SOUL Methusalem

A Data Management Plan created using DMPonline.be

Creators: Clio Gielen, Pablo Huijse https://orcid.org/0000-0003-3541-1697, n.n. n.n.

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: n.n. n.n.

Project Administrator: Rose Stanley https://orcid.org/0000-0002-7822-3234

Grant number / URL: METH/24/012

ID: 207129

Start date: 01-01-2024

End date: 31-12-2030

Project abstract:

The long-term Methusalem project SOUL is situated in the field of astrophysics. Its global aim is the development of novel three-dimensional (3D) models for close binary and star-planet systems. We will develop computational tools to simulate the evolution of such systems, from birth to death. Our 3D evolutionary models will be unique in that they will maintain consistency between the physical and chemical ingredients of the stellar interiors, atmospheres, and outflows. Additionally, our models will have asteroseismically calibrated rotation and magnetic field profiles inside the stars, while also considering tidal interactions between the two components. The outflows and mass and angular momentum exchange of the models will meet observations of the highest precision. Our open-source SOUL deliverables will have major impact for the whole of astrophysics, including gravitational wave physics and exoplanet research.

Last modified: 02-07-2024

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 / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
WP1.1	Asteroseismic models of close binary pulsators subject to tidal forces	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP1.2	Asteroseismically-calibrated models for the transport processes in exoplanet hosts	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP1.3	Code to perform asteroseismically-calibrated 3D simulations of tidal dissipation	New	Digital	Software	.c, .cpp, .f90**	<100GB	
WP2.1	3D theory of stellar flaring for single and binary stars	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP2.2	Multiplicity analysis of low-mass stars and brown dwarfs based on MARVEL spectroscopic data	New	Digital	Numerical	.txt, .csv, .fits, .hdf5*	<100GB	
WP2.3	Asteroseismic models of stellar coronal magnetic fields based on detected quasi-periodic pulsations from stellar flares	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP2.4	Stellar coronae models based on uniturbulance with tidal effects from companion	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP3.1	Integrated 3D chemo-dynamical model of cool evolved intermediate-mass binary stars	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP3.2	3D radiation-magnetohydrodynamic model of red supergiants and their predicted mass-losses	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP3.3	3D dynamical models of planetary atmospheres under winds from evolved host stars	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
WP4.1	Temporal evolution of the multiplicity properties for a statistically meaningful sample of massive binaries	New	Digital	Numerical	.txt, .csv, .fits, .hdf5*	<5TB	
WP4.2	Asterosesmically-calibrated binary population synthesis models and their chemical yields	New	Digital	Model	.txt, .csv, .fits, .hdf5*	<5TB	
	TESS photometry	Existing	Digital	Numerical	.fits		
	Kepler photometry	Existing	Digital	Numerical	.fits		
	Gaia photometry/spectroscopy	Existing	Digital	Numerical	.fits		
	MARVEL spectroscopy	Existing	Digital	Numerical	.fits		
	PLATO photometry	Existing	Digital	Numerical	.fits		

*or other open file format for multidimensional data.

**or other high-performant low-level code

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:

- TESS archival photometric data is openly available at https://archive.stsci.edu/missions-and-data/tess
- Kepler archival photometric data is openly available at https://archive.stsci.edu/missions-and-data/kepler
- Gaia archival data is openly available at https://gea.esac.esa.int/archive/ for data release (DR) 3. DR4 data is expected by the spring of 2026.
- MARVEL data is not yet available. Expected to begin operations during 2025
- PLATO data is not yet available. Expected to begin operations in 2027

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.

- No

Will you process personal data? 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).

- 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 or 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

Software codes and archival observations are free to use but need to be acknowledged and/or cited in publications following the instructions of the original authors/source.

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

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).

Uploads to code and data repositories will be accompanied by README files providing detailed instructions for usage, installation, contribution and citation. Code developed during the project will be commented inline when appropriate. The addition of examples and interactive notebooks will be evaluated as they facilitate adoption considerably.

Shared data products will be accompanied by relevant keywords and metadata. Open data repositories, such as the Zenodo platform, support the mandatory and recommended terms of the DataCite and allow users to search by metadata terms. Tables and databases will include detailed descriptions of their columns and values.
All newly developed source code materials for simulations and numerical models will be described in detail in open-access journal publications. Journal publications will use dedicated journal standards (e.g. keywords) to identify research context.

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.

