12E3623N An efficient framework for reliable calibration and uncertainty estimation of novel self-consistent cross-field transport models for plasma edge simulations of fusion devices

A Data Management Plan created using DMPonline.be

Creators: Stefano Carli, First Name Surname, First Name Surname

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Stefano Carli

Project Administrator: First Name Surname, First Name Surname

Grant number / **URL:** 12E3623N

ID: 193731

Start date: 01-11-2022

End date: 31-10-2025

Last modified: 19-01-2023

12E3623N An efficient framework for reliable calibration and uncertainty estimation of novel self-consistent cross-field transport models for plasma edge simulations of fusion devices

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, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
solps code	Fortran source code for simulation	• Generate new data • Reuse existing data	digital	software	text files in syntax of programming language (typically .F or .F90)	<100MB	
postprocessing tools	Matlab/Python source code for postprocessing of simulation data and creation of figures/tables	• Generate new data • Reuse existing data	digital	software	text files in syntax of programming language (typically .m and .py)	<100MB	
code documentation	Text files describing code usage and code options	• Generate new data • Reuse existing data	digital	Other	Latex .tex and .pdf files	<100MB	
simulation data	input files and computed solutions using simulation software	• Generate new data	digital	Simulation data	output specific of solps, Matlab files (.mat), hdf5, csv	<1TB	
postprocessed data	figures and tables	• Generate new data	digital	Other	.eps, .png, .csv, .xls, .odp	<1GB	
experimental data	Experimental data from TCV, JET and ASDEX Upgrade	• Reuse existing data	digital	Experimental	.txt, .dat, .csv, magnetic equilibrium files .equ, vessel geometry .ogr	<1GB	
research output	Files for publications and conferences	• Generate new data	digital	Other	Latex files .tex, .pdf, powerpoint .ppt	<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:

Experimental data from TCV, JET and ASDEX Upgrade are stored on the local servers of such experimental devices

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

- SOLPS code: agreement with ITER does not allow dissemination/redistribution of the code developed
- Experimental data: JET data cannot be redistributed, TCV and ASDEX data potentially needs permissions for reditribution

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).

- solps code: version control on ITER repository or KU Leuven gitlab servers
- postprocessing tools: code comments, readme files, KU Leuven RDR repository for reproducibility
- Code documentation: version control on KU Leuven gitlab servers
- simulation data: document the simulations and their outcome for reproducibility (code version using git hash in inputs/outputs, scripts, data files, and workflow). Use of standardized storage scripts for archiving of important data. Store on KU Leuven RDR repository inputs and relevant outputs for reproducibility
- postprocessed data: readme files with descriptions, store on KU Leuven RDR
- experimental data: this data is stored on the specific experimental device's server. If data can be shared, then readme files and scripts for reproducibility will be placed on KU Leuven RDR repository
- research output: version control on KU Leuven gitlab servers and archiving of associated research data

Use of KU Leuven platform "ManGO" will be exploited.
An archive space for long term storage of data on the VSC cluster is already in place for the research group.

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.

- No

For KU Leuven RDR and ManGO the format of the metadata will be tailored to our specific needs of computational experiments.

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

Where will the data be stored?

- File synchronization (devices/collaborators) and backup: for version-controlled files, we use the KU Leuven gitlab service or ITER git repository. For other files, we use KU Leuven OneDrive.
- Large simulation results are stored on a dedicated archive on the VSC cluster.
- When possible, all relevant data will be made publicly available with the paper. For this, we will use a coupling betweeen ManGO 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 will the data be backed up?

All solutions described above are equipped with automated back-up procedures that are provided by the central university services.
Personal laptops of researchers are synced with their home directory on the departmental servers. These servers are backed up as per the above.

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.

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 ManGO will be read-only, also avoiding accidental changes by authorized persons.

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

We expect data management costs to be eligible costs from the working budget of this or other research projects within the research group.

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 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 ManGO and archives will be retained there for as long as useful (and certainly for the legally required time).

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

- Simulation inputs and relevant outputs, postprocessing tools will be published on the KU Leuven RDR repository in open access
- 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.

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

Publishing on RDR requires no costs.
We expect data management costs to be eligible costs from the working budget of this or other research projects within the research group.
Archive storage at the VSC currently costs 100 EUR/TB/year (with 3TB available to the project).

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

- SOLPS code: restricted access repository
- postprocessing tools: restricted access repository (solps postprocessing tools) and open access (additional postprocessing tools)
- code documentation: restricted access repository
- simulation data: open access repository
- postprocessed data: open access repository
- experimental data: restricted access repository
- research output: restricted access repository

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

- SOLPS code: accessible by other/future members of the research group as well as from any institution with an agreement with ITER Organization, contact Xavier Bonnin at ITER
- postprocessing tools withing solps: same as above
- code documentation: same as above
- experimental data: accessible by other/future member of the group within the EUROfusion framework
- research output: accessible by other group's members with the KU Leuven gitlab repository

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.

- Yes, Intellectual Property Rights

- SOLPS code: software is subject to IMAS agreement for use, contact Simon Pinches or Xavier Bonnin at ITER
- experimental data: JET data not shareable, data from ASDEX/TCV potentially sharable

Where will the data be made available? If already known, please provide a repository per dataset or data type.

- SOLPS code: ITER git repository, access to be granted by the institution
- postprocessing tools: either available within SOLPS code or made available in KU Leuven RDR open access repository together with the simulation data. also stored on the group's archive
- code documentation: either available within SOLPS code or in KU Leuven's gitlab repository
- simulation data: KU Leuven RDR open access repository for inputs and relevant outputs, also group's VSC archive
- postprocessed data: U Leuven RDR open access repository and group's archive
- experimental data: already stored by other institutions. possibly share together with postprocessing tools and simulation data on open access repository when linked to a publication and when sharing is allowed
- research output: gitlab repository of KU Leuven, group's archive

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.

Datasets: CC-BY-NC-SA-4.0
Code: AGPL-3.0-or-later

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?

Open access on RDR is free of charge for KU Leuven affiliates.
We expect other data management costs to be eligible costs from the working budget of research projects.

6. Responsibilities

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

Stefano Carli, compliance monitored by the supervisor

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

Stefano Carli, compliance monitored by the supervisor

Who will manage data preservation and sharing?

Stefano Carli. The supervisor and/or the responsible for data management in the group if/when Stefano Carli will eventually leave the research group.

Who will update and implement this DMP?

Stefano Carli. The supervisor and/or the responsible for data management in the group if/when Stefano Carli will eventually leave the research group.