Quantum Chemistry on Interstellar Grains

Summary

Project acronym

QUANTUMGRAIN

Project number

865657

Provide a dataset summary

The purpose of QUANTUMGRAIN is to obtain new data on the role of interstellar grains in the chemistry occurring in the Universe by means of theoretical methods based on computational chemistry. This new data includes structures of newly developed ice grain models, unexplored synthetic routes for the formation of the molecules, and the related structures energetics and dynamics. Thus, it merges quite different simulation activities that are producing data in heterogeneous formats . Also, the size of datasets varies greatly as a function of the associated source of simulation, from a few Mbytes up to hundreds of Gbytes.The expected size of the data should not exceed 1 Tb.

The data provided from computer simulations are mainly output files from different specific computational chemistry programs and packages. Some of them are commercial (Gaussian, CRYSTAL17 and VASP) while others are free of charge (CP2K, QuantumESPRESSO, ORCA and XTB). The generated files and extensions are well known within the computational chemistry community. They will be readable by any standard text editor (vi for Linux, standard text-editor for OSX and Windows operating systems).

The new data will complement the literature data because they are new and will serve the Astrochemical, Astronomical, and computational chemistry communities. Internally we reuse data provided in each WP. For instance, new grain structures will be reused to run reactivity simulations.

All data are meant to be available for long-term usage as they can be adopted by other researcher organizations as a starting point for different purposes than the one of QUANTUMGRAIN. The computational data will be useful to other modelers involved in the gas-grain fields, such as surface scientists, chemical physicists and astrochemical modelers.

FAIR data and resources

1. Making data findable

Outline the discoverability and identifiability of data produced/used:

For discoverability, those data not already stored in a specific repository, QUANTUMGRAIN will deposit and describe data in the QUANTUMGRAIN Community of Zenodo, which provides descriptors of each file to be searchable within the Zenodo database.

Datasets are not easily classifiable through existing metadata schema, as there are no standards to describe the results of quantum mechanical calculations. The customized metadata schema adopted to define our data includes the following fields:

- work package of provenience
- program name that generated the data
- what is simulated
- labels for the adopted quantum mechanical methods
- labels identifying the kind of outcome (structure, energy, spectroscopy, molecular dynamics)

For identifiability, data deposited in the Zenodo QUANTUMGRAIN Community have an automatic assigning DOI. Zenodo will also handle the versioning of each file.

Outline naming conventions do you follow:

File naming will follow MIT library rules. That is:

date (YYYYMMDD)_work package #_ESR/R# in the QUANTUMGRAIN project_associated program_short but meaningful description of the file content_version number_format extension.

Example: 20210228_WP2_ESR1_CP2K_amorphousice_V001.out.

More extended description are included in read.me files.

Outline the search keywords are provided that optimize possibilities for re-use:

Data are fully searchable as provided by the Zenodo facility (name, authors, keywords, etc.).

Specify what metadata will be created:

We are not aware of standard metadata to describe our dataset. Metadata are created manually by depositors in the deposit form at the repository. We decorate our data by the fields provided by the Zenodo database that are highlighted in each file description.

Outline the approach for clear versioning:

Versioning is automatically handled by the Zenodo facility.

Outline the identifiability of data and refer to standard identification mechanism:

Identifiability of data and standard identification mechanism is provided in the previous sections: data deposited in the Zenodo QUANTUMGRAIN Community have an automatic assigning DOI.

2. Making data openly accessible

Specify how the data will be made available (e.g. by deposition in a repository):

Data will be archived by assigning a DOI. The DOI will be shared when data will be part of the supplementary materials for publications. Data will be openly available without any embargo period. Prior publication, data produced on a daily based will be backed up by each researcher on One Drive cloud service provided by the QUANTUMGRAIN project. They will circulate only internally. Upon publication, datasets will be made openly available in Zenodo.

Specify what methods or software tools are needed to access the data:

Generated data are standard text ASCII files (see Data summary for details) and are accessible through standard text editors available for all main operating systems. To be reused they can be re-edit and piped to the specific program (see Data summary for the list of adopted computer programs).

Specify where the data and associated metadata, documentation and code be deposited:

Data will be deposited in Zenodo, within the QUANTUMGRAIN Community to gather all the material (data, documentation, software, metadata schemas, publications). Zenodo will retain the data for the next 20 years. Data produced on a daily based will be backup up by each researcher on One Drive cloud service provided by the QUANTUMGRAIN project.

Specify how access will be provided if there are restrictions on use:

At the moment there are no restrictions.

Specify which data will be made openly available:

All data (raw data and published papers, conferences, posters, etc.) will be openly available.

3. Making data interoperable

Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability:

As most outputs will be in standard ASCII files, the level of interoperability is high.

Specify whether you will be using standards vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability:

Neither standard vocabulary nor mapping will be provided to our data as they are specific to the computational chemistry community who is used to handle standard output files from the adopted computational packages as listed in the Data summary.

4. Increase data reuse

Specify how the data will be licenced to permit the widest reuse possible:

Data will be released under the Creative Commons Attribution (CC-BY). This can be specified within the Zenodo record description.

Describe data quality assurance processes:

Data will be mostly computer-generated, so manual data entry and relative mistake risk is low. Zenodo provides a MD5 checksum of the file content for all data files. Files are regularly checked against their checksums to assure that file content remains constant. Data discussion will be supervised by the PI. Data in Zenodo will be available for reviewers also before publication, to allow a thorough and sound peer review.

Specify the length of time for which the data will remain re-usable:

data will remain re-usable up to the existence of the corresponding computer codes. Most of the adopted codes are available for more than twenty years ago and are very likely to be available for the next 20 years.

Specify when the data will be made available for re-use:

data will be available after the related scientific paper will be published. We do not plan an embargo of the data.

5. Allocation of resources and data security

Estimate the costs for making your data FAIR in your project:

Costs related to data FAIRification require an in-kind contribution by the QUANTUMGRAIN project. In case data will remain within the 50 Gb limit of Zenodo there will not be extra costs.

Clearly identify responsibilities for data management in your project:

The DMP is under the responsibility of the PI and researchers of the project. Responsible for data management will be the PI, ESRs (Ph.D. students) and Rs (postdocs). Each researcher is informed by the PI that all data should at least be regularly backed up locally on the One Drive cloud service provided by the QUANTUMGRAIN project.

Describe costs and potential value of long term preservation:

The Zenodo long term preservation of data is provided under the tag Longevity.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 865657) for the project "Quantum Chemistry on Interstellar Grains" (QUANTUMGRAIN).