SNF DMP Template

This is a data management plan (DMP) template for a SNF project in computational materials science using the Materials Cloud Archive as a data repository.

Further SNF guidelines on DMPs can be found here.

Please adapt the template to your project!

1. Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

Briefly describe the data you will collect, observe or generate. Also mention any existing data that will be (re)used. The descriptions should include the type, format and content of each dataset. Furthermore, provide an estimation of the volume of the generated data sets. (This relates to the FAIR Data Principles F2, I3, R1 & R1.2.)

1.2 How will the data be collected, observed or generated?

Explain how the data will be collected, observed or generated. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated measurements, data recording standards, usage of controlled vocabularies, data entry validation, data peer review, etc. Discuss how the data management will be handled during the project, mentioning for example naming conventions, version control and folder structures. (This relates to the FAIR Data Principle R1.)

1.3 What documentation and metadata will you provide with the data?

Describe all types of documentation (README files, metadata, etc.) you will provide to help secondary users to understand and reuse your data.
Metadata should at least include basic details allowing other users (computer or human) to find the data. This includes at least a name and a persistent identifier for each file, the name of the person who collected or contributed to the data, the date of collection and the conditions to access the data. Furthermore, the documentation may include details on the methodology used, information about the performed processing and analytical steps, variable definitions, references to vocabularies used, as well as units of measurement. Wherever possible, the documentation should follow existing community standards and guidelines. Explain how you will prepare and share this information. (This relates to the FAIR Data Principles I1, I2, I3, R1, R1.2 & R1.3.)

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

No confidentiality agreements apply.
Where commercially licensed crystal structures were used as input, appropriate identifiers have been put in place to preserve provenance.

2.2 How will data access and security be managed?

[Please adapt if this does not apply to how you manage your data] During research, data is stored on supercomputer premises, protected by access control lists (ACLs) and transferred to workstations using the secure copy protocol (scp).
No personal or other sensitive data are collected in this project.

2.3 How will you handle copyright and Intellectual Property Rights issues?

Outline the owners of the copyright and Intellectual Property Right (IPR) of all data that will be collected and generated, including the licence(s). For consortia, an IPR ownership agreement might be necessary. You should comply with relevant funder, institutional, departmental or group policies on copyright or IPR. Furthermore, clarify what permissions are required should third-party data be reused. (This relates to the FAIR Data Principles I3 & R1.1.)

Data and metadata will be licensed under the Creative Commons Attribution (CC-BY) 4.0 license.

Code written within this project will be released under an open source license (MIT, BSD or GPL).

3. Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

Please mention what the needs are in terms of data storage and where the data will be stored. Please consider that data storage on laptops or hard drives, for example, is risky. Storage through IT teams is safer. If external services are asked for, it is important that this does not conflict with the policy of each entity involved in the project, especially concerning the issue of sensitive data. Please specify your back-up procedure (frequency of updates, responsibilities, automatic/manual process, security measures, etc.)

Data will be stored on the /project file system of Piz Daint at the Swiss National Supercomputing Centre (CSCS) in Lugano. Its features include:

- redundancy provided by the GPFS file system
- daily snapshots for the last seven days
- versioning of files for the last 90 days

For details, see https://user.cscs.ch/storage/recovery/

Data stored locally are backed up weekly to a central backup server.

3.2 What is your data preservation plan?

Please specify which data will be retained, shared and archived after the completion of the project and the corresponding data selection procedure (e.g. long-term value, potential value for reuse, obligations to destroy some data, etc.). Please outline a long-term preservation plan for the datasets beyond the life-time of the project. In particular, comment on the choice of file formats and the use of community standards. (This relates to the FAIR Data Principles F2 & R1.3.)

Curated data will be archived on the Materials Cloud Archive repository, which guarantees preservation of data sets for at least ten years after deposition, irrespective of future funding for the repository.

Temporary files residing on the (super)computers, where calculations were performed, will not be preserved.

4. Data sharing and reuse

4.1 How and where will the data be shared?

Consider how and on which repository the data will be made available. The methods applied to data sharing will depend on several factors such as the type, size, complexity and sensitivity of data. Please also consider how the reuse of your data will be valued and acknowledged by other researchers. (This relates to the FAIR Data Principles F1, F3, F4, A1, A1.1, A1.2 & A2.)

All curated data will be shared on the Materials Cloud Archive, a non-profit data repository for computational materials science hosted in Switzerland (https://archive.materialscloud.org/).

Researchers can find the deposited data via persistent links from the publications or by searching the Materials Cloud Archive.

4.2 Are there any necessary limitations to protect sensitive data?

Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output. Restrictions may be only due to legal, ethical, copyright, confidentiality or other clauses. Consider whether a non-disclosure agreement would give sufficient protection for confidential data. (This relates to the FAIR Data Principles A1 & R1.1.)

Data will be shared at the latest at the time of publication.

4.3 I will choose digital repositories that are conform to the FAIR Data Principles [CHECKBOX]

[X]

Explanation: The Materials Cloud Archive complies with the FAIR Data Principles as defined by the SNF.

4.4 I will choose digital repositories maintained by a non-profit organisation [RADIO BUTTON]

[X] yes [ ] no

Explanation: The Materials Cloud Archive is maintained and supported by non-profit organisations.