From photons to 1 m/s precision radial velocities of exo-worlds with the MARVEL facility

A Data Management Plan created using DMPonline.be

Creator: Mariam Haidar

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 11PRP24N

ID: 206498

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Ultra high resolution spectroscopy is revealing unprecedented dynamical, atmospheric, and chemical composition information of both exoplanets and their stellar companions. And while facilities exist to provide pointed radial velocity (RV) follow-up on exoplanets detected via transit, a large sweeping survey has yet to be undertaken. The international MARVEL project will accomplish this with 1 m/s precision, exploiting an ultra stable, high resolution echelle spectrograph fiber fed by an array of 4 independent telescopes. Using a prism cross disperser, the telescopes can either work in unison focusing on the same star, or be used to observe four separate objects simultaneously.

This PhD project is fully embedded into the MARVEL project and I will focus on the final commissioning of the facility, and its scientific exploitation following first light in late 2024. I will investigate a novel wavelength calibration method involving a rubidium locked Fabry Pérot étalon, and integrate it into the MARVEL instrument. As part of the end-to-end instrument evaluation, I will identify, quantify, and mitigate sources of noise, in order to help the spectrograph reach 1 m/s precision. After commissioning, I will lead an RV follow-up of promising exoplanet candidates around F, G, K, or M type stars detected by the TESS satellite, demonstrating the scientific capabilities of MARVEL.

Last modified: 26-04-2024

From photons to 1 m/s precision radial velocities of exo-worlds with the MARVEL facility
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.

				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
		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	
Observational spectroscopic datasets	Observational spectroscopic data, design data, and characterization data collected by the MARVEL telescope array	Generate new data	Digital	Observational	.fits, .png, .csv	<1TB	
Python code	Python code used for simulations, data reduction, or measurement/statistics.	Generating new data	Digital	Other	.py, .ipynb	<1GB	
Simulation datasets	Data collected from simulations of the MARVEL telescope array performance	Generate new data	Digital	Simulation data	.fits	<1TB	
Statistical plots	Outputs of simulations, or derived from observational data	Generate new data	Digital	Compiled data	.png	<100GB	
Measurements, numerical results and statistics	Results of physical lab measurements, results, and statistics.	Generate new data	Digital	Experimental	.csv, .xml	<1GB	
Publications	Published scientific articles, proceedings, and other findings	Generate new data	Digital	Other	.pdf	<1GB	

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 is reused

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.

- No

This is not applicable to this project

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

This is not applicable to this project

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 observational spectroscopic datasets collected by the MARVEL array will be stored in the commonly used FITS format, with information regarding the day/time of acquisition, weather conditions, observational setup, size, resolution etc., and accessible by the public upon publication. During analysis and post-processing, the resulting files will be saved as either PNG files in the case of images and spectra, or as CSV files in the case of tables.

Computer code developed over the course of the project for data reduction, simulations, and processing will be documented and shared publicly via GitHub.

Observational spectroscopic data, design data, and characterization data will be stored within the MARVEL Archive on the KU Leuven's Institute of Astronomy (IoA) servers.

Image processed data (saved as PNG files): usually do not have standard metadata associated with them. The title generally includes a

summary of the techniques used to obtained them, and the explanation is given in the final publication files.
Measurements, Numerical results and Statistics. The CSV tables have explanatory title and comments, including content of the table, units of measurements, etc. For ASCII files that will be associated to a
publication and made available through the CDS SAGA database system, a README file will be provided.
Python codes usually have documentation in the file itself, like explanatory comments in code, procedure to use/run them etc.
Publication files (PDF) generally include date of submission, acceptance and publication.

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

STANDARD METADATA:
Astronomical data acquired from MARVEL observations: FITS header with information regarding the day/time of acquisition, weather conditions, observational setup, size, resolution etc.

NON STANDARD METADATA:
Image processed data (saved as PNG files): no METADATA is necessary in order to reuse them. All the information will be store in the image filename and in the publication description.
Measurements, Numerical results and Statistics. The CSV tables have explanatory title and comments, including content of the table, units of measurements, etc. For ASCII files that will be associated to a
publication and made available through the CDS SAGA database system, a README file will be provided.

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

Where will the data be stored?

Simulation data and lab measurement data, and observational data will be stored in the MARVEL archive within the KU Leuven's Institute of Astronomy servers.
Computer code developed over the course of the project for data reduction, simulations, and processing will be documented, stored and shared publicly via GitHub.
A copy will also be made and kept on a personal device.

How will the data be backed up?

KU Leuven's Institute of Astronomy servers have an automatic daily back-up procedure.
GitHub, where simulation code will be made available, also undergoes a daily back-up.
A copy will also be made and kept on a personal hard disk device.

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

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

The MARVEL archive stored within the KU Leuven Institute of Astronomy's server is only accessible by approved and authorized users, so there is no concern with unauthorized users accessing our data.

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

No costs are expected for data storage and back up during the projects as the facilities are already available at the host institute.
A small fraction of the first year bench fee will be used to purchase an external hard drive for a personal copy of the backup.

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...).

Observational spectroscopic data, design data, and characterization data, will be stored for at least 5 years.

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

Observational spectroscopic data, design data, and characterization data will be stored on both MARVEL telescope array servers and KU Leuven's Institute of Astronomy servers, which will remain in operation far beyond 5 years.

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

Costs for data preservation will be handled internally by the KU Leuven's Institute of Astronomy. Datasets for this project are not expected to incur significant costs for data preservation

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

Observational spectroscopic datasets collected by the MARVEL array will be stored in the commonly used .fits format, and released in an open access repository.
Code to produce simulation datasets will also be available as open access via GitHub.

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

For one year following our observations, data will subject to one year proprietary access by our MARVEL consortium members, so that we have time to work with and publish results using it. Following the proprietary period, observational data will become open access.

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.

Data will be made available without cost on CDS, Strasbourg astronomical Data Center (https://cds.unistra.fr/). This allows open access to astronomers and the public worldwide.
If published in a scientific article, it would be possible to access through the NASA's Astronomical Database System (https://ui.adsabs.harvard.edu/).
Code to produce simulation datasets will also be available as open access via GitHub.

When will the data be made available?

The observational spectroscopic datasets collected by the MARVEL array will be accessible by the public upon publication via databases such as NASA ADS (https://ui.adsabs.harvard.edu/), CDS (https://cds.unistra.fr/), and KU Leuven Lirias.
If not published within scientific articles, observational data will be made publicly available one year proprietary access following our observations on CDS.

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

Public Domain Mark (PD)

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?

Deposition of data for open access on CDS or ADS will be covered by the repository.

6. Responsibilities

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

Mariam Haidar

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

IT team of KU Leuven's Institute of Astronomy

Who will manage data preservation and sharing?

Hans Van Winckel

Who will update and implement this DMP?

Mariam Haidar