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

**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?**  *Example answer:*  The consortium PI is in charge of data management activities and will support the consortium using the [guidance](https://www.aalto.fi/en/services/introduction-to-research-data-management-rdm-and-open-science) for research data handling and management provided by Aalto University. The project will use and generate qualitative data (such as models, guidelines and policies), quantitative data (such as [example of data]) and our own measured data (such as [example of data]). Members of the consortium will also conduct surveys and interviews as well as potentially record some sessions during interaction workshops. Interviews will be documented using audio and video recorders. Survey data will be collected using [online survey tools/ other methods, please specify]. The collected data will include direct identifiers, such as a person’s voice, and indirect identifiers of the participants. No sensitive topics are addressed (**OR** Sensitive personal data are collected from or of the participants: data concerning [racial or ethnic origin; political opinions; religion or philosophical beliefs; trade union membership; data concerning health; sexual orientation or activity; genetic and biometric data for identifying the person]). The members of the consortium may use either publicly available background data (e.g. publications or open access databases) and own data, or they may purchase background data e.g. from [name of database from which data is purchased].  The consortium will also generate and store project administrative data, such as meeting minutes, materials for reporting, cost sheets and legal documents. We expect about 10 GB of data collected yearly by the consortium. Experimental data is both in form of numbers and images. Some is gathered manually, but mostly via computers. 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 [] will be stored in*.tiff* format. Other data will consist of generally accepted formats (*.docx* and *.pdf* for documents/manuscripts, *.png*, *.jpg*, or *.tiff* for illustrations *etc.*). Recorded audio (e.g. from interviews) will be saved in *.aif* or *.wav* format and videos in *.avi* or *.mj2*.  **1.2 How will the consistency and quality of data be controlled?**  *Example answer:*  All data will be obtained using professional software, which ensures the consistency and quality of the data. Standardized protocols for the methods and treatments with clear instructions will be prepared, if not existing already, to ensure the repeatability of measured data. Quantitative data will be checked for missing data. The subproject leaders [**OR** WP leaders] 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?**  The consortium 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.  *Example answer 1 (if no ethical issues are foreseen):*  No studies will be conducted in which ethical issues may arise, such as research carried out with human beings or animals, genetic information, sensitive patient information, GPS locations or other personal data. 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 personal data is handled in the project.  *Example answer 2 (explain in detail the ethical issues within your research):*  The project collects personal data for research purposes (see section 1). The individual participants will give their written consent for participating in the project, and they will be informed of data collection, and subsequent storage and analysis in the project in accordance with the GDPR. All participants are legally competent adults (**OR** Permission will be sought from the guardian of a minor or a person under guardianship for their participation in the study), the participation is voluntary and all the personal data will be managed according to the EU’s General Data Protection Regulation ([GDPR](http://ec.europa.eu/justice/data-protection/index_en.html)). The collection of identifiers are minimized. Individuals are entitled to review the data concerning them, request deletion of their data, and stop participating in the data collection at any point in time. Only data that can be efficiently anonymized will be deposited in public repositories. As no sensitive personal data is collected nor the research design requires ethical review, no review will be sought from the Aalto Research Ethics Committee (**OR** Ethical review will be sought from the Aalto Research Ethics Committee. Data protection impact assessment (DPIA) is included in the request for statement to be submitted to the Ethics Committee’s review).  **2.2 How will data ownership, copyright and IPR issues be managed?**  *Example answer:*  Each consortium party will own the data and results that it will generate. During mobility and cooperative activities, data and results might be jointly generated together with consortium parties or with partners. These data and results will be owned either by the organizations in question in the proportion in which they have contributed to the creation, or jointly owned. The terms of exercising this joint ownership will be specified in a consortium agreement. The aim is to publish the data as openly as possible (**OR** In this project, data includes personal data and will be treated as mentioned above. **OR** In this project, data includes confidential data, which will not be published). However, before publishing the data (not valid for personal data and other confidential data), the possibilities of commercial exploitation will be evaluated, for instance with the support of the innovation services of the consortium parties (such as [Aalto 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**  *Example answer:*  The consortium is committed to following the FAIR data principles in order to make our data findable, accessible, interoperable and reusable. Such strategy will promote joint use of data during the research and ensure the interpretation of data and their re-usability after the project. During the project kick-off meeting, under the guidance of the PI, the consortium will define in details the common practices on how to implement the FAIR principles in details. Overall, all data will be documented in notebooks (e.g., parameters applied, the date, the code for the labels and abbreviations used, *etc.*) as well as in 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/ type of data. 