DMP TITLE

DMP for CELSA project CELSA/22/020

ADMIN DETAILS

Project Name: A robust and efficient framework for constrained optimization with complex, multi-physics simulation models
Project Identifier: CELSA multi-physics optimization
Grant Title: CELSA/22/020
Principal Investigator / Researcher: Martine Baelmans (KUL), Leon Kos (UL), Wouter Dekeyser (KUL)
Project Data Contact: Wouter Dekeyser, wouter.dekeyser@kuleuven.be
Description: Enabling efficient adjoint optimization for nuclear fusion divertor design with realistic plasma edge models and realistic design constraints requires fundamental progress to handle the interrelated sensitivities between multi-physics modules and constraints. The development of such optimization framework and the necessary numerical tools is the central objective of this proposal. With a successful proof-of-concept, we anticipate a wider adoption of the methodology in various related scientific disciplines.
Institution: KU Leuven, UL

1. GENERAL INFORMATION

Name applicant
Baelmans Martine (KU Leuven, Department of Mechanical Engineering, Applied Mechanics and Energy Conversion)
Kos Leon (UL, Department of Mechanical Engineering, LECAD Laboratory)
Dekeyser Wouter (KU Leuven, Department of Mechanical Engineering, Applied Mechanics and Energy Conversion)

CELSA Project Number & Title
CELSA/22/020
A robust and efficient framework for constrained optimization with complex, multi-physics simulation models

Affiliation
- KU Leuven
- UL

2. DATA DESCRIPTION

Will you generate/collect new data and/or make use of existing data?
- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

- Source code for simulation software: text files in syntax of programming language (Matlab, Python, C/C++, Fortran)
- Code documentation: text files & pdf files
- Research manuscripts and publications: LaTeX files & pdf files
- Simulation data: computed solutions using simulation software: Matlab data file format (.mat), hdf5, csv files, simulation-code-specific output formats
- Postprocessed data: figures & tables: pdf & text, along with scripts to reproduce from simulation data

Size of data is not an issue at all, and is directly determined by the parameters used in the computational methods and by the number of simulations. At most a few terabytes over the course of the project.

3. LEGAL AND ETHICAL ISSUES

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:
Short description of the kind of personal data that 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, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. DOCUMENTATION AND METADATA

What documentation will be provided to enable reuse of the data collected/generated in this project?

- Source code for simulation software: version control
- Code documentation: version control
- Research manuscripts and publications: version control and archiving of associated research data
- Simulation data & postprocessing data: document the experiments and their outcome for reproducibility (code version, scripts, data files). Use of standardized storage scripts for archiving of important data.

For version control of the simulation software, repositories on the KU Leuven gitlab server are already in place and used within the research group. We will also extend an (existing) continuous integration workflow that ensures software integrity as the software evolves. A reproducibility policy (keeping all scripts, data and postprocessing together to ensure reproducibility) is already in place.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

We use code-specific archiving procedures and will explore the iRods solution recently offered at KU Leuven. The format of the metadata is tailored to our specific needs of computational experiments.

5. DATA STORAGE AND BACKUP DURING THE CELSA PROJECT

Where will the data be stored?

- File synchronization (devices/collaborators) and backup: for version-controlled files, we use the KU Leuven gitlab service. For other files, we use KU Leuven OneDrive.
- Large simulation results are stored on a dedicated archive on the cluster.
- When possible, all relevant data will be made publicly available with the paper. For this, we will use a coupling betweeen iRods and the new KU Leuven RDR system that is designed for this purpose.
- Manuscripts and all corresponding data are stored on an archive drive in the research group and/or on an archive drive at the VSC depending on the nature of the research data.

How is backup of the data provided?

All solutions described above are equipped with automated back-up procedures that are provided by the central university services.

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

KU Leuven OneDrive services and dedicated archive storage on the VSC (presently 3TB archive storage available to the project) suffice to store the data.
Extensions to the archive can be obtained throughout the project via the VSC if needed.

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

Costs are covered by the standard KU Leuven internal ICTS rates and archive storage rates on the VSC cluster. We expect data management costs to be eligible costs from the working budget of research projects.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All systems described above allow configuration to prohibit access to unauthorized persons. Additionally, all data of computational experiments stored in iRods will be read-only, also avoiding accidental changes by authorized persons.

6. DATA PRESERVATION AFTER THE CELSA PROJECT

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All simulation code, simulation data, scripts and figures/tables that are needed to reproduce published results will be available in open access repositories and findable via a link in the publication.
Additionally, all data in archives and/or iRods will be retained for as long as useful (and certainly for the legally required time).

Where will the data be archived (= stored for the longer term)?

- Manuscripts and all corresponding data are stored on an archive drive in the research group and/or on an archive drive at the VSC depending on the nature of the research data.
- We are involved in a pilot to use iRods for this purpose.
- We are committed to making the data available in open access repositories (see above).

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Costs are covered by the standard KU Leuven internal ICTS rates and archive storage rates on the VSC cluster. We expect data management costs to be eligible costs from the working budget of research projects.

Archive storage at the VSC currently costs 100 EUR/TB/year (with 3TB available to the project).

7. DATA SHARING AND REUSE

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

- When possible, all relevant data will be made publicly available with the paper. For this, we will use a coupling between iRods and the new KU Leuven RDR system that is designed for this purpose.

Where/how will the data be made available for reuse?

- In an Open Access repository

At least in KU Leuven RDR. Potentially also data repositories that are coupled to the journal in which the publication appeared.

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

Full data is publicly available. Acknowledgement of will be required via license.

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

Costs are covered by the standard KU Leuven internal ICTS rates and archive storage rates on the VSC cluster. We expect data management costs to be eligible costs from the working budget of research projects.

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?

The data management plan will be executed by the individual researchers and compliance monitored by the PIs.

Who will be responsible for data storage & back up during the project?

The data management plan will be executed by the individual researchers and compliance monitored by the PIs.

Who will be responsible for ensuring data preservation and reuse ?

The data management plan will be executed by the individual researchers and compliance monitored by the PIs.

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsibility of updating & implementing this DMP.