Plan Overview

A Data Management Plan created using DMPonline

Title: MatFlow: a computational workflow management platform for reproducible digital materials research

Creator: Joao Quinta da Fonseca

Principal Investigator: Joao Quinta da Fonseca

Affiliation: University of Manchester

Template: EPSRC Data Management Plan

ORCID iD: 0000-0001-6063-8135

Project abstract:

MatFlow is an open-source Python package for designing, running, and sharing reproducible hybrid computational workflows in materials science and digital manufacturing. We plan to develop new functionality, develop comprehensive user/developer documentation and introduce MatFlow to new user communities. Currently, MatFlow can only be used at the University of Manchester, but by the end of the project it will be possible to install and run it in other institutions and businesses. Digital manufacturing relies on an ecosystem of existing modelling and data processing tools, however we lack a framework for creating and sharing transparent and reproducible workflows that seamlessly connect these packages. With the extension ecosystem of MatFlow, researchers can efficiently develop simulation and data analysis pipelines, regardless of their expertise in programming or high-performance computing, and without having to re-write code that is often poor quality and lacks community oversight. MatFlow is designed to enable researchers to focus on developing the science and facilitating and accelerating its impact.

ID: 85006

Start date: 01-04-2022

Last modified: 28-09-2021

MatFlow: a computational workflow management platform for reproducible digital materials research

UoM Project Details

Proposal title

MatFlow: a computational workflow management platform for reproducible digital materials research

Is the project already funded?

- No

Will the project make use of data (please select all that apply)?

- Acquire new data
- Re-use existing data (please list below)

Where will the data be stored and backed-up during the project lifetime?

- Other repository or storage system (please provide details below)

Data will be stored on Zenodo

How much data storage will you require?

- < 8TB

Data Collection

What data will you collect or create?

Experimental and computational simulation data:
- text files

- binary filkes
- images
More detail to be added later.

How will the data be collected or created?

Several techniques.

Documentation and Metadata

What documentation and metadata will accompany the data?

We are going to use at the beginning a simple set of metadata which will be the bare minimum needed by Zenodo data repository. We are going also to define a set of metadata specific to the project but following as much as possible metadata standard with eventually additional informations needed for the specificities of the project.

Ethics and Legal Compliance

How will you manage any ethical issues?

There not ethical issue for this project.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Individual agreements with collabortators.

Storage and Backup

How will the data be stored and backed up during the research?

During data processing analysis the data will be stored in the Manchester RDS.
Once completed, data and analysis will be uploaded after creation and a first curation to be sure that they are consistent and with the proper metadata to the data repository (Zenodo). The upload will be eased by the usage of a dedicated software which will verify the presence of the needed metadata.

How will you manage access and security?

Most of the data will be place under open data license. For the one which will need to have an embargo or to be closed, Zenodo is providing the necessary tool to allows that. Sharing can be done specifically by asking the data creators through Zenodo.
Security is provided by Zenodo.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Most of the data created by the project will be kept for future usage or to be use by other members of the collaboration. Zenodo is providing the space which allows that conservation.

What is the long-term preservation plan for the dataset?

Zenodo via the community facility

Data Sharing

How will you share the data?

Data will be shared between the members of the project as soon as they are produced. At the end of the project most of the data under embargo will be shared under an open data license. Some data, coming from industrial partners, could be keep closed but with a contact person if needed.

Are any restrictions on data sharing required?

For some porojects with substantial industrial funding, yes.

Responsibilities and Resources

Who will be responsible for data management?

Joao Fonseca and Project RSE

What resources will you require to deliver your plan?

RSE time and intermediate data storage (RDS).