Quantum Chemical Design of Molecular Magnets

A Data Management Plan created using DMPonline.be

Creators: Maristella Alessio, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: European Commission (Horizon)

Template: Horizon Europe DMP +

Principal Investigator: Maristella Alessio

Data Manager: Maristella Alessio

Project Administrator: Maristella Alessio

Grant number / **URL:** 101062717

ID: 198638

Start date: 01-11-2022

End date: 31-10-2024

Project abstract:

A reliable *ab initio* description of molecular magnets is key to developing a new era of quantum devices that will be more efficient and easier to tune by structural modification of their building units. However, quantum mechanical treatment of such systems is challenging due to their multi-configurational wavefunctions, requiring a well balanced description of their constituent electronic configurations. Furthermore, these systems are often large magnetic molecules or atoms deposited on supports whose models include hundreds of atoms, hampering the application of accurate *ab initio* methods; yet small energy gaps (from tens to hundreds of wavenumbers) call for quantitative accuracy. The aim of this project is to design new molecular magnets, practical for real-world applications. To this end, we will employ a new and affordable computational strategy that combines accurate equation-of-motion coupled-cluster (EOM-CC) theory on the magnetic center with more approximate density functional theory (DFT) on the remainder, avoiding costly EOM-CC calculations on the full system. We will employ interdisciplinary approaches, EOM-CC-in-DFT for open-shell species and tools computing magnetic properties from *ab initio* calculations, to determine how microscopic interactions (spin-orbit and Zeeman) contribute to macroscopic magnetic properties and how these are optimized in two model systems: (i) a cobalt(II) single-molecule magnet and (ii) single cobalt atoms on the MgO(001) and Cu(111) surfaces.

Last modified: 17-04-2023

Quantum Chemical Design of Molecular Magnets
Grant proposal

1. Data summary

1.1 Types of data/other research outputs

Dataset Name	Description	New or Reused	Primary or Secondary	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)
Inputs_Q-Chem	Input files of the Q-Chem quantum chemistry software	Generate new data	Primary	Digital	Manually created	.txt	< 100 MB
Outputs_Q-Chem	Output files of the Q-Chem quantum chemistry software	Generate new data	Primary	Digital	Simulation data	.txt	< 10 GB
Inputs_ezMagnet	Input files of the ezMagnet sofware	Generate new data	Primary	Digital	Manually created	.txt	< 100 MB
Output_ezMagnet	Output files of the ezMagnet software	Generate new data	Primary	Digital	Simulation data	.txt	< 100 MB
Input_VASP	Input files of the VASP quantum chemistry software	Generate new data	Primary	Digital	Manually created	.txt	< 100 MB
Output_VASP	Output files of the VASP quantum chemistry software	Generate new data	Primary	Digital	Simulation data	.txt	< 10 GB
Summaries	Tables and text summarising, processing, and analyzing the raw data	Generate new data	Primary	Digital	Compiled/aggregated data	.ods and .pdf	< 100 MB
Visualisations	Images and graphs visualising results and conclusions drawn from them	Generate new data	Primary	Digital	Simulation data	.jpg, .png, and .svg	< 100 MB
Scripts	Scripts for post-processing and analysing output files	Generate new data	Primary	Digital	Software	Python and ReadMe.txt	< 100 MB

2. FAIR principles

2.1 Findability of data/research outputs

As soon as conclusive data analysis will be achieved, data will be deposited in the Research Data Repository (RDR) of KU Leuven, https://rdr.kuleuven.be/. To make data findable, RDR assigns a unique digital object identifier (DOI) to the uploaded data, and makes use of metadata files.

2.2 Accessibility of data/research outputs

Data deposited in the RDR repository will be publicly available. Metadata files will explain the procedures for accessing the data under the open access license (Creative Common Attribution International Public License (CC-BY)). Therefore, data will be available to all external users for any purpose, provided that they give appropriate credit to the creators. Additionally, manuscripts for publications will be placed in a preprint server before submission (ChemRxiv), the peer-reviewed publication will then be available through open access CC-BY license, and will also be deposited in KU Leuven's repository system ("Liras").