- Yes

Interoperability and data reuse will be facilitated by using metadata and technical data formats that are non-proprietary and widely adopted by the community. As a general rule, files will be stored unencrypted. The DataCite scheme for metadata will be used when possible.
For technical data, preference will be given to open hierarchical data formats that allow for the inclusion of metadata (self-describing), such as the Flexible Image Transport System (FITS) format, which is commonly used in astronomy, or the HDF5 format, which is widely used by the scientific community. The FITS format has the advantage of being readily compatible with the Strasbourg Astronomical Data Center (CDS) public repository. All these formats maintain official C, C++, Fortran and Java APIs to read and manipulate data. Support for Python, Rust, Julia, MATLAB and many other languages is provided by a collection of well-established third-party public libraries.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Personal network drive (I-drive)
- Other (specify below)

All material and codes will be stored on the shared internal network of the Institute of Astronomy and Centre of Plasma Astrophysics of KU Leuven.
The source codes of the simulators and tools developed during the project will be version-controlled using git and stored on private Github or Gitlab repositories. The repositories will be made public upon publication of the relevant papers.

How will the data be backed up?

- Other (specify below)

Continuous data preservation (beyond the 10-year requirement) is possible thanks to the well-developed in-house storage and backup system of the Institute of Astronomy of KU Leuven.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

The Institute of Astronomy of KU Leuven has an automatic periodic backup (daily and off-site) on the network server of raw and processed data according to KU Leuven and in-house security standards: all data are secured with access restrictions on file-system level.

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

Data storage and backup needs are estimated at 1500TB, this totals to 120keuro.
These costs are included in the granted SOUL budget.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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 preserved for 10 years according to KU Leuven RDM policy

All the data generated during the project will be preserved for 10 years or longer. This is possible thanks to the well-developed in-house storage and backup system of the Institute of Astronomy.

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

- Other (specify below)

All data will be stored long-term on the Institute of Astronomy's backup system (onsite and offline).
Additionally public repositories such as Zenodo, CDS and the virtual observatory will be used for the long-term storage of numerical models/simulations and final software releases. Archives in these platforms are projected to be maintained for at least twenty years. Metadata will be included to facilitate discovery by prospective users through these open platforms.

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

The costs for backup and storage are included in the granted project budget and the general working budget of the Institute of Astronomy. The aforementioned external repositories are free of charge.

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.

- Yes, as open data

All the data generated during the project, i.e. numerical simulations, software tools, databases, etc. and their corresponding metadata will be publicly available upon publication of associated research results. All journal publications will be open access via the ArXiv.org and Lirias servers.
To facilitate re-usability by the public, the following licenses are considered:

- Computational codes will be open-source and licensed using the GNU General Public License v3 or more permissive licenses.
- Databases and simulations will be licensed using Creative Commons Attribution-ShareAlike (CC-BY-SA).
- Accompanying metadata and documentation will be licensed using Creative Commons Zero (CC0).

In principle, we aim to maximally facilitate adoption by the scientific community at large, and in particular for those astrophysical projects that rely on modern numerical modelling.

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

NA

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.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)
- Other (specify below)

All the codes will be hosted in Github or Gitlab repositories and will be accessible to the general public after the associated research results are published.
The outputs of the models, final release versions of software and relevant tables/databases will be submitted to open online repositories such as the Virtual Observatory (ivoa.net), Zenodo (zenodo.org) and the CDS (https://cds.unistra.fr/data-publication/). These repositories are browsable, facilitating discoverability by prospective users. Additionally, they are highly interoperable and can be accessed directly from their websites or programmatically using APIs.
All journal publications will be open access via the ArXiv.org and KU Leuven Lirias servers.

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- GNU GPL-3.0 (code)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

Journal publications will receive a DOI. Large data sets and selected software releases that are published on the CDS/Zenodo data center will also receive a dedicated DOI, complementary to the article DOI.

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

Publication costs are budgeted in the project. There are no associated costs with making the data available on the Virtual Observatory, Zenodo and CDS servers.

Responsibilities

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

The PIs and team members will be responsible for guaranteeing data quality, documentation, storage, and backup in collaboration with the IT team of the Institute of Astronomy of KU Leuven.

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

The PIs and team members will be responsible for the garanteeing data quality, documentation, storage and backup, in collaboration with the IT team of the Institute of Astronomy of KU Leuven.

Who will manage data preservation and sharing?

The PIs and team members will be responsible for uploading the data to the relevant back-up and online sites. The IT team of the Institute of Astronomy of KU Leuven will be responsible for the continued back-up after the project.

Who will update and implement this DMP?

The PIs, with assistance from the institute's project coordinator, are ultimately responsible for updating and implementing this DMP.