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	**Thomas Jagau, 0000-0001-5919-424X**
Contributor name(s) (+ ORCID) & roles	**Wojciech Skomorowski, 0000-0002-0364-435X, collaborator at University of Warsaw**
Project number[1] & title	Autoionizing Rydberg states using non-Hermitian quantum chemistry
Funder(s) GrantID[2]	CELSA/22/014
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:
Please provide a short project description	The project seeks to establish novel computational protocols for the study of molecular Rydberg states, with a focus on autoionization. To this end, we will apply various methods of non-Hermitian quantum chemistry, including complex basis functions, complex absorbing potentials, and Feshbach-Fano projection. All of them will be combined with equation-of-motion coupled-cluster theory – an electronic structure method capable of describing various types of excited states with high and controllable accuracy. The generated data will mostly consist of software outputs, possibly in addition some post-processing scripts.

---

[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/

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
Outputs	Output files of Q-Chem quantum chemistry software	☒ Generate new data	☒ Digital	☒ Simulation data	☒ .txt	☒ < 100 GB	
Inputs	Input files for Q-Chem quantum chemistry software	☒ Generate new data	☒ Digital	☒ Other Manually created	☒ .txt	☒ < 100 MB	
Summaries	Tables and text summarizing and analyzing 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-processing and analyzing output files	☒ Generate new data	☒ Digital	☒ Software	☒ .txt	☒ < 100 MB	

---

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

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.	No reuse of data planned, only generation of new data.
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:
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:
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:

3. Documentation and Metadata

---

[6] See Glossary Flemish Standard Data Management Plan

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).	All Q-Chem output files of calculations include by default the input file and the Q-Chem version number used to generate them. This is already worth a great deal to ensure usability and reproducibility of the data. In addition, in all data folders, separate text files will be saved with information about the data and their context. This includes: – Overview of files in the respective folder. – Overview of the performed computations and their relation to each other. – Purpose of the performed computations and their relation to other datasets and publications (if applicable)
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.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	While the project is carried out, 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. The data stored locally on the computer of the PI and all other involved researchers are backed up weekly onto an external hard drive.

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). – Manual weekly backups of the data stored locally on the computer of the PI and the other involved researchers 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 ☐ No If yes, please specify concisely: The cluster "Dirac" currently provides a default storage capacity of 400 GB per user. This is expected to be sufficient for the present project but can be increased if necessary. The external hard drives provide a capacity of 1 TB per researcher. If during the course of the project this turns out to be insufficient (unlikely), additional hard drives can be purchased.
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 the computer cluster "Dirac" is restricted to people who have an account. This comprises all current members of the division Quantum Chemistry and Physical Chemistry. Read, write, execute access to the data are regulated via unix file permissions, by default every user can only access the files stored in their personal folder on "Dirac". All external hard drives are protected by password and antivirus software.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The computer cluster "Dirac" can be used for free. Possible costs include the purchase of new external hard drives. This could be covered by the CELSA grant if necessary.

---

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

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 funds 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 PI will be provided in the datasets for those who are further interested. The PI 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:
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.

7. Responsibilities
---	---
Who will manage data documentation and metadata during the research project?	The PI, Thomas Jagau, is responsible for collecting, processing, analysing and documenting the generated data and metadata.

---

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

Who will manage data storage and backup during the research project?	The PI, Thomas Jagau, is responsible for regular data storage and backups.
Who will manage data preservation and sharing?	The PI , Thomas Jagau, is responsible for data preservation and sharing.
Who will update and implement this DMP?	The PI , Thomas Jagau, is responsible for implementing this DMP and updating it when needed.