2.3 Interoperability of data/research outputs

Data and metadata will be deposited in a format that can be accessible for everyone (open-file format). For textual documents, mostly .txt, .odt, and .pdf formats will be used. Format for spreadsheets will be mainly ODS. For figures, we will use .jpg, .png, and .svg formats. Scripts will be provided as Python files. Within RDR repository, metadata standards and controlled vocabularies will be automatically applied upon deposition of the research data. Data files will be clearly organised into a logical hierarchy of folders, which will reflect the workflow in the project. Systematic file nomenclature including dates (YYYYMMDD), project name, and designer name will also be applied.

2.4 Reusability of data/research outputs

For each dataset deposited in the RDR repository, documentation and metadata files will be created, describing provenance of the research data and outputs, and explaining their accessibility under a CC-BY license. Such research data description includes information on how, why, and by whom is the data created and/or processed, specifying working protocols and adopted software.

3. Resources and responsibilities

3.1 Curation and storage/preservation costs

Data will be archived in a trusted repository, the RDR repository of KU Leuven. In addition to RDR repository, the Jagau lab is equipped with in-house storage servers: i) external disk for weekly backups of personal computers (up to 1 TB for user) and ii) the "Dirac" computer cluster for daily backups of the produced Q-Chem output files (up to 400 GB for user). Similarly, up to 500 GB for user will be available for backups of VASP output files in the Vienna Super Computer (VSC) center in Austria (secondment period). In this project, there will no be cost to make data and research output FAIR, as adopted repositories (RDR, ChemRxiv, and Liras) and in-house storage units, e.g., Dirac and VSC computer clusters, and external hard drives, are free of charges for small/medium datasets (less that 50 GB) as the one expected in this project.

3.2 Person/team responsible for data management and quality assurance

The research fellow is responsible for the data management during the duration of the project. In addition, the supervisors of the project, Prof. Jagau at KU Leuven and Prof. Grüneis at TU Wien, will ensure the long-term storage and preservation of the data and outputs of this project in their in-house storage servers. Contact information and ORCID of the research fellow and of the supervisor will be provided in the datasets for those who are further interested.

Quantum Chemical Design of Molecular Magnets
DPIA

DPIA

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

- Not applicable

Quantum Chemical Design of Molecular Magnets
GDPR record

GDPR record

Have you registered personal data processing activities for this project?

- Not applicable

Quantum Chemical Design of Molecular Magnets
Full DMP

Version information

Action number

101062717

Action acronym

ezEmbedMagnet

Action title

Quantum Chemical Design of Molecular Magnets

DMP version number

v1.0

Date

16.04.2023

1. Data summary

1.1 Will you re-use any existing data and what will you re-use it for?

No re-use of data is planned, only generation of new data.

1.2 What types and formats of data and other research outputs will the project generate or re-use?

Dataset Name	Description	New or Reused	Primary or Secondary	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)
Inputs_Q-Chem	Input files of the Q-Chem quantum chemistry software	Generate new data	Primary	Digital	Manually created	.txt	< 100 MB
Outputs_Q-Chem	Output files of the Q-Chem quantum chemistry software	Generate new data	Primary	Digital	Simulation data	.txt	< 10 GB
Inputs_ezMagnet	Input files of the ezMagnet sofware	Generate new data	Primary	Digital	Manually created	.txt	< 100 MB
Output_ezMagnet	Output files of the ezMagnet software	Generate new data	Primary	Digital	Simulation data	.txt	< 100 MB
Input_VASP	Input files of the VASP quantum chemistry software	Generate new data	Primary	Digital	Manually created	.txt	< 100 MB
Output_VASP	Output files of the VASP quantum chemistry software	Generate new data	Primary	Digital	Simulation data	.txt	< 10 GB
Summaries	Tables and text summarising, processing, and analyzing the raw data	Generate new data	Primary	Digital	Compiled/aggregated data	.ods and .pdf	< 100 MB
Visualisations	Images and graphs visualising results and conclusions drawn from them	Generate new data	Primary	Digital	Simulation data	.jpg, .png, and .svg	< 100 MB
Scripts	Scripts for post-processing and analysing output files	Generate new data	Primary	Digital	Software	Python and ReadMe.txt	< 100 MB

