FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

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	Daniel Groselj Orcid: 0000-0002-5408-3046
Contributor name(s) (+ ORCID) & roles	Fabio Bacchini Orcid: 0000-0002-7526-8154
Project number [1] & title	Project number: 12B1424N How relativistic particles shine in astrophysical collisionless plasma turbulence
Funder(s) GrantID [2]	
Affiliation(s)	KU Leuven ROR identifier KU Leuven: 05f950310
Please provide a short project description	Coronae of black-hole accretion disks in X-ray binaries and at the cores of active galaxies are extremely bright sources of high-energy radiation. The radiation is produced in a hot ionized gas (plasma) that shines through emission from relativistic particles. The energization of particles is mediated by collisionless plasma processes, among which electromagnetic plasma turbulence is a prime contender. A comprehensive understanding of the mechanisms for high-energy emission therefore requires detailed kinetic models for how the particles are energized and how they pass their energy to radiation. This project aims to make important breakthroughs on the subject by performing the most detailed kinetic plasma simulations of turbulence coupled with radiative processes, including ab initio modeling of Compton scattering, production and annihilation of electron-positron pairs. The simulations will be directly compared with observations, particularly with the hard spectral state in X-ray binaries. These results will deepen our understanding of the production of high-energy emission in extreme astrophysical environments and significantly advance the state-of-the-art in modeling turbulent radiative plasmas.

---

[1] "Project number" refers to the institutional project number. This question is optional. 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
Simulation outputs	The research in this proposal will produce output from large 2D and 3D simulations of kinetic plasma turbulence (electromagnetic fields and particle data) and restart files that allow to restart the simulation at a given point. The data from the simulation will be post-processed and analyzed for scientific plots and visualizations. The key simulation data supporting publications will be made publicly available. The rest of the data will be kept for the entire duration of the project and for 5 years thereafter in archival storage.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.hdf5 .txt binary	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☒ > 5 TB ☐ NA	

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.	
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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:
Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:

---

[3] Add rows for each dataset you want to describe.
[4] 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.	☐ Yes ☒ No If yes, please comment:
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). _RDM guidance on documentation and metadata_.	The output data is stored hdf5 format, which includes metadata, which describes the layout of the data formats and other relevant data attributes. The code used for the project is well documented on the publicly available wiki page: https://princetonuniversity.github.io/tristan-v2/index.html The wiki explains also the structure of the outputs and how they can be read.
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 yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: HDF5 If no, please specify (where appropriate per dataset or data type) which metadata will be created:

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

Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☒ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☐ Other:
How will the data be backed up? WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☐ Other (specify)
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 We expect that our production runs + test runs will amount to a maximum of ~10 TB of data at the end of the FWO. There is enough space on the Large Volume Storage to store this for long-term storage. During the FWO, we will use the shared network drive for data we are working on and the Large Volume Storage for data that is not used actively. We will strive to have a maximum of 2TB of data on the shared network drive at any time and move the rest to the Large Volume Storage. If no, please specify:

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.* *Guidance on security for research data*	Myself (Daniel Groselj) and the project suprevisor (Fabio Bacchini). Only these two persons will maintain regular access to the data, which requires relevant credentials (username + password + 2-factor verification).
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	We estimate to need only 500GB of memory on the shared network drive for the first year since it will be mainly dedicated to development and so not very data heavy. For the second and third year, we will purchase 2TB on the shared network drive. We will also increasingly need space on the Large Volume Storage as we accumulate data from simulations. We estimate 5TB for the second year and 10TB for the third year. Hence, we estimate the costs to be: 1st year : ~250 euros 2nd year : ~1,500 euros 3rd year : ~2,000 euros These expenses will be covered by the bench fee of 4,000 euros yearly that has already been approved by the FWO. We might ask for an extra 2,000 euros (over three years) to the FWO to ensure that the bench fee is enough to cover for travel and data storage.

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...). *Guidance on data preservation*	☐ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☒ Certain data cannot be kept for 10 years (explain) We will keep the data to a maximum of 5 years for budget issue. Moreover, with the always increasing performance of numerical codes and computing hardware, it is more costly (money-wise and environmentally/energy-wise) to store the data than to redo a simulation with a code and hardware whose performances might have doubled or tripled over 5 years. Of course, we will keep the configuration files of each simulations to ensure that we are able to rerun the exact same simulation. Those files are of the order of the KB.
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☒ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The purchase of 1TB of data on the Large Volume Storage for five years will cost 500 euros, which we will pay in advance with our bench fee provided by the FWO.

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, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify:
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.	☒ KU Leuven RDR ☐ Other data repository (specify) ☐ Other (specify)

When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
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.* Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
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, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No The data will have the DOI of the journal publication.
What are the expected costs for data sharing? How will these costs be covered?	None.

7. Responsibilities

Who will manage data documentation and metadata during the research project?	The FWO postdoctoral fellow – Daniel Groselj
Who will manage data storage and backup during the research project?	Daniel Groselj
Who will manage data preservation and sharing?	Daniel Groselj
Who will update and implement this DMP?	Daniel Groselj