Data Management Plan - 1S66623N

Project Name DMP for FWO-SB - Muhammad Hafeez Saeed - Data Management Plan - 1S66623N
Principal Investigator / Researcher Muhammad Hafeez Saeed
Institution KU Leuven

1. General Information

Name applicant

Muhammad Hafeez Saeed

FWO Project Number & Title

File number: 1S66623N

Title of FWO-SB Ph.D: Stochastic control of stacked flexibility in district-level multi-energy systems (MES)

Affiliation

- KU Leuven

ESAT/ELECTA

2. Data description

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

- Generate new 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).

Type of data	Format	Volume	How created
Detailed models from Modelica software	.fmu	1-5 GB	Physics-assisted models developed/used for buildings and districts, in Modelica language-based software i.e. OpenModelica
Hyper-parameters of the trained machine-learning models	.json	10 kB - 10 MB	The reinforcement learning agent that is trained for controlling a multi-energy system can involve neural network modules. The .json files contain hyper-parameters of such neural network modules.
Simulation results	.csv	2-10 GB	Data generated due to control simulations run in python
Simulation images	.png	5-10 GB	Graph images showing the control signals for the devices of the buildings and districts, and compared with setpoints of the temperature or energy consumption
Textual output	.tex, .PDF	20-100 MB	Documented text for research papers, project reports, and the thesis manuscript in LaTeX

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?

- Yes

The research outcomes will be shared with scientific and general audiences
through regular publications. Sharing the district control algorithms and techniques for tech transfer will depend on the valorization potential of the research outputs, which is an important exercise during this Ph.D.

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?

1. The building and district models used from OpenModelica software will be documented properly through the built-in software documentation feature, and also a version history will be maintained for each iteration of model development.

2. Hyper-parameters of the trained machine learning models will be maintained in separate folders for each use-case that is used for simulating control of the districts. These files will be .json extensions and will contain the parameter values used to run the simulation.

3. Simulation results will be stored in separate folders for each simulation use-case. The results will be recorded in the form of .CSV files containing state values on each control step. Depending upon which energy carriers and which size of the cluster is considered, further sub-folders will be made. There will be a ReadMe file containing a crisp explanation of how to re-run the simulation and analysis.

4. Simulation images will be maintained in the respective folder and sub-folder of simulation use-cases, as mentioned in the above points.

5. Textual output will be stored in respective folders and sub-folders of use-cases, and an indication will be made in the title of the textual output according to the purpose: papers, specifications, clarifications, and the thesis manuscript. The files will carry the date of their generation to facilitate version management.

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

5. Data storage and backup during the FWO project

Where will the data be stored?

1. The main copy of the data will be stored in our research group's central storage facility. Copies can be made and kept on personal devices.

2. The open-source data used during the research will be made available on the GitHub repository for a wider research audience.

3. Since we will collaborate with researchers from other research units within KU Leuven and EnergyVille (energy research center), we will use EnergyVille OneDrive for active use of the data during the project.

How is backup of the data provided?

The data will be stored in the central storage servers of our research group (ELECTA/ESAT), which have automatic daily backup procedures. The data on the personal directories of all the users/researchers is backed-up each day. Moreover, periodic snapshots of personal directories are also kept, which contain information about the data that users have recently removed.

Two kinds of backups are considered for our research project:

1. Full backup

Every three months, a complete backup of the data on our department's servers is made.

2. Incremental backup

The everyday changes made in data since the last daily backup are stored on backup tapes.

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

The research project's data is not expected to exceed 20 GB, which is well within our department's storage and backup limits for a single user.

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

The separate cost calculation of data storage and backup for this project was not made, because they are covered within the general operating expenses of our research group.

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

All the user accounts on servers are password protected, which means only the account holder can access the working data.

6. Data preservation after the FWO 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, ...).

The data on the central server and EnergyVille SharePoint will remain for at least 5 years after the end of the project. The working data on the user's individual account is normally removed once the user leaves. Therefore, should there be any necessary working data that needs to be maintained, will be transferred to the main PI's user account before the end of the project.

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

Main PI's user account.

EnergyVille SharePoint.

Please refer to the backup strategies mentioned in section 5 above.

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

No additional costs are foreseen at the moment.

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?

Internally, the data will be available for the research group. Research findings and information will be made externally available through publications. The release of specific algorithms or information may depend upon their valorization potential, and it is an important exercise of

this research project.

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

- In a restricted access repository

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?
Anyone within the research group will be able to access the data.

What are the expected costs for data sharing? How will the costs be covered?
No costs are associated with data sharing.

8. Responsibilities

Who will be responsible for data documentation & metadata?
Co-workers on the project together with the PI, are responsible for data documentation and metadata.

Who will be responsible for data storage & back up during the project?
The responsible for data storage & backup is the IT responsible at ESAT, KU Leuven, Veronica Lucero Ortega.

Who will be responsible for ensuring data preservation and reuse ?
Co-workers on the project together with the PI, are responsible for data preservation and reuse.

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