1.3 What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The aim of the project is to design more efficient molecular magnets with potential application in quantum information science. To do so, it is important to use accurate yet affordable quantum chemical methods capable of describing their electronic structure, and to develop useful computational protocols for predicting their magnetic properties. Data and outputs of this project will determine the performances of quantum embedding theories in describing magnetic behavior of a set of cobalt-based molecular magnets.

1.4 What is the expected size of the data that you intend to generate or re-use?

The overall size of the generated data is expected to be less than 30 GB. The two datasets including Q-Chem and VASP output files are the ones expected to have larger volume (maximum 10 GB). While generated input, generate scripts, summaries, and visualisation datasets are expected to have a total volume less than 1 GB.

1.5 What is the origin/provenance of the data, either generated or re-used?

The output datasets are originated by the research fellow using the Q-Chem, VASP, and ezMagnet software. The Q-Chem and VASP software are fully available in the groups of Port. Jagau and Prof. Grüneis (secondment), respectively. The ezMagnet software is an open-access software generated by the research fellow in previous projects and it is available for download to all users at http://iopenshell.usc.edu/downloads/ezmagnet/ (DOI:doi.org/10.1021/acs.jctc.1c00430). Input files will be generated manually using text editors by the research fellow as well as additional post-

processing scripts, which will be generated using Python. Summaries and visualization datasets will be generated using open-access software, such as LibreOffice, Gnuplot, Inkscape, and Jmol.

1.6 To whom might your data be useful ('data utility'), outside your project?

The data will be useful for those interested in the design of improved molecular magnets for quantum information science and quantum computing. This involves academic scientists as well as leaders in quantum computing (e.g., Google AI Quantum or IBM Quantum). In addition, these research outputs will be used by other academic and non-academic scientists for the development of other embedding methods, e.g., periodic embedding theories for strongly correlated systems, and to describe other complex chemical systems, e.g, plasmonic catalysts or paramagnetic defects in solids.

2.1 FAIR data: Making data findable, including provisions for metadata

2.1.1 Will data and other research outputs be identified by a persistent identifier?

- Yes: describe below

Data will be archived in KU Leuven's RDR repository which assigns a unique persistent identifier (DOI) to the deposited data. When needed, these identifiers will be provided in the related publication. All authors involved in the project will also be linked to their respective ORCID IDs.

2.1.2 Will rich metadata be provided to allow discovery?
What metadata will be created?
What disciplinary or general standards will be followed?
In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Information helping to document the provenance of the data generated during this research project will be initially recorded in ReadMe.txt files. Discoverability of output data will then be possible by depositing output data into the RDR repository at KU Leuven. Within the RDR repository, general metadata standards are automatically applied upon upload of each dataset, which facilitates identification of the deposited data, allowing long-term findability, accessibility and usability of the data. The RDR metadata model includes required, recommended and optional fields to make the dataset more findable and reusable. Required fields are: title, author name and contact information, identifiers, description of the dataset, keywords, format of the resource, and access right information.

2.1.3 Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

- Yes: describe below

A set of three key terms is provided that describes important aspects of the dataset. These keywords make the dataset more findable for other researches working in the field of quantum chemistry and molecular magnetism. To increase discoverability of the data, controlled vocabularies will be used.

2.1.4 Will metadata be offered in such a way that it can be harvested and indexed?

- Yes: describe below

The RDR repository is a trusted data repository that ensure metadata fields and digital object identifiers (DOIs) to uploaded data. In such way, research data will be archived and indexed in a controlled manner, and thus, easy to discover.

2.2 FAIR data: Making data accessible

2.2.1 Will the data and other research outputs be deposited in a trusted repository?

- Yes: describe below

Data will be archived in the KU Leuven's RDR repository.

