

## 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	<b>Thomas Jagau, 0000-0001-5919-424X</b>
Contributor name(s) (+ ORCID) & roles	<b>Koushik Chatterjee, postdoc</b> <b>Cansu Utku, PhD student</b>
Project number <sup>1</sup> & title	Computational investigation of chemical reactions and spectroscopy of unbound electrons
Funder(s) GrantID <sup>2</sup>	C14/22/083
Affiliation(s)	X KU Leuven Provide ROR <sup>3</sup> identifier when possible:
Please provide a short project description	The project seeks to advance quantum chemistry of electronic resonances. Building on recent methodological progress in non-Hermitian electronic-structure methods, we will explore chemistry and spectroscopy of unbound electrons driven by different types of resonance states. Specifically, we will computationally investigate 1) dissociative electron attachment to unsaturated organohalides, to pyrrole, and to nucleobases, 2) photodetachment spectra of molecular alkyl halide clusters and polycyclic aromatic hydrocarbons with a focus on the signatures of temporary anions in these spectra, 3) Auger electron spectra and near-edge X-ray absorption fine structure spectra of small and medium-sized organic molecules with biological relevance such as formamide, pyridine, furan, and uracil, 4) intermolecular Coulombic decay of such species in aqueous solution. The generated data will mostly consist of software outputs, post-processing scripts, and scientific manuscripts for publication.

<sup>1</sup> "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.

<sup>2</sup> 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.

<sup>3</sup> 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<sup>4</sup>.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Outputs	Output files of quantum chemistry software, mostly Q-Chem but also Turbomole and CFOUR	<input checked="" type="checkbox"/> Generate new data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Simulation data	<input checked="" type="checkbox"/> .txt	<input checked="" type="checkbox"/> < 100 GB	
Inputs	Corresponding input files for quantum chemistry software	<input checked="" type="checkbox"/> Generate new data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Other Manually created	<input checked="" type="checkbox"/> .txt	<input checked="" type="checkbox"/> < 100 MB	
Summaries	Tables/text/sketches summarizing and analyzing the raw data	<input checked="" type="checkbox"/> Generate new data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Compiled/aggregated data	<input checked="" type="checkbox"/> .pdf <input checked="" type="checkbox"/> .txt <input checked="" type="checkbox"/> other: xlsx	<input checked="" type="checkbox"/> < 100 MB	
Visualizations	Images and graphs visualizing results and the conclusions drawn from them	<input checked="" type="checkbox"/> Generate new data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Simulation data	<input checked="" type="checkbox"/> other: jpg, svg, mp4	<input checked="" type="checkbox"/> < 100 MB	
Scripts	Scripts for post-processing and analyzing output files	<input checked="" type="checkbox"/> Generate new data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Software	<input checked="" type="checkbox"/> .txt	<input checked="" type="checkbox"/> < 100 MB	

<sup>4</sup> Add rows for each dataset you want to describe.

<p><b>GUIDANCE:</b></p> <p>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.</p> <p>EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA<sup>5</sup> (E.G. TEXT &amp; DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.</p> <p>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 &amp; COMPUTATIONAL SCRIPT.</p> <p>DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.</p> <p>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).</p>	
<p>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.</p>	<p>No reuse of data planned, only generation of new data.</p>
<p>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.</p>	<p> <input type="checkbox"/> Yes, human subject data  <input type="checkbox"/> Yes, animal data  <input type="checkbox"/> Yes, dual use  <input checked="" type="checkbox"/> No          If yes, please describe:       </p>

---

<sup>5</sup> These data are generated by combining multiple existing datasets.

<p>Will you process personal data<sup>6</sup>? 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.</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No          If yes:</p> <ul style="list-style-type: none"> <li>- Short description of the kind of personal data that will be used:</li> <li>- Privacy Registry Reference:</li> </ul>
<p>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.</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No          If yes, please comment:</p>
<p>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.</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No          If yes, please explain:</p>
<p>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.</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No          If yes, please explain:</p>

---

<sup>6</sup> 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 <b>data understandable and usable</b> , 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).	<p>All outputs files produced by Q-Chem, Turbomole, and CFOUR include by default the input files and the version number of the software used to generate them. This ensures usability and reproducibility of the data.</p> <p>In addition, in all data folders, we will save separate text files with information about the data and their context. This includes: 1) overview of the files in the respective folder, 2) overview of the performed computations and their relation to each other, 3) purpose of the performed computations and their relation to other datasets and publications</p>
<p>Will a metadata standard be used to make it easier to <b>find and reuse the data</b>?</p> <p>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.</p> <p><i>REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.</i></p>	<p><input checked="" type="checkbox"/> Yes  <input type="checkbox"/> No</p> <p>If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used:</p> <p>The data will be archived in KU Leuven's RDR repository. This repository provides an appropriate metadata standard which will be used.</p>

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

Where will the data be stored?	<p>During the execution of the project, the generated data will be 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 the other involved researchers are backed up weekly onto an external hard drive.</p>
--------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<p><b>How will the data be backed up?</b></p> <p><i>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.<sup>7</sup></i></p> <p><i>REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.</i></p>	<p>Two procedures are foreseen:</p> <ul style="list-style-type: none"> <li>– Automatic daily backups of the data stored on “Dirac” to the central ICT systems (using rsync protocol).</li> <li>– Manual weekly backups of the data stored locally on the computer of the PI and the other involved researchers to an external hard drive.</li> </ul>
<p><b>Is there currently sufficient storage &amp; 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.</b></p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>If yes, please specify concisely: The cluster “Dirac” currently provides a 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.</p>
<p><b>How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</b></p> <p><i>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. <sup>7</sup></i></p>	<p>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 (QCPC). 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.</p>
<p><b>What are the expected costs for data storage and backup during the research project? How will these costs be covered?</b></p>	<p>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 C1 grant if necessary.</p>

<sup>7</sup> 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 (QCPC).

### 6. Data Sharing and Reuse



<p>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.</p> <p><i>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 &amp; RESTRICTED ACCESS. FOR MORE INFORMATION:</i> <a href="https://wiki.surfnet.nl/display/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS">HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS</a></p>	<p><input checked="" type="checkbox"/> Yes, in an Open Access repository  <input type="checkbox"/> Yes, in a restricted access repository (after approval, institutional access only, ...)  <input type="checkbox"/> No (closed access)  <input type="checkbox"/> Other, please specify:</p> <p>All data generated throughout the project will be made available in the open-access RDR repository.</p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>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.</p>
<p>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.</p>	<p><input type="checkbox"/> Yes, privacy aspects  <input type="checkbox"/> Yes, intellectual property rights  <input type="checkbox"/> Yes, ethical aspects  <input type="checkbox"/> Yes, aspects of dual use  <input type="checkbox"/> Yes, other  <input checked="" type="checkbox"/> No</p> <p>If yes, please specify:</p>
<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<p>All data will be archived in the open-access RDR repository.</p>
<p>When will the data be made available?</p> <p><i>THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.</i></p>	<p>Upon publication of the results. Unpublished results will be made available upon request by email.</p>

<p>Which data usage licenses are you going to provide? If none, please explain why.</p> <p><i>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.</i></p> <p><i>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." <sup>8</sup></i></p>	<p>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.</p>
<p>Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.</p> <p><i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i></p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No  If yes:</p>
<p>What are the expected costs for data sharing? How will these costs be covered?</p>	<p>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.</p>

## 7. Responsibilities

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

<sup>8</sup> 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.