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	Pablo Marchant, 0000-0002-0338-8181
Contributor name(s) (+ ORCID) & roles	
Project number[1] & title	Understanding mixing and chemical enrichment in massive stars
Funder(s) GrantID[2]	12ZY523N
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:
Please provide a short project description	As the observed sample of massive stars increases in size and quality, long standing problems in stellar astrophysics can now be addressed. One such problem is the so called Humphreys-Davidson limit, an observational feature of stellar populations that indicates that stars above a certain mass do not expand as expected from theoretical models. This has far reaching implications, impacting the feedback we expect from stars in their galactic environments, their properties before a potential supernova, and the way we think binary stars form gravitational wave sources. Recent work has provided a tantalizing solution to this problem: increased mixing in the stellar interior, beyond that predicted by the standard theory of stellar evolution, can prevent such stars from expanding beyond the observed limit. In this proposal I describe a method to use observations of chemically enriched stars to directly constrain these mixing processes, providing insight on the feasibility of this explanation of the Humphreys-Davidson limit as well as on the physical processes behind it. Owing to the inherent scarcity of massive stars, I will apply statistical methods to properly quantify the strength of these mixing processes in different stellar environments, as well as touch upon the consequences on our understanding of the formation of gravitational wave sources.

[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/

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].

				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
MESA simulation results	MESA is an open source code to model stellar evolution. Results are in the form of time series and structure profiles at given points in time.	X Generate new data ☐ Reuse existing data	X Digital ☐ Physical	☐ Observational ☐ Experimental ☐ Compiled/ aggregated data X Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml X other: Plain text machine readable tables ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB X < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

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

MESA simulation inputs	Files specifying inputs to all simulations performed in the project.	X Generate new data ☐ Reuse existing data	X Digital ☐ Physical	☐ Observational ☐ Experimental ☐ Compiled/ aggregated data X Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml X other: Plain text ☐ NA	X < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Software analysis tools	Code will be developed to analyze the result of large sets of simulations and released in an open source format.	X Generate new data ☐ Reuse existing data	X Digital ☐ Physical	☐ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data X Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml X other: Plain text ☐ NA	X < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 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).*	
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.	No existing data will be used
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 X No If yes, please describe:

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

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 X 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 X 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 X 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 X No If yes, please explain:

[6] See Glossary Flemish Standard Data Management Plan

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).	Simulation tools being used are designed to include proper metadata in output. Additional description will be provided in the form of README files to describe sets of simulations and their different inputs.
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 X No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: If no, please specify (where appropriate per dataset or data type) which metadata will be created: Simulation output of tools being used include human readable metadata in a self-explanatory format.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	During development of the project we will rely on the IT infrastructure of the institute of astronomy of KU Leuven, which is managed by Maarten Dirickx. This includes regular backups and sufficient storage space for all envisioned data products. For long term archival storage of data we will rely on the zenodo platform (zenodo.org), which is administered by CERN and allows free uploads of scientific data products.
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.*	During development of the project, internally using the IT infrastructure of the institute of astronomy of KU Leuven. As results are published all data products will be made publicly available and backed up using the zenodo platform.
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.	X Yes ☐ No If yes, please specify concisely: Storage space was guaranteed at the institute of astronomy during preparation of the project. If no, please specify:

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

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]	Data security during development of the project will be taken care of by the IT department of the institute of astronomy at KU Leuven. For long term archival storage access will be open, and data shared through the zenodo platform is made static after submission.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Costs are covered internally by the data infrastructure of the institute of astronomy of KU Leuven. Archival storage through the zenodo platform is free of cost.

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…).	All simulation outputs as well as inputs to reproduce the simulations will be made openly available through the zenodo infrastructure, which ensures their preservation (beyond the 5 year period).
Where will these data be archived (stored and curated for the long-term)?	The zenodo infrastructure that will be used by the project is archived and backed up together with the data from CERN, ensuring long term availability.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The zenodo infrastructure is free to use.

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*	X Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, 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.	Data access will not be restricted
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 X No If yes, please specify:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Data will be shared openly through the zenodo platform.

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'.*	In conjunction with any publication that makes use of the data.
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]	None, data coming from astrophysical simulations is not generally licensed.
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.*	X Yes ☐ No If yes: DOIs are automatically generated by the zenodo platform we will use.
What are the expected costs for data sharing? How will these costs be covered?	The zenodo platform we will use is free.

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

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Data documentation and metadata will be managed by me (Pablo Marchant) personally.
Who will manage data storage and backup during the research project?	During development of the project data will be stored and backed up using the IT insfrastructure of the institute of astronomy of KU Leuven.
Who will manage data preservation and sharing?	For long-term archival purposes I (Pablo Marchant) will be resonsible of uploading the data to the zenodo platform.
Who will update and implement this DMP?	Myself (Pablo Marchant)