2.2.2 Have you explored appropriate arrangements with the identified repository where your data and other research outputs will be deposited?

- Yes

2.2.3 Does the repository ensure that the data and other research outputs are assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes, RDR repository uses DOIs as identifiers of datasets.

2.2.4 Will all data and other research outputs be made openly available?

- Yes

Data deposited in the RDR repository will be available as open data (public access under the CC-BY license). Additionally, manuscripts for publications will be placed in a preprint server before submission (ChemRxiv), the peer-reviewed publication will then be available through open access CC-BY license, and will also be deposited in KU Leuven's repository system ("Liras").

2.2.5 Is an embargo applied to give time to publish or seek protection of the intellectual property (e.g. patents)?

- No

2.2.6 If an embargo is applied (see question 2.2.5), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

There will be no embargo applied.

2.2.7 Will the data and other research outputs be accessible through a free and standardized access protocol?

- Yes: describe below

Accessibility is ensured by the RDR repository that will be used in this project.

2.2.8 If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

There will be no restrictions applied.

2.2.9 How will the identity of the person accessing the data be ascertained?

There will be no need to know who will access the data, thus, we will not integrate any access request procedure.

2.2.10 Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

- No

2.2.11 Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

- Yes

2.2.12 Will metadata contain information to enable the user to access the data?

- Yes

2.2.13 How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

KU Leuven's RDR repository ensures availability of the deposited dataset for a period of 10 years after the end of the project.

2.2.14 Will documentation or reference about any software needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Documentation files (ReadMe.txt) will include a list of the software (Q-Chem, VASP and ezMagnet) used in this project with references. In addition, all Q-Chem, VASP, and ezMagnet output files of calculations include by default the version number of the software used to generate them, which facilitate reproducibility. ezMagnet is an open-source software developed by the research fellow in previous projects and it is available for download at http://iopenshell.usc.edu/downloads/ezmagnet/. ezMagnet will not be included in the RDR archive since this does not represent an output of this project and it is already made available elsewhere, but link and DOIs will be provided in the documentation. On the contrary, additional scripts generated in this project for post-processing data will be uploaded in the RDR repository and will be open access.

2.3 FAIR data: Making data interoperable

2.3.1
What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines?
Will you follow community-endorsed interoperability best practices? Which ones?

Data and metadata will be deposited in a format that can be understood across multiple disciplines to reach a broad scientific community. Data will be available in open-file formats: .txt, .odt, and .pdf for textual documents; .ods for spreadsheets; .jpg, .png, and .svg for figures; and Python files for scripts. A metadata standard will be automatically applied upon deposition of the data into the RDR archive. Controlled vocabularies are already integrated in the RDR repository, allowing meta(data) to be combined and exchanged. Moreover, a logical hierarchy of folders will be used together with a systematic file nomenclature including dates (YYYYMMDD), project name, and designer name.

2.3.2 In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies:
Will you provide mappings to more commonly used ontologies?
Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

If and when specific vocabularies will be generated during the project, these will be openly available.

2.3.3 Will your data and other research outputs include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

- Yes

If and when applicable.

2.4 FAIR data: Increase data re-use

2.4.1 How will you provide documentation needed to validate data analysis and facilitate data re-use?

In all data folders of each dataset, separate files will be created (ReadMe.txt files) reporting information about the data, their context and provenance, and their analysis. This includes: i) overview of files in the respective folder, ii) overview of the performed computations/analysis and their relation to each other, iii) purpose of the performed computations and their relation to other datasets and publications (if applicable). In addition, all Q-Chem, VASP, and ezMagnet output files deposited in the trusted repository include by default the input file and the version number of the software used to generate them. This ensures usability and reproducibility of the data.

2.4.2
Will your data and other research outputs be made freely available in the public domain to permit the widest re-use possible?
Will your data and other research outputs be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Yes, data will be made freely available on the RDR repository under standard re-use licenses (CC-BY).

2.4.3 Will the data and other research output produced in the project be useable by third parties, in particular after the end of the project?

