JYU's supplementary organisational instructions to the Academy of Finland DMP template [retrieved on September 13, 2021: Data management plan - Academy of Finland (aka.fi)]

1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

Guidance: No need to repeat your Methods section here; make appropriate reference to the Research Plan where possible. In 1.1., data is understood in the broadest sense: they can be e.g. code, survey, interview, observation, machine or instrument collected, models, physical samples, compiled databases, etc. Using standardized and interchangeable data formats ensures the long-term usability of data.

The most important thing here is to show that you understand what kind of formats your data is contained in, and how it can be utilized, stored and described. As this varies considerably, we cannot give exact model answers, only some examples. Note that your data may go through several formats from collection to storage to analysis to archival to publication to re-use: try to describe all relevant steps. For example, perhaps you first gather observations or interviews or measurements in hand-written notes, then transcribe them using a spreadsheet program and save the file as .csv, then perform some conversions using R and save the work as .r data while you're working on it, and finally archive and publish it in simple and secure .txt. If you use some equipment that produces or manipulates the raw data, describe shortly what kind of data it provides (and if possible, that the data is compatible with some metadata standards).

Model clauses:

"The interactions will be coded with standardized CLASS and ECCOM observation instruments. Parts of the interactions will also be coded and analysed by using conventions used in discourse and conversation analysis. Recall interviews will be transcribed and content coded. Eye-tracking data will be coded according to the teacher fixations and saccades on areas of interests."

"Data consist of video and audio recordings, eye-tracking recordings, interview and observational transcripts in generally accepted formats. Most of the data will be quantified with SPSS program. The data formats for SPSS-data are in .sav format and for video and audio recordings .mp4 and .mp3 format. Transcripts and content codings will be in .docx format. Eye-tracking recording are in json and .mp4 format (Tobii Analysis Pro software is needed for analyzing the eye-tracking data). Quantitative data will be processed and analyzed by using SPSS and MPlus statistical programs; Atlas.ti to analyze and handle qualitative data. All other materials, such as questionnaires are saved as .pdf-files. Used software and formats are based on open standards to enable data reuse, interoperability and sharing. The data content and specific methods as well as analyses are described in more detail in the research plan."

"The research data created in this project comprises mainly the database on politicians' wives. The data consists of generally accepted formats (.csv, .rtf). I will compile excel tables with information on all the study subjects in the research area and period. I will also make individual files on those study subjects that I select for a closer study based on the availability of source material. The file of one politician's wife consists of a biographical document and individual documents that include transcriptions from original archival documents. To make the source material in old, mainly handwritten Swedish usable, I will transcribe the relevant text from the archival documents and contemporary literature. The size of the data produced in the project will be max. 50 GB."

"The primary form of data is analytical mathematical models in the form of equations. All analytical models will be included in peer-reviewed publications and will therefore not require external data storage. Secondarily, if in some cases analytical modelling proves impossible, simulation models will be utilised. In such cases the simulation code itself will need to be stored, along with instructions on how to run the code. Simulation models will be implemented in matlab, using the .m file format. Finally, a literature review will produce data sourced from previously published literature. This data will be stored as a combination of text (stored in .txt, .doc, or .pdf format) and statistical data in excel format."

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

Explain how the data collection, analysis and processing methods used may affect the quality of the data and how you will minimise the risks related to data accuracy.

Model clauses:

"Files will be stored using checksums that are used to ensure that data is not corrupted when copying, transmitting and saving it. We will also ensure that conversions to new file-formats (e.g. .docx to .pdf) are made maintaining the original information."

"University of Jyväskylä provides GitLab version control environment for research projects, hosted on university server. This can be used to secure documentation, consistency and recoverability of changes to data files, even where multiple persons are involved in handling them in different stages of research."

2. Ethical and legal compliance

2.1 What legal issues are related to your data management (for example, GDPR and other legislation affecting data processing)?

