Innovative digital twin concept of complex microstructure evolution in multi-component materials GDPR Record

GDPR record

Have you registered personal data processing activities for this project?

- No

Innovative digital twin concept of complex microstructure evolution in multi-component materials DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Innovative digital twin concept of complex microstructure evolution in multi-component materials ERC DMP +

Project information

Project Acronym

muTWIN

Project Number

101123107

Data summary

Summary

Following data types will be generated within the project
1) **Outcomes from phase-field simulations** used for training of the digital TWIN models

- The data will be generated as output of phase-field simulations using C++ MOOSE software (resulting in csv. data sets) and matlab (resulting in .mat files)
- The total amount of simulation data will be of the order of 0.5-2Tb

2) **Scripts for the computation of** the factor matrices representing the **TWIN models** based on the outcomes from phase-field simulations, including scripts to load and format the output of phase-field simulations, scripts to perform the least-squares optimization of the elements of the factor matrices and scripts for validation and visualization of the TWIN models. These scripts will bewritten in Matlab and use subroutines of the Matlab toolbox TensorLab (publicly available: http://www.tensorlab.net/)

- Matlab scripts based on and extending subroutines of the toolbox TensorLab. data type : .m
- All scripts together will have a size of the order of kb's.

Depending on the interest of companies, the patentability of the approach and possible further use of the scripts in follow-up projects, these data may not or only be made publicly available after a certain embargo time.
3) **Optimized digital TWIN models** of microstructure evolution in the form of sets of factor matrices from which the full tensor description can be calculated

- The data factor matrices will be generated using matlab routines
- Data format : .mat files (standard) and other data types for further use are possible
- A total size of the order of kb's up to maximum a few Mb is expected

Depending on interest of companies and possible further use of the data in follow-up projects, these data may not or only be made publicly available after a certain embargo time.
4) **Documents and reports** describing the methodologies, their validation and accuracy and applications

- The name of the documents will reveal the aim, version and date of the document
- Data type: Word, LaTeX and ppt, pdf documents
- A total size of max 1 Mb is expected

Depending on interest of companies, the patentability and possible further use of the approaches in follow-up projects, part of the reports may not or only be made publicly available after a certain embargo time.

FAIR data

1. Making data findable

1) The **raw simulation data from phase-field simulations** will be collected in separate folders per simulation test including a txt file with a clear description of what the data represent and how they were generated. The input-files used for the simulation will be kept insight the same folder. The name of the folder will contain the composition, temperature and a reference to the loading conditions of the considered material. A .txt file explaining the naming will be maintained and kept together with the data sets. Reduced key datasets underlying a publication (up to 50 GB) will be made available through Mendeley with a mentioning that the full data set can be accessed upon request. DOIs will be appointed to these key datasets.

2) The **scripts implementing the calculation of the tensor TWIN models** will get unique names referring to the procedure they implement. An extensive explanation of the input and output variables and the outline of the program will be given at the top of the file. The new procedures will be documented in the reports (data set 4) and a clear reference to the report name and version will be given in the description in the matlab file (and vice versa, the supporting documents will be clearly mentioned in the scripts). The procedures to calculate the tensor models will initially not be made publicly available until a decision has been made about the patentability of the approach. In a later stage they may still be made publicly available through Mendeley with a manual describing the approaches and how to apply them on output from phase-field simulations (possibly, after negotiation with the developers of the TensorLab toolbox (http://www.tensorlab.net/), they may also be added as part of the TensorLab toolbox and described in the on-line manual). When the procedures are publicly published, an associated journal paper will also be published refering to the data set on Mendeley (and or the function in the TensorLab toolbox). When published on Mendeley, a doi will be appointed to the whole set of scripts.

3) **The digital TWIN models** of microstructure evolution consist of numerical data represented in the form of matrices ('the factor matrices of the tensor decomposition').

They will be saved in folders per simulation test following the same structure as those used for the raw simulation data 1) and using file names consistent with those used for the raw simulation data. The tensor descriptions will be proceded by metadata containing information on the elements included, the temperature and composition range and the name of the procedure used to obtain them. A procedure to automatically generate these metadata will be implemented. The models that can be published will be published together with the key data set of raw materials (on Mendeley and hence recieve a unique doi)

4) When possible, reports will be published in the form of journal papers. They will be open access true the gold or green route (by uploading the postprint in LIRIAS, the open access repository of KU Leuven). For presentations, a pdf will be uploaded in LIRIAS.