- Yes

2.4.4 Will the provenance of the data and other research outputs be thoroughly documented using the appropriate standards?

- Yes

2.4.5 Describe all relevant data quality assurance processes.

The quality of data will be ensured during raw data generation, data processing, and analysis. To do so, well-established computational methods and protocols will be adopted which are in agreement with common requirements in our scientific discipline. Such procedures will be well documented in the Method and Computational Details section of each open-access publication. Regular meeting with project supervisors and dissemination/exploitation actions at conferences and workshop will furthermore ensure quality checks of both data collection and data analysis processes. Access to data will be free but modification will not be possible by others to ensure data quality and security (i.e., primary access protected by password).

3. Other research outputs

3.1 Do you have any additional information, that was not addressed in the previous sections, which you wish to provide regarding other research outputs that are generated or re-used throughout the project?

Not applicable.

4. Allocation of resources

4.1 What will the costs be for making data and other research outputs FAIR in your project?

Data will be archived in a trusted repository, i.e., the RDR repository of KU Leuven. Additional storage resources will be made available to the research fellow for backups of the produced raw and processed data during the project. The Jagau lab at KU Leuven is equipped with in-house storage servers: i) external disk for weekly backups of person computers (up to 1 TB for user) and ii) the "Dirac" computer cluster for daily backups of the generated Q-Chem output files (up to 400 GB for user). Furthermore, for backups of the VASP output files generated during secondment (TU Wien), the Grüneis group owns a fraction of 30 computer nodes at the Vienna Scientific Cluster (VSC) center, offering the possibility to store up to 500 GB for user, which will be shared with the research fellow. There will no be cost to make data and research output FAIR, as adopted repositories (RDR, ChemRxiv, and Liras) and in-house storage servers, e.g., "Dirac" computer cluster, external hard drives, and storage units at the VSC center in Austria, are free of charges. Additionally, the Jagau and the Grüneis groups have full access to the Q-Chem and to the VASP software, which will be provided to the research fellow for the execution of the project. On the contrary, ezMagnet is an open-source software, developed by the research fellow in previous projects, which is available for download at http://iopenshell.usc.edu/downloads/ezmagnet/. Furthermore, to learn how to more efficiently manage research data, I will attend two training sessions of two hours each offered by KU Leuven (05.2023).

4.2 How will these be covered?

There are not expected costs for storing small/medium datasets (less than 50 GB).

4.3 Who will be responsible for data management in your project?

The research fellow is responsible for the data management during the duration of the project. In addition, the supervisors of the project, Prof. Jagau at KU Leuven and Prof. Grüneis, will ensure the long-term storage and preservation of the data and outputs of this project in their in-house computer clusters, Dirac and VSC, respectively.

4.4 How will long term preservation be ensured?

All data generated throughout the research project will be retained in the RDR repository for at least 10 years in accordance with the FAIR principle and KU Leuven policies.

5. Data security

5.1 What provisions are or will be in place for data security?

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 of the Department of Chemistry at KU Leuven. A daily back-up of this data is stored at the central ICT service (using rsync protocol). The cluster "Dirac" currently provides a default storage capacity of 400 GB per user, which is expected to be sufficient for the present project (estimated storage volume is less than 30 GB). The data stored locally on the computer of the research fellow, of the supervisor, and of all other involved researchers are manually backed up weekly onto an external hard drive. The latter provide capacity of 1 TB per researcher, and so, it is expected to be sufficient for the present project. Access to the computer cluster "Dirac" is restricted to the people who have an account. This comprises all current members of the QCPC division. Read, write, execute access to the data are regulated via unix file permission, by default every user can only access the files stored in their personal folder in "Dirac". All external hard drive are protected by password and antivirus software. Similarly, all generated data during secondment period will be stored in the VSC center in Austria providing up to 500 GB for user, which also limits its access account holders only.

5.2 Will the data be safely stored in trusted repositories for long term preservation and curation?

- Yes

6. Ethics

6.1 Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

- No

6.2 Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

- Not applicable

7. Other issues

7.1 Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

- No