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 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?**  The consortium will ensure that all data are stored, accessed and backed up according to a pre-defined common strategy, which will be detailed during the kick-off meeting. We will use data storage systems that take into account the appropriate security level of the data in question and the needs born from our multidisciplinary collaboration. Each consortium party will use their IT services to store data during the project. During the project, data will be stored primarily on the institutional secured servers, which are supported by a snapshot feature and regular backups that make file versions automatically to recover from unwanted deletions – tape backups provide also system-level disaster recovery.  *Example answer 1 (no sensitive or confidential data):*  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.  *Example answer 2 (sensitive or confidential data):*  The data will be transferred to other systems only if necessary, such as when archiving selected anonymous data.  **4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?**  *Example answer:*  All consortium institutions hold high security standards, and data is stored in password-protected environments that ensure secured access to the data only to authorized users involved in the research project. Laptops include automatic data encryption with e.g. Bitlocker, and secure file transfer over the network with a VPN solution. During data analysis, the data will be accessible only to the researchers involved in the project. The consortium PI will control access to the project data. Each subproject PI [**OR** WP leader] will manage the data within each subproject [**OR** WP]. The data can also be stored in a repository that secures their backup and where you can choose the level of openness.  If personal data is transferred from one consortium institution to another, or the consortium institutions otherwise process personal data jointly in the project, the consortium institutions will make appropriate agreements with one another to ensure compliance with the GDPR.  **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?**  The consortium is committed to sharing the data and other relevant materials, e.g. scientific articles, with society and the academic community in order to validate the results and to foster societal impact. However, prior to the publishing of data and results, the requirements related to confidentiality, legal, ethical and security issues will be assessed, and the possibilities of commercial exploitation (e.g. IPR) will be reviewed. Each Subproject PI is responsible for managing data publishing and sharing in each subproject and for the protection of IPR. [**OR** WP]  *Example answer 1 (no personal data or confidential data):*  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 a Creative Commons license chosen case by case by the owner of the results (e.g. 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 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. We will use institutional information systems of the consortium parties, such as [ACRIS](https://www.aalto.fi/en/services/introduction-to-acris) (Aalto Current Research Information System),to enter the resulting publication information and the metadata of data, e.g. for internal and external reporting and to improve the visibility of research outputs. The open access publications will also be available in such institutional repositories. All research outputs will be publicly available in institutional platforms, such as [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/).  *Example answer 2 (personal data or confidential data):*  To protect the identity of the participants, the data that cannot be efficiently anonymized will not be released. However, the metadata of all data sets will be entered to institutional information systems of the consortium parties, such as ACRIS (Aalto Current Research Information System), and the national [Qvain Research Metadata Tool](https://qvain.fairdata.fi/) systems for internal and external reporting purposes and to improve the findability and visibility of the datasets. The open access publications will also be available in such institutional repositories. The project will adhere to the default policy of the Academy of Finland by either sharing data via FSD’s services or by sharing metadata of research data though national catalogues.  **5.2 Where will data with long-term value be archived, and for how long?**  *Example answer 1 (no sensitive or confidential data):*  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.  *Example answer 2 (sensitive or confidential data):*  The anonymized data with recognized long-term value will be archived at [the Finnish Social Science Data Archive (FSD)](https://www.fsd.uta.fi/en/). The archive is responsible for curating, preserving and disseminating the dataset. The storage period will be decided during the project.  **6. Data management responsibilities and resources**  *Example answer:*  The consortium PI will be responsible for data management during the research project life cycle and will provide overall guidance to all collaborative parties. Each Subproject PI [**OR** WP leader] is responsible for managing data publishing and sharing in each subproject [**OR** WP] and for the protection of IPR. In general, data management is integrated into research practices, and it is difficult to make a 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 of the data are free of charge and do not require additional help from experts. However, data handling and documentation during data processing will need additional work by the researchers and data experts of the consortium members, in order to ensure understandability and findability of the published and preserved data. We estimate that the extra effort on data management will require altogether 4 weeks. | 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?  For example, in handling sensitive data, protecting the identity of participants, or gaining consent for data sharing.  Are there any copyrights, licenses or other restrictions that prevent you from using or sharing the data?  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?  How will the data be backed up / recovered in the event of an incident?  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  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. |