Lifting stellar structure and evolution to higher dimensions in the era of space asteroseismology
ERC DMP +

Project information

Project Acronym

4D-STAR

Project Number

101071505

Data summary

Summary

Fulfilling the objectives of the 4D-STAR project requires developing new scientific software, collecting existing data, and generating new data products. The expected data and software products are described in what follows.

- **Asteroseismic calibration dataset:** A catalog of thousands of single and detached eclipsing binary pulsators from existing public data sources, including time-series photometry from Kepler and TESS missions, Gaia astrometry, and ground-based spectroscopy. This catalog will be used to obtain new asteroseismic calibration as described in WP2. Data is in FITS format and the expected size is less than 100GB.
- **Clustering modeling dataset:** A dataset of public photometry from the Hubble Space Telescope crossed with surface abundances from published spectroscopy. This catalog will be used to model star clusters as described in WP3. Data is in FITS format and the expected size is less than 100GB.
- **3D stellar oscillation software:** A software package for magnetohydrodynamical oscillation in 3 spatial dimensions to produce simulations of 3D waves as described in WP4, extending existing software that does not consider magnetism. The resulting simulations will be the input of the evolution models in WP5. The software will be released using open-source licenses and made publicly available with no restrictions. The package's core will be written in low-level languages (e.g. *.c/*.cpp/.f90) to ensure high performance. The expected size of this package will be 100MB or less.
- **3+1D stellar evolution software:** A newly developed software package that can describe the evolution of a star in three spatial dimensions and one time dimension. This corresponds to the final result of WP5, which is based on the novel theory of WP1, calibration results from WP2, and cluster modeling from WP3. Currently, no other software with these capabilities exists, hence we expect that this newly developed software will be an enormous contribution to the scientific community and for the further advancement of the astrophysical theories. The software will be released using open-source licenses and made publicly available with no restrictions. The package's core will be written in low-level languages (e.g. *.c/*.cpp/*.f90) to ensure high performance. Interfaces with high-level languages and notebook examples (e.g. .py,* .ipynb) will also be considered to facilitate usability. The expected size of this package will be 100MB or less.
- **Training and outreach material:** New multimedia material produced for the expected summer schools and online video training of the 4D-STAR software (e.g. *.pdf/*.mp4/*.ogg). This corresponds to the result of WP6. This material will be released under the CC0 license, published on the 4D-STAR website, and in relevant broadcasting platforms.

FAIR data

1. Making data findable

How can users find the data?

Prospective users can find, track and cite the data products listed above through their unique and persistent Digital Object Identifier (DOI). Datasets and software releases will be uploaded to Zenodo, automatically granting them a DOI and a version number. Zenodo is an open and free repository hosted by CERN and operated by OpenAIRE.

Uniquely identified data products will be accompanied by relevant keywords and metadata following the DataCite schema (https://schema.datacite.org/). The Zenodo platform supports the mandatory and recommended terms of DataCite and allows users to search by

metadata terms. Additionally, the metadata can be exported to other schemas, such as Dublin core and MARCXML.

2. Making data openly accessible

Which data products will made openly available?

All the data and research software listed before will be made publicly available.

Where can users find the data?

Databases and their corresponding metadata will be stored in Zenodo. Research software and their documentation will be version-controlled and stored in public GitHub repositories. Software releases will also be stored in Zenodo through the existing GitHub-zenodo interfaces. The usage of additional repositories, such as the institutional KU Leuven Research Data Repository, will be evaluated as the project progresses. Additionally, the public availability of the dataset and software will be announced on the 4D-STAR project website.

How can users access the data?

No registration or special tools are needed to access data deposited in Zenodo and users can download the files through their browser. Users can download the data programmatically using Zenodo's API if they prefer. The 4D-STAR website will also maintain a concise index of the project's relevant data and software products.

3. Making data interoperable

Which standards will be used?

Interoperability and data reuse will be facilitated by using technical file formats that are non-proprietary and based on open standards that are widely adopted by the scientific community in general or the astronomical community in particular. As a general rule, files will be stored unencrypted.

For example, multidimensional arrays will be stored using the Flexible Image Transport System (FITS) format, a standard specifically designed for astronomical data incorporating a human-readable metadata header. Libraries to read and manipulate FITS files can be found in C, Fortran, Java, Python, and many other programming languages.

4. Increase data re-use

What data/software will remain reusable, and for how long?

All the research software developed during the execution of the 4D-STAR project, i.e., entirely new software or extensions of existing software, aims to be publicly and widely used by the scientific community. Users can freely obtain the latest version of the source codes from the project's Github repositories or specific releases from the Zenodo archives.

All the data collected or generated during the execution of the project will be made publicly available through Zenodo, at the very least, at the moment relevant publications are published. Software and data will remain reusable beyond the duration of the project. Archives in Zenodo are projected to be maintained at least for the next twenty years.

What licenses will be used?

- The software will be released open-source and licensed under GNU General Public License 3 or more permissive licenses (e.g., MIT).
- Datasets will be licensed under Creative Commons Attribution-NonCommercial-ShareAlike (CC-BY-NC-SA), and metadata will be licensed under CC0.
- Outreach material and video tutorials will be licensed under CC0.

How is quality assured?

To facilitate usage and adoption, datasets and software will include user documentation, citation instructions and examples in interactive notebooks. Integration with package managers will also be considered to facilitate the installation of the 4D-STAR software. To ensure quality, software products will include development/deployment documentation, contribution guidelines, and test suites.

5. Allocation of resources and data security

What are the costs of making the data open, and how is data secured for long-term preservation?

As mentioned, data and software releases will be stored in the Zenodo repository at the CERN data center. The lifetime of the repositories is equivalent to the lifetime of its host institution, whose program is defined for at least the next twenty years. This repository is free of charge, and although the maximum file size is 50GB, there is no restriction on the number of records. In summary, no additional costs are expected to make the data open or for its long-term preservation. Files in Zenodo are backed up in multiple replicas on a nightly basis. No special security provisions are needed as no sensitive or personal data exists in the databases or software products.

How are the DMP responsibilities distributed across the collaboration?

The responsibilities of the day-to-day implementation and future adjustments to the DMP fall on Pablo Huijse, computational support postdoc at the KU Leuven site. Documentation, metadata generation, and data archiving will be carried out by the computational support postdoc in collaboration with the respective WP responsible. Long-term preservation, beyond the duration of the project, will be managed by the PIs.