

---

## Kinetic Modelling of Novel solar wind Observations.

*A Data Management Plan created using DMPonline.be*

**Creator:** Luca Pezzini

**Affiliation:** KU Leuven (KUL)

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

**Template:** FWO DMP (Flemish Standard DMP)

**ID:** 206843

**Start date:** 01-09-2023

**End date:** 01-09-2027

### **Project abstract:**

Recent observations from various solar wind missions have revealed that the non-collisionality of solar wind plasma leads to fascinating behavior in plasma kinetics. Electrons and protons in the plasma often exhibit nonequilibrium features in their velocity distribution functions, resulting in anisotropies that can give rise to kinetic microinstabilities and electromagnetic waves. Plasma particles can also gain energy through heating and scattering phenomena caused by these waves. Furthermore, the adiabatic radial expansion of the solar wind and the ubiquitous presence of magnetic field switchbacks influence the thermodynamics of the system, thereby changing the configuration of equilibria. In this proposal, we aim to study the kinetics of solar wind plasma through cutting edge numerical simulations, developing a new model that extends microscale simulations of plasma to the macroscale of switchbacks. We will use advanced numerical methods that are suitable for studying plasma physics at both the microscopic and macroscopic scales to accomplish this goal. Our study will bridge the gap between the physics of microprocesses and the behavior of plasma at a larger scale that is currently lacking in state-of-the-art models. By conducting numerical simulations of solar wind plasma based on Parker Solar Probe in situ observations, we will gain a deeper understanding of the behavior of solar wind plasma and contribute to fundamental advancements in plasma physics.

**Last modified:** 26-04-2024

**Kinetic Modelling of Novel solar wind Observations.**  
**DPIA**

---

**DPIA**

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

Question not answered.

**Kinetic Modelling of Novel solar wind Observations.**  
**GDPR**

---

**GDPR**

**Have you registered personal data processing activities for this project?**

Question not answered.

## Kinetic Modelling of Novel solar wind Observations.

### 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)**

My research does not involve dealing with personal data. Simulation data produced by ECSim3D/iPIC3D is saved in a .h5 format. The field files occupy roughly 200 GB per simulation, while the particle files occupy roughly 4 TiB per simulation. I will use Python scripts to analyse the data, typically producing .png figures, which I will include in my scientific publication in LaTeX. These types of files occupy negligible space on the disk compared to the simulation data. I will archive the existing data for future scientific research or extensions of the current research.

**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)**

Dr. Jorge Amaya, a former research expert at the CmPA, has implemented a comprehensive storage solution to manage and preserve our team's data. This multi-level solution includes a local NAS, a KU Leuven iRODS repository, shared disk space at VSC, and a freely accessible data repository for publications. To ensure the reproducibility of our research, all data that leads to publications will be stored for at least 5 years after the project's completion.

**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)**

N.A.

**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)**

N.A.

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

N.A.

# Kinetic Modelling of Novel solar wind Observations.

## 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
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• Generate new data</li> <li>• Reuse existing data</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• Digital</li> <li>• Physical</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• Observational</li> <li>• Experimental</li> <li>• Compiled/aggregated data</li> <li>• Simulation data</li> <li>• Software</li> <li>• Other</li> <li>• NA</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ...</li> <li>• NA</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• &lt;100MB</li> <li>• &lt;1GB</li> <li>• &lt;100GB</li> <li>• &lt;1TB</li> <li>• &lt;5TB</li> <li>• &lt;10TB</li> <li>• &lt;50TB</li> <li>• &gt;50TB</li> <li>• NA</li> </ul>	
ECSim model		Generate new data	Digital	Software	.c and .c++	<100 MB	N/A
ECSim - Fields	Simulation data containing the simulated field	Generate new data	Digital	Simulation data	.h5	<1TB	N/A
ECSim - Particles	Simulation data containing the simulated particles quantities	Generate new data	Digital	Simulation data	.h5	<5TB	N/A

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)? 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).

The ECSim model will come with a comprehensive user guide detailing the process of compiling the code and running simulations. Furthermore, to handle the output data, which is formatted in .h5, a Python script will be available for post-processing purposes.

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

Yes, a metadata standard will be employed to facilitate the discovery and reuse of the data. Specifically, we will add the input file related to each run. In the input file all the specific characteristic of the run are specified parameter wise but also with a small description of the physics is currently modelled.

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

Where will the data be stored?

Dr. Jorge Amaya, a former research expert at the CmPA, has implemented a comprehensive storage solution to manage and preserve our team's data. This multi-level solution includes a local NAS, a KU Leuven iRODS repository, shared disk space at VSC, and a freely accessible

data repository for publications. To ensure the reproducibility of our research, all data that leads to publications will be stored for at least 5 years after the project's completion.

**How will the data be backed up?**

Regular backups will be performed by the IT staff.

**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

To acquire computational hours, we conduct both weak and strong scaling tests on the VSC cluster before initiating each simulation. This procedure aids us in gauging the required computational power and estimating the necessary disk storage for archiving the database. As a result, prior to each simulation, we accurately assess the disk space required, ensuring optimal resource allocation.

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

All data and model source codes will be managed in strict compliance with IT guidelines, including password protection for all computers and external drives. Our IT staff continuously implements security upgrades to ensure robust protection. When sharing datasets, read-only permissions will be enforced to maintain data integrity and security.

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

We do not anticipate incurring any supplementary expenses for data storage or backup, aside from the ongoing operational costs associated with the IT department, which is covered by KU Leuven.

**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 will be retained for a minimum of five years, with the possibility of longer retention periods.

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

The KU Leuven iRODS repository, shared disk space at VSC, and a freely accessible data repository for publications.

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

The KU Leuven iRODS repository is free of charge.

**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
- Yes, in a restricted access repository (after approval, institutional access only, ...)

The published results will be made available in an open access repository. However, the model source code will only be made available in a restricted access repository.

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

The code will be shared upon mutual agreement, following a 'rules of road' agreement to be established through email exchanges.

**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.**

It is yet to be decided.

**When will the data be made available?**

Publication datasets will be made available upon manuscript acceptance. Access to the model source code will be restricted after the project concludes.

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

We do not intend to share our ECSim kinetic model as a community model. However, we will utilize this model with external collaborators under mutual agreement. As a result, we will not provide a license for its distribution.

**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.**

- No

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

The iRODS repository is free of charge.

## **6. Responsibilities**

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

Luca Pezzini



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

Luca Pezzini

**Who will manage data preservation and sharing?**

Luca Pezzini

**Who will update and implement this DMP?**

Luca Pezzini

.