**Data management plan for the Academy of Finland September 2019 call**

**Remember to always adapt example answers to describe research data in your project and how it is actually used!**

**PI’s name: fill in**

**Research topic: title of the project**

**Application number: application number from SARA**

**Date: 20 September 2019**

**1. General description of data**

**1.1 What kinds of data is your research based on?**

*Answer these questions in your reply:*

What data will be collected, produced or reused?

Give a rough estimate of the size of the data produced or collected.

What file formats will the data be in?

Is it necessary to use uncommon software to read or use the data?

*Possible template Experimental research project*

Overall, I will follow the [guidance](https://www.aalto.fi/en/services/introduction-to-research-data-management-rdm-and-open-science) for the research data management provided by the Aalto University in data handling and managing procedures. The project data will be obtained from the experimental research. The data content is described in more detail in the research plan. Most of the collected data is measurement data by instruments such as scanning and tunneling electron microscopy (SEM, TEM), confocal Raman microscopy (CRM), and chromatography (GL, LC). The volume of the measurement data is preliminary estimated to be less than 1 GB. Additionally, images from investigations *via* SEM, TEM, and CRM will be collected with a total volume estimated to less than 5 GB. The existing data, for instance experimental structural parameters, will be reused to avoid spending time to do duplicate efforts. Measurement data will be stored as excel files (*.xls*, *.xlsx*) as well as in comma-separated values (*.csv*) or text format (ASCII) that can be easily converted into other formats. Images generated by SEM or CSM will be stored in the *.tiff* format. Other data will consist of generally accepted formats (*.docx* and *.pdf* for documents/manuscripts, *.png*, *.jpg*, or *.tiff* for illustrations *etc.*).

*Possible template for Computational research project*

Overall, I will follow the [guidance](http://www.aalto.fi/en/research/research_data_management/) for the research data management provided by the Aalto University in data handling and managing procedures. The project data will be obtained from computational simulations. The data content is described in more detail in the research plan. The volume of the archives with important files from computational codes can vary up to 100GB for the whole project. The existing data, for instance experimental structural parameters, will be used to justify chosen computational approach. All the data will be in standard data types and generally accepted format (*.docx* and *.pdf* for documents/manuscripts, *.png*, *.jpg*, or *.tiff* for illustrations *etc.*).

**1.2 How will the consistency and quality of data be controlled?**

*Possible template for Experimental research project*

Standardized protocols for the methods and treatments with clear instructions will be prepared, if not existing already, to ensure the repeatability. Quantitative data will be checked for missing data. The PI and the research group leader will review processed data files regularly and before any release.

*Possible template for Computational research project*

All data from computational simulations will be obtained by using professional software which ensures the consistency and quality of the data. The PI and the research group leader will review processed data files regularly and before any release.

**2. Ethics and Legal Compliance**

**2.1 What ethical issues are related to your data management?**

In my research project, I foresee no data related ethical issues, such as research carried out with human cell types, genetic information or personal data. In general, Aalto University is committed to follow the [guidelines](http://www.tenk.fi/en/responsible-conduct-of-research) issued by the Finnish Advisory Board on Research Integrity (“*Responsible* *conduct of research and procedures for handling allegations of misconduct in Finland*”) on good scientific practice, how to handle violations against it, as well as valid legislation. We are also following [the European Code of Conduct for Research Integrity by ALLEA](http://www.allea.org/wp-content/uploads/2017/03/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017-1.pdf). Aalto University has a [Research Ethics Committee](https://www.aalto.fi/en/services/research-ethics-committee) whose tasks include ex-ante evaluations of the ethical nature of research projects, and resolving any problems related to research ethics arising at the university. Furthermore, the EU’s General Data Protection Regulation’s ([GDPR](http://ec.europa.eu/justice/data-protection/index_en.htm)) and the Finnish Personal

Data Act requirements will be applied if necessary.

**2.2 How will data ownership, copyright and IPR issues be managed?**

*Answer these questions in your reply:*

Are there any copyrights, licenses or other restrictions that prevent you from using or sharing the data?

As in all externally funded research projects of Aalto University, the data will be owned by the university. However, during the mobility phase, data and results might be jointly generated together with the researchers from the [] University, and, possible, with the third parties. These data and results will be owned either by the organisation in question in the proportion in which they have contributed to the creation or jointly owned. Before starting the co-operation, the terms of exercising the joint ownership will be specified by agreement, if necessary. The aim is to publish the data as openly as possible. However, before the publishing of the data, the possibilities of commercial exploitation will be evaluated together with [the Innovations Services](https://innovation.aalto.fi/). If the research results in patentable inventions, the IPR of the inventions will be protected and treated according to the law applied to academic researchers.

**3. Documentation and metadata**

*Answer these questions in your reply:*

How will you document your data to make them findable, accessible, interoperable and reusable for you and others?

