DMP title

Project Name My plan (FWO DMP) - DMP title
Principal Investigator / Researcher Daniel Escudero
Institution KU Leuven

1. General Information

Name applicant
Daniel Escudero Masa

FWO Project Number & Title
UNRAVELING COMPLEX EXCITED STATES PROPERTIES FROM FIRST-PRINCIPLES (COMPLEXES)
G079122N

Affiliation

- KU Leuven

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).

The main datatypes generated by research of this type are:

1) Computer codes and scripts: For the research scheduled in this project most often we will use comercial and open source softwares. However, the Project involves up to some extent, coding and implementation efforts. The latter codes and scripts will be developed by the members of the group.

2) Inputs and outputs of computer simulations: The main datatype generated in this project will consist of inputs and outputs (as well as intermediate and temporary files) of the simulations performed using both comercial and open source softwares. The temporary file are often very voluminous and they are only useful to restart some failed simulations. These are often automatically deleted if the simulations are completed succesfully.

Key outputs and codes can be readily made available through the groups' webpages, through the Supporting Information of publications, and through the research repository of KU Leuven, LIRIAS, as the volume of data is relatively modest (5-10 GB). Full output from all the simulations is very voluminous, but the field has standard protocols for collecting the key data.

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?

During research, data will be stored on computers belonging to the Quantum Chemistry and Physical Chemistry Division (QCPC) and to the Flemish Supercomputing Centre.

Key data will always be backed up regularly to avoid loss of data (i.e., this is done on a daily basis in our cluster at the QCPC). Generally to unambiguously trace and organize data; the data arising from the simulations (inputs&outputs) 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. A common protocol to name and store data for this project will be set up among all project members. Scripts, codes, input&output files will unambiguosly display date of creation and who generated the data.

After research, as much as possible the data will be published through Supporting Information files and the Research Repository LIRIAS.

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

The applicant has considerable expertise in curating computational chemistry research data with these types of approach. All data (appropriately collected and classified with a clear description of what the data represent and how they were generated) will be preserved for at least 10 years (this is the standard protocol at the QCPC division.

5. Data storage and backup during the FWO project

Where will the data be stored?

Full output from all calculations is very voluminous, but the field has standard protocols for collecting the key data. Specifically, data will be stored in our computers belonging to the Quantum Chemistry and Physical Chemistry Division (QCPC, KU Leuven) and on the data storage services of the Flemish Supercomputing Centre (VSC).

How is backup of the data provided?

Automatic daily back-up procedures are ensured at both our QCPC cluster as well as at the VSC facilities.

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

Our data storage capabilities are currently on the TB scale, and therefore sufficiently large for the project needs.

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

No additional cost in our QCPC facilities. At VSC, for a modest user data storage, in line with our needs for this project, no additional costs are scheduled.

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

The data will be stored in the university's secure environment. Only authorized persons have access to the 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 will be stored on our QCPC facilities (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

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

The data will also be stored on our QCPC facilities

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

Given the current infrastructures which are totally sufficient for the data produced in this project, no extra costs are scheduled for data preservation

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?

Computer codes and scripts will be made available at e.g., GitHub and/or the jupyter notebook and through the groups' webpages, under CC-BY licenses. Relevant data will be collected at the supporting Information of publications, but also through the research repository of KU Leuven, LIRIAS.

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

- In an Open Access repository

When will the data be made available?

- Immediately after the end of the project
- Upon publication of the research results

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

We prefer to make our data available under a CC-BY license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creator.

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

No costs are expected for this

8. Responsibilities

Who will be responsible for data documentation & metadata?

The responsible person will be Prof. Escudero, the PI of this project

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

The responsible person will be Prof. Escudero, the PI of this project; together with Hans Vansweevelt, the responsible person for our computer cluster at the QCPC division

Who will be responsible for ensuring data preservation and reuse ?

The responsible person will be Prof. Escudero, the PI of this project

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

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