage available in the servers/machines in our department. The simulation data to be produced by this project are actually minimal in comparison with other much larger datasets which are produced by colleagues. The fact that we focus on steady-state simulations on grids not exceeding a few millions elements guarantees rather limited and manageable needs in terms of storage.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All data will be stored in accounts for which an authentication is strictly required. If data are stored on external hard drives, any action related to using these devices will be centralized and monitored by the data manager.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

At this point, we don't expect specific costs for storage during the research project, since we have free access to storage within the department in available machines. However, if necessary, a few thousands euros from the operational budget (32k EUR/year) will be more than sufficient to satisfy our storage needs.

4. 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...).

All data corresponding to operational code and valid/published numerical results will be preserved.

Where will these data be archived (stored and curated for the long-term)?

Github (for source code and some simple test-case setups) and Zenodo (simulation data) as mentioned before. In addition, some publications will be open access and therefore available on the corresponding scientific journal websites.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs foreseen, since both Github and Zenodo are free.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

The source code extensions will be available at all times through local or public repositories. All the other data will be available for reuse by people outside the project only after the project or after scientific publications via public repositories.

If access is restricted, please specify who will be able to access the data and under what conditions.

During the project only members of the team and possible external collaborators (if any) will be able to access the numerical/simulation data. Updates to the source code will be available at all times via local and public repositories.

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 in the comment section 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.

Github (for source code and some simple test-case setups) and Zenodo (simulation data) as mentioned before. In addition, some publications will be open access and therefore available on the corresponding scientific journal websites.

When will the data be made available?

The source code extensions will be available at all times, while the numerical/simulation data will only be made available after publication of related results.

Which data usage licenses are you going to provide? If none, please explain why.

In general we won't provide any license, since they will be made public on repositories only after they have been already used for scientific publications which will protect our copyright.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

This will be provided when uploading data on repositories, after publishing the corresponding results.

What are the expected costs for data sharing? How will these costs be covered?

Zero costs, since the mentioned repositories provide free data sharing platforms.

6. Responsibilities

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

The postdoc/phd researcher(s) working on the project.

Who will manage data storage and backup during the research project?

The postdoc/phd researcher(s) working on the project.

Who will manage data preservation and sharing?

Stefaan Poedts (Co-PI)

Who will update and implement this DMP?

Stefaan Poedts (Co-PI)

WEAVE project G.0025.23N
GDPR

GDPR

Have you registered personal data processing activities for this project?

Question not answered.

WEAVE project G.0025.23N
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable