Ab initio spin-flip methods for theoretical photochemistry
DPIA

DPIA

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

- Not applicable

Ab initio spin-flip methods for theoretical photochemistry
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Ab initio spin-flip methods for theoretical photochemistry
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Input files	Input files to initiate electronic structure calculations with various softwares (Q-Chem, Turbomole, CFOUR...)	Generate new data	Digital	Simulation data	.in	<100 MB	
Output files	Output files resulting from electronic structure calculations with various softwares (Q-Chem, Turbomole, CFOUR...)	Generate new data	Digital	Simulation data	.out	<100 GB	
Q-Chem code	Q-Chem source code implemented during the PhD	Generate new data	Digital	Software	.C or .h	<100 GB	

Scripts for processing data	Scripts written in bash or python to extract/ visualize/ summarise information	Generate new data	Digital	Software	.sh or .py	<100 MB	
Result collections	Collections of results obtained with electronic structure softwares, collected with bash/python	Generate new data	Digital	Compiled/aggregated data	.txt or .csv	<100 MB	
Figures / plots	Figures and plots representing the results	Generate new data	Digital	Compiled/aggregated data	.png, .jpg, .pdf or .svg	<100 MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The source code will be part of the commercial software package Q-Chem.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes
- Q-Chem code: Development in Q-Chem requires access to the source code for which a non-disclosure agreement (NDA) is

signed. It is prohibited to share parts of the code that you do not have written yourself or other proprietary and confidential information disclosed by Q-Chem.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Appropriate README files will be included to explain the structure and content of the datasets.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

The metadata standard of KU Leuven's research data repository (RDR) will be used.

3. Data storage & back-up during the research project

Where will the data be stored?

Input, output and code files are stored on the local cluster ("Dirac") of the chemistry department of KU Leuven. Once the code is correctly implemented, it will also be stored at the Q-Chem servers.
Processed data will be stored on the PhD student's computer and an external hard drive.

How will the data be backed up?

Dirac is backed-up daily at the central ICTS services. The processed data are backed-up weekly to the external hard drive. The Q-Chem code is under SVN (subversion) version control.

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

On Dirac, 400 GB of storage is available per user. The total amount of storage is not expected to exceed 300 GB, see table under 1. Research Data Summary. The external hard drive has 1 TB of storage which is more than enough to accommodate all data.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Dirac is secured by a password and needs a secure connection through the KU Leuven network or a VPN. All output from software is by default unreadable for all other users. The computer and external hard drive are password secured and the office is always locked when unsupervised.

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

There are no additional cost for the PhD student for using the storage and back-up of Dirac of Q-Chem. If additional storage is required, an extra external hard drive can be bought with the bench fee.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data will be stored for at least 10 years after the project.

Where will these data be archived (stored and curated for the long-term)?

All data will be archived at KU Leuven's research data repository (RDR).

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

The RDR is free to use for the PhD student.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

The RDR is open access.

If access is restricted, please specify who will be able to access the data and under what conditions.

NA

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Intellectual Property Rights

The parts of the Q-Chem code that are not written by the PhD student cannot be shared.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The data will be made available in the RDR.

When will the data be made available?

The data will be made available upon publication of the PhD thesis. The data supporting scientific papers that result from the PhD, are made available upon publication of these papers.

Which data usage licenses are you going to provide? If none, please explain why.

The data can be shared, and modified as long as the original authors are given credit and changes are indicated (CC-BY).

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

Not yet available.

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

There are no expected cost for data sharing since the use of the RDR is free for the PhD student.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

The PhD student.

Who will manage data storage and backup during the research project?

The PhD student.

Who will manage data preservation and sharing?

The PhD student during the research project and the supervisor afterwards.

Who will update and implement this DMP?

The PhD student.