In all cases, the METADATA will also contain links to relevant publications and manuals, as well as a relevant contacting address (nele.moelans@kuleuven.be).

2. Making data openly accessible

1) The simulation data themselves will be saved on archive space at the VSC-KU Leuven (https://icts.kuleuven.be/sc/onderzoeksgegevens/HPC-storage). Because of their huge size, it is difficult to find an appropriate place to make them publicly available. We will give people access upon request. When needed, we can put the data temporarily on a place with fast access. We will make them available under a CC-BY license. Reduced key datasets underlying a publication (up to 50 GB) will be made available through Mendely under CC-BY license (or in a repository advised by the journal if possible).

2) Depending on the interest of companies, the patentability of the approach and possible further use of the scripts in follow-projects, these data may not or only be made publicly available after a certain embargo time. the scripts will be saved in a joint folder (on the KU Leuven institutional drives) to which a limited number of people have access. The scripts can be used within Matlab and using the TensorLab toolbox (open source toolbox: http://www.tensorlab.net/license.txt)

3) Depending on the interest of companies, the patentability of the approach and possible further use in follow-projects, of the data representing the tensor models may not or only be made publicly available after a certain embargo time. The data will be saved in a joint folder to which a limited number of people have access. They will be saved in different format and hence can be used in various codes.

4) Also the reports and presentations will be stored in the joint folder on a KU Leuven drive with limited access, until they can be published.

When 2) ,3),4) can be made publicly available, this will be done through Mendeley, most likely under CC-BY license.

3. Making data interoperable

1) The raw data will be saved using csv format, which can be loaded by most simulation and visualization software.

2) The scripts for calculating the tensor TWIN models are generally applicable to microstructure data sets. The scripts use MATLAB in combination with the open source Matlab toolbox TensorLab (http://www.tensorlab.net/license.txt).

3) The datasets representing the tensor TWIN models are limited in size and will be stored using different file formats. They can be used within any software. Example scripts (matlab, c++) will be provided showing how to use the data.

4) Reports and presentations will be made using word, powerpoint and LaTeX, but a .pdf of all files will be stored as well.

4. Increase data re-use

1) The large raw data sets will be available until at least 10 years after the end of the project under CC-BY license. The selected data sets stored in a Mendeley repository will be permanently available under CC-BY. A csv file format is used which can be loaded by most simulation and visualization software.

2)+3)+4) Depending on the interest of companies, the patentability of the approach and possible further use of the scripts in follow-projects, these data may not or only be made publicly available after a certain embargo time or under restricted licensing. When publication of the data is allowed, they will be published through Mendeley and remain permanently available. The computed TWIN models are system specific, but the scripts to compute the models are generally applicable to microstructure evolution data for any type of material.

Publications will as much as possible refer to the open data and their locations. We will make sure that a possible embargo time for the publications is not longer than 6 months. We will advertise the publications and data locations on the group website and at conference presentations

5. Allocation of resources and data security

Most of the material, except for the raw simulation data, can be saved and made publicly available at now cost (through the Mendeley repository).

The raw simulation data (as well as a copy of all the other data generated within the project) will be archived on the VSC-KU Leuven (https://icts.kuleuven.be/sc/onderzoeksgegevens/HPC-storage) at a price of 20 euro per Terrabyte per year. The price also includes 'staging' space, to where the data can be copied (relatively fast) temporarily to use it in computations. It is possible to give external researchers a temporary vsc account so that they can access and use these data on the VSC, avoiding the time and difficulties associated with copying a large data set to external infrastructure. We expect that by the end of the project, we will need approx. 2-3 Terrabyte, i.e. a cost of 60 euro per year. During the project, this cost will be booked on the project. Since this project will most probably feed into future related projects, we will foresee budget allocation in these projects for long-term storage of the data.

The data 2)+3)+4) will also be stored on the local institutional drive (Storage facilities of the research unit) which is backed up. The foreseen disk space will be connected to the research group and the personnel number of prof. Nele Moelans and is at no cost. A copy of all the data will also be kept at the VSC-KU Leuven archive/staging space, which is backed-up daily.

The PI, Nele Moelans, will be responsible for implementing the DMP and assures that all data will be saved until at least 10 years after the end date of the project according to KU Leuven data management rules. She will update the DMP anytime conditions change. A final reviewed DMP will be sent along with the final report.