Revolutionizing spectral synthesis: From isolated massive stars to unresolved stellar populations in the Early Universe

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.

Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)
		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, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA
Stellar Evolution Models	I will calculate an extensive set of detailed stellar evolution models that is the basis of the spectral synthesis code.	Generate new data	Digital	Simulation data	.txt	<2TB
Stellar Atmosphere Models	Based on the abovementioned evolutionary models, I will calculate a consistent set of stellar atmosphere models.	Generate new data	Digital	Simulation data	.txt	<1TB
Spectral Synthesis Code	I will combine the data in a new tool that creates synthetic spectra of stellar populations.	Generate new data	Digital	Software	.py, .cpp, .exe	<2GB
Stellar Atmosphere Fitting tool	I will use the stellar atmosphere models to create a fitting tool	Generate new data	Digital	Software	.py, .cpp, .exe	<2GB
Spectroscopic Data of Stars and Galaxies	I will test my software using spectroscopic data acquired with various telescopes using their archival data.	Reuse existing data	Digital	Observational	.fits	<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:

The spectroscopic data used in this work will be downloaded from the various instrument's archives:
http://archive.eso.org/eso/eso_archive_main.html
https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html
https://hla.stsci.edu/

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.

- Yes

The used spectra will be from public archives. The employed public archival astronomical data has to be properly acknowledged with citations of scientific publications. The use of public software tools such as MESA and PoWR has to be acknowledged with citations of scientific publications.

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 information on how the stellar evolution models as well as the stellar atmosphere models were computed will be documented for each model individually, as well as in papers associated with the publication. A README.txt file will be provided to explain the structures of the saved models, such that it is easy to find a specific model.
For the codes developed during the project, the code will be commented on and a README.txt will be created that explains the fundamental concepts of the code and the different scripts needed to execute it.
The used astronomical observations already include a metadata description of the observational strategy, program description, and instrumental set-up in the .fits headers.

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

All observations used will be available from the dedicated archives of the observatories
and use the astronomical metadata standard (.fits).
Databases of model outputs will have a description of the different parameters in the model
headers (README.txt and a description of the different parameters in the file headers, as is the
standard for when using MESA).

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

Where will the data be stored?

The newly calculated grids will be stored locally on personal and networked storage devices hosted by the Institute of Astronomy of KU Leuven. Input files used to create the stellar evolution models will be made publicly available on Zenodo.
The source codes of the new tools that will be developed at the end of the project will be made publicly available via GitHub.
The used spectroscopic data will be stored on the archives of the different observatories and is publicly available.

How will the data be backed up?

The Institute of Astronomy (IvS) has an automatic backup (daily and off-site) on the network server of raw and processed data according to KU Leuven security standards. The data is saved periodically on a daily and off-site basis. The data is secured with access restrictions on the file-system level. The backup and recovery procedures are handled by the IvS's IT team.
Observational data is archived and backed up by the observatories.

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

The IvS has a large volume storage server system with very large capacities (~1PB). The data created during the project will be only a small fraction of the available storage capacities.

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

The data is secured with access restrictions on the file-system level which is maintained by the Institute of Astronomy's IT team.

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

The data will be saved on the network servers of the IvS which is covered by shared funds of the institute.

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 at least five years.
Newly created data will be stored on the IvS servers, while archival data will be preserved on the publicly available archives of the different observatories.

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

All data will be stored long-term on the in-house storage system of the Institute of Astronomy of the KU Leuven. Additionally, the input and source files are made publicly available and will be stored on publicly accessible archiving services like Zenodo and GitHub.

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

Given the small size of the data created during the project compared to the increasing storage capacities, the estimated costs for preservation are low and will be covered by shared funds of the institute. The uploaded input and source files are small and are expected to be uploaded free of charge to publicly available archiving services.

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

Input files needed to recreate the produced science products will be made available on Zenodo. Source files of the code will be made available on GitHub.
Output data from the stellar evolution models and stellar atmosphere models will be made available upon request.

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

N/A

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.

Input files for publicly available codes and the software's source codes will be made publicly available on Zenodo and GitHub.

When will the data be made available?

The data will be made available after the publication of the research results.

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

Software datasets produced during the project will be published under the GNU General Public License version 3.0.
Other types of datasets will be published under the Creative Commons Attribution 4.0 International license.

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

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

Given the small size of the publicly available data no additional costs are expected, besides journal publication costs, which are already included in the project budget.

6. Responsibilities

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

The researcher.

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

The researcher with assistance from the Institute of Astronomy's IT team.

Who will manage data preservation and sharing?

The researcher during the project, and the supervisor with assistance from the Institute of Astronomy's IT team afterwards.

Who will update and implement this DMP?

The researcher during the project, and the supervisor with assistance from the Institute of Astronomy's IT team afterwards.