What kinds of metadata standards, README files or other documentation will you use to help others to understand and use your data?

*Possible template for Experimental research project*

All data will be documented in laboratory notebooks (e.g., parameters applied, the date, the code for the labels and abbreviations used, *etc.*) as well as in the excel files. The resulting raw data will be recorded either in the same laboratory notebooks or in electronic form, depending on the used measurement technique. The data generated during the project will be converted to an electronic format (*.xlsx*, *.docx*, *.pdf*, *.txt*, *.tif*). The folder structure will separate the different sample sets. The filenames will contain all descriptive and necessary information (the date at which the data was collected and, an abbreviation for the respective method or treatment applied, sample type) so that it will be easy to recognize them afterward. Datasets will be documented with the descriptive metadata (e.g., title, year of publication, dataset´s creator, description, keywords, *etc*.) which ensures understandability and findability of the data in the future. If needed, README files will be created for datasets to ascertain their re-usability, reading and interpretation.

*Possible template for Computational research project*

The filenames will contain all descriptive and necessary information (an abbreviation for the method and basis set applied, name of the studied structure, *etc.*) so that it will be easy to recognize them afterward. The folder structure will separate the different datasets. All relevant information from simulations will be documented in the excel files. Datasets will be documented with the descriptive metadata (e.g., title, year of publication, dataset´s creator, description, keywords, *etc*.) which ensures understandability and findability of the data in the future. If needed, README files will be created for datasets to ascertain their re-usability, reading and interpretation.

**4. Storage and backup during the research project**

**4.1 Where will your data be stored, and how will they be backed up?**

*Answer these questions in your reply:*

How will the data be backed up / recovered in the event of an incident?

All data will be stored on the institutional servers at Aalto University and will be backed up by the Aalto University IT Services; more details on [the Aalto Storage services for research data](https://www.aalto.fi/en/services/storage-services-for-research-data). The services include a snapshot feature and regular backups that make file versions automatically to recover from unwanted deletions - tape backups provide also system-level disaster recovery. If needed, cloud storages services will be used for storing non-confidential data. Aalto University provides such services as OneDrive, Google Drive, Dropbox with an unlimited capacity. All mentioned services are free of charge.

**4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?**

All the Aalto University laptops include automatic data encryption with Bitlocker, and secure file transfer over the network with a VPN solution, that will be utilized in the project. During data analysis, the data will be accessible only to the members involved in the research project. The PI of the project will control access to the project data. The data can also be stored in a repository that secures their backup and where you can choose the level of openness.

**5. Opening, publishing and archiving the data after the research project**

**5.1 What part of the data can be made openly available or published?**

*Answer these questions in your reply:*

Where and when will the data, or their metadata, be made available?

If your data or parts of them cannot be opened, explain why and describe where and how the metadata will be available

The data related to the published materials will be made available for example in [Zenodo](https://zenodo.org/), in [IDA Storage Service](https://openscience.fi/ida) or in [B2SHARE](https://b2share.eudat.eu/) repositories under the Creative Commons license CC BY 4.0 to maximize the re-use of the data. In the repositories, the deposited data will be included with the required standard metadata to ensure the re-usability. The services provide also persistent identifiers (e.g. DOI, URN) to promote data citation. In addition, the repositories give the option of including README files to make the uploaded data files understandable. ACRIS (Aalto Current Research Information System) system will be used to enter the resulting publication information and the metadata of data, e.g. for the internal and external reporting purposes and to improve the visibility of research outputs. The open access publications will also be available in the institutional ACRIS repository. All research outputs will be publicly available in [the research.aalto.fi/en/](https://research.aalto.fi/en/) platform and the publications in the national [Juuli portal](http://www.juuli.fi/). In addition, the metadata of data will be showcased in the national [Etsin Research Data Finder](https://etsin.fairdata.fi/).

**5.2 Where will data with long-term value be archived, and for how long?**

The data with recognized long-term value will be archived in the national [Fairdata PAS](https://www.fairdata.fi/en/fairdata-pas/) service provided by the Finnish IT Center for Science. The storage period will be decided during the project.

**6. Data management responsibilities and resources**

*Answer these questions in your reply:*

Who will be responsible for specific tasks of data management during the research project life cycle?

Specify the extra management costs in the budget and explain it in the application text.

The PI will be responsible for data management during the research project life cycle. In general, the data management is integrated to the research practices, and it is difficult to make detailed estimation about the use of time, costs or other efforts put alone on management. Furthermore, all facilities to be used for managing, preservation, sharing, and publishing the data are free of charge and do not require additional help from experts. However, the data handling and documentation during the data processing will benefit from additional aid to ensure the understandability and findability of the published and preserved data. I approximate that the extra effort on data management will require altogether 4 weeks.