Initial DMP

Project Name Computational and Conceptual Density Functional Theory for Electronic Resonances (FWO DMP) - Initial DMP

Grant Title FWO Aspirant PhD Fellowship - 1103922N

Principal Investigator / Researcher Charlotte Titeca

Description The research project focusses on the description of temporary anions using conceptual and computational DFT, in terms of bonding, forces on the nuclei and energy. The generated data will mainly comprise computer code and software output.

Institution KU Leuven

1. General Information
Name applicant

Charlotte Titeca

FWO Project Number & Title

1103922N
Computational and Conceptual Density Functional Theory for Electronic Resonances.

Affiliation

- KU Leuven
- Vrije Universiteit Brussel

Joint PhD at KU Leuven (main institution) and VUB.

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).

New data will be generated throughout the project, it is not expected that any existing data will be reused.

The generated data consists mainly of:

Type of data	Format	How created
Input files for Q-Chem	text files	Manually created.
Output files of Q-Chem, some serve as input for AIMAll	text files	Generated by Q-Chem software.
Output files of AIMAll, some serve as input for Multiwfn	text files	Generated by AIMAll software.
Output files of Multiwfn	text files	Generated by Multiwfn software.
Tables, summarizing the results	.xlsx	Manually created, contains results from the various output files.
Images and graphs, saved as figures	.jpg .svg	Created using GaussView, VMD, AIMAll.

Upon creation of new data not yet reported in the table, the table will be updated. The total data

volume is estimated to not exceed 5 TB and the individual file volumes will not exceed 10 GB.

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: NA
Short description of the kind of personal data that will be used: NA

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

Part of the planned research consists of code development for Q-Chem, which may result in data with potential for tech transfer and valorisation. As this part of the research has not yet started, it is not yet possible to assess the potential IP restrictions. This will be discussed in due course and added to the DMP.

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?

- Yes

The code development for Q-Chem requires access to its source code. A Non-Disclosure Agreement (NDA) has been signed which prohibits dissemination of the source code or any other proprietary and confidential information disclosed by Q-Chem.

4. Documentation and metadata
What documentation will be provided to enable reuse of the data collected/generated in this project?
In the data folders, separate text files will be saved which contain information about the data and the methodology used to obtain the data. Such information includes:

- Overview of files in the corresponding data folders.
- Overview of performed calculations and their purpose.
- Settings for the calculations.

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.

- Yes

The data will be archived in KU Leuven's RDR repository. This repository provides an appropriate metadata standard which will subsequently be used.

5. Data storage and backup during the FWO project
Where will the data be stored?

During the research activities, data is stored on the central computer cluster ("Dirac") of the Quantum Chemistry and Physical Chemistry division at KU Leuven. A daily back-up of this data is stored at the central ICT services.

Once per week, the data stored on the local computer of the PI is back-upped onto an external hard drive.

How is backup of the data provided?

Two ways are foreseen to back-up the generated data:

- Automatic daily back-ups to the central ICT systems of the data stored on Dirac .
- Weekly back-ups of local data to an external hard drive.

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

Dirac provides a storage capacity of 400 GB, while the external hard drive provides a capacity of 1 TB. If during the course of the project this turns out to be insufficient, additional hard drives can be purchased.

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

Dirac and the RDR repository can be used for free. Possible expected costs include the purchase of additional hard drives, if necessary. These costs can be covered by the allocated fellowship bench fee.

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

Access to Dirac is restricted to people having an account. The permissions for the data are regulated via unix file permissions, by default only the PI can access the files stored in her personal Dirac account.

The external hard drive (and any additional ones) will be protected with a password and anti-virus software.

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, ...).

All data generated throughout the research project will be retained.

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

The data will be stored on Dirac, with daily automatic back-up to the central ICT servers.

Additionally, the data will be archived in the RDR repository.

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

The Dirac cluster and RDR repository can be used for free. If there are any unforeseen costs, these can be covered using the allocated fellowship bench fee. Alternatively, the supervisors or the research groups can provide for the costs if necessary.

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)?

- Yes. Specify:

The NDA with Q-Chem prevents sharing of source code disclosed by them. Any data resulting from the use of the code can be shared.

Which data will be made available after the end of the project?

All data generated throughout the project will be made available.

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

- In an Open Access repository
- Upon request by mail

The data will be archived in the open-access RDR repository. Contact information of the

supervisors will be provided in the datasets for those who are further interested. The supervisors will be responsible for the follow-up of the data reuse and requests.

When will the data be made available?

- Upon publication of the research results

For results that are not published, the data will be made available upon request by e-mail.

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

The data will be open to all external users under an open licenses generated by the RDR platform. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.

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

The RDR repository is free, provided that the data does not exceed 50 GB per year. It is expected that this limit will not be exceeded.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The PI is responsibe for collecting, processing, analysing and documenting the generated data and metadata.

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

The PI is responsibe for the regular data storage and backups.

Who will be responsible for ensuring data preservation and reuse ?

The promotors (Thomas Jagau and Frank De Proft) are responsible to ensure the data preservation and sharing.

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

The PI, under the supervision of the promotor, bears the overall responsibility for updating & implementing this DMP.