FWO DMP Template - Flemish Standard Data Management Plan

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	**Nguyen Ha Quyen, 0000-0002-1459-137X**
Contributor name(s) (+ ORCID) & roles	**Thomas Jagau, 0000-0001-5919-424X, promotor at KU Leuven** **Lars Goerigk, 0000-0003-3155-675X, promotor at University of Melbourne**
Project number[1] & title	Density functional methods for unusual electronic structures
Funder(s) GrantID[2]	GPUM/22/011
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel X Other: The University of Melbourne Provide ROR[3] identifier when possible:

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	The emission of an Auger electron is the predominant relaxation mechanism of core-vacant states in molecules composed of light nuclei and thus of great importance to all settings that involve ionizing radiation such as X rays. Auger decay is a non-radiative process in which one valence electron fills the core vacancy while a second valence electron is emitted into the ionization continuum. Because of this coupling to the continuum, core-vacant states represent electronic resonances that can be tackled with standard quantum-chemical methods only if they are approximated as bound states, meaning that Auger decay is neglected. In recent years, several theoretical methods based on non-Hermitian quantum mechanics have been suggested for modelling Auger decay but these approaches remain limited to small systems. The aim of the present project is to combine non-Hermitian quantum mechanics with time-dependent density functional theory in order to treat Auger decay in complex systems comprising up to 100 atoms. The project thus combines theoretical developments with the implementation into a computer code and applications to chemistry. The first 15 months of the project will be based at KU Leuven followed by 12 months in Melbourne. The project will then conclude at KU Leuven.

2. 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[4].

				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
Inputs	Input files for Q-Chem quantum chemistry software	☒ Generate new data	☒ Digital	☒ Other manually created	☒ .txt	☒ < 100 MB	
Outputs	Output files for Q-Chem quantum chemistry software	☒ Generate new data	☒ Digital	☒ Simulation data	☒ .txt	☒ < 100 GB	
Summaries	Tables and text summarizing and analysing the raw data	☒ Generate new data	☒ Digital	☒ Compiled/ aggregated data	☒ other: .xlsx ☒ .pdf	☒ < 100 MB	
Visualizations	Images and graphs visualizing results and conclusions drawn from them	☒ Generate new data	☒ Digital	☒ Simulation data	☒ other: .jpg, .svg, .mp4	☒ < 100 GB	
Scripts	Scripts for post-progressing and analysing output files	☒ Generate new data	☒ Digital	☒ Software	☒ .txt	☒ < 100 MB	

[4] Add rows for each dataset you want to describe.

GUIDANCE: *DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, …). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.* *EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[5] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.* *EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR,. SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG,. GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.* *DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.* *PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).*	
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.	Mostly generation of new data Reuse of data from the following sources: N. K. Jayadev, A. Ferino-Pérez, F. Matz, A. I. Krylov, T.-C. Jagau: The Auger spectrum of benzene, J. Chem. Phys. 158, 064109/1-17 (2023). DOI: 10.1063/5.0138674
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, please describe these issues further and refer to specific datasets or data types when appropriate.	☐ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☒ No If yes, please describe:

[5] These data are generated by combining multiple existing datasets.

Will you process personal data[6]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☐ Yes ☒ No If yes: - Short description of the kind of personal data that will be used: - Privacy Registry Reference:
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 ☐ No If yes, please comment: 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 exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☒ Yes ☐ No If yes, please explain: 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.
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 to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

[6] See Glossary Flemish Standard Data Management Plan

3. 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).	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 to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☒ Yes ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: The data will be archived in KU Leuven's RDR repository. This repository provides an appropriate metadata standard which will be used. If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	While the project is carried out in Leuven, the generated data is stored on the computer cluster "Dirac" of the Quantum Chemistry and Physical Chemistry (QCPC) division at the Department of Chemistry of KU Leuven. A daily back-up of this data is stored at the central ICT services. While the project is carried out in Melbourne, the generated data is stored on the computer cluster in the Department of Chemistry of the University of Melbourne.
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.*[7] *REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.*	Two procedures are foreseen: – Automatic daily backups of the data stored on “Dirac” to the central ICT systems (using rsync protocol). – The data stored locally on the computer of the PI is backed up onto One 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 ☐ No If yes, please specify concisely: 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. If no, please specify:

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* [7]	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.
What are the expected costs for data storage and backup during the research 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 bench fee.

5. 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 generated throughout the research project will be retained for 10 years in accordance with KU Leuven policies.
Where will these data be archived (stored and curated for the long-term)?	All data will be stored on the computer cluster "Dirac" with daily automatic backups to the central ICT servers. Additionally, the data will be archived in the RDR repository of KU Leuven.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The computer cluster "Dirac" and the RDR repository can be used for free. If there are any unforeseen costs, these can be covered by the bench fee of the PI and the division Quantum Chemistry and Physical Chemistry.

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☒ Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify: All data generated throughout the project will be made available in the open-access RDR repository.
If access is restricted, please specify who will be able to access the data and under what conditions.	Contact information of the KU Leuven promotor will be provided in the datasets for those who are further interested. The KU Leuven promotor will be responsible for following up requests for data access and reuse.
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 per dataset or data type where appropriate.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☒ Yes, other ☐ No If yes, please 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.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	All data will be archived in the open-access RDR repository.

When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	Upon publication of the results. Unpublished results will be made available upon request by email.
Which data usage licenses are you going to provide? If none, please explain why. *A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [8]	The data will be open to all external users under the license generated by the RDR platform, e.g. CC-BY 4.0 Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *Indicate whether you intend to add a persistent and unique identifier in order to identify and retrieve the data.*	☐ Yes ☒ No If yes:
What are the expected costs for data sharing? How will these costs be covered?	No costs are expected. The RDR repository is free, provided that the data does not exceed 50 GB per year, which is not expected to happen.

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The PI is responsible for collecting, processing, analyzing and documenting the generated data and metadata.
Who will manage data storage and backup during the research project?	The PI is responsible for regular data storage and backups.
Who will manage data preservation and sharing?	The promotors (Prof. Thomas Jagau, KU Leuven and Prof. Lars Goerigk, University of Melbourne) are responsible for data preservation and sharing.
Who will update and implement this DMP?	The PI, under the supervision of the promotors, is responsible for implementing this DMP and updating it when needed.