JYU policies and guidelines:

JYU Data privacy guide for researchers

Quick links to privacy notice and a description of processing forms

Jyväskylän yliopiston ihmistieteiden eettinen toimikunta

The University of Jyväskylä Human Sciences Ethics Committee

Considerations and examples:

a. Does the research include processing of personal data?

Note: The researcher should identify situations when the collected and processed data include personal data.

If personal data are processed in the research, data protection legislation will apply to it. You can identify a research subject with personal data directly or indirectly. Personal information includes not only strong personal identifying details such as one's name, personal identity number, or photo; for example, social, economic, cultural, physical, physiological, psychological or genetic factors can together or individually form personal data. Personal information is evaluated from the point of view

of whether even a small group (e.g., the research subject and/or his/her relatives) could use the collected information to identify who an individual person is.

Model clauses:

"The research does not include processing of personal data."

"The research includes processing of personal data. Personal data are processed in the research (for more information, go to http://www.fsd.uta.fi/aineistonhallinta/fi/tunnisteellisuus-ja-anonymisointi)."

Direct, strong indirect identifiers and/or indirect tags are collected from the subjects. Collected personal data include, for example, the subject's name, contact information, place of residence, educational background as well as psychological and financial information."

b. Does the project include personal data that belong to a specific personal data group, i.e., sensitive personal data?

Note: Identify and describe situations when collected personal data are sensitive as well as the legal basis for processing sensitive personal data.

Sensitive personal data include the following:

- racial or ethnic origin
- political opinions
- religious or philosophical beliefs
- trade union membership
- genetic data produced by analysing a biological sample
- biometric data for the purpose of uniquely identifying a natural person
- data concerning health
- data concerning a natural person's sexual behaviour or sexual orientation.

If sensitive personal data are processed in the research, the processing must be based on Article 9 of the General Data Protection Regulation (GDPR). Consent from the data subject and scientific purpose of the research are emphasised as a lawful basis when processing sensitive personal data as part of research.

Model clause:

"Sensitive personal data are processed in the research (data concerning health). The data are processed with the explicit consent of the data subject."

c. Criminal convictions and offences

Note: Identify and describe situations when collected personal data include criminal convictions and offences or related precautionary measures. Also identify the legal basis for processing the data. Processing criminal convictions and offences in research is based on consent of the data subject.

Model clause:

"The project includes processing of data concerning criminal convictions and offences with the consent of the data subjects."

d. Data processing and related roles

Note: Identify and describe possible roles in the research and take care of the possible agreements required by the GDPR.

Controller refers to the natural or legal person, who, alone or jointly with others, determines the purposes and means of the processing of personal data. Where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers. Processor refers to a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. The processor cannot decide how personal data are used or what information is collected.

For example, when two universities together as joint controllers determine what personal data are collected and used for scientific purposes, both universities are controllers. When these joint controllers submit personal data to a third party, such as to a transcriber or database administrator, the transcriber and database admin are both processors.

The GDPR requires controllers and joint controller to conclude data processing agreements with all the processors. Joint controllers should conclude an agreement showing the responsibilities and roles of data controllers in the processing of personal data.

Model clause:

"University A is the data controller. The processors are internet-based survey service provider Webropol Oy and University B, which conducts the analysis of data, as well as Tietoarkisto, where the data will be archived at the end of the research. Data processing agreements have been made with the processors."

e. Where are the personal data obtained from?

Note: Describe where the processed personal data are being collected. The data can be obtained from the data subject directly or, for example, from another register.

Model clause:

"The processed data are obtained from the data subject directly on the basis of consent."

f.   Safeguards to protect personal data

Note: Describe what safeguards will be implemented to protect the personal data.

You must follow appropriate and risk-based safeguards in the processing of personal data. The research data must be processed anonymously if the objectives of the research allow it. If anonymisation is not possible, use pseudonymization of data whenever possible. If pseudonymisation is not used, justify your reason for it.

Safeguards regarding the processing of personal data may include the following:

- measures to improve the skills of staff dealing with personal data (training, advice and guidelines)
- appropriate data security (pre-defined where the data are stored)
- measures to prevent access to personal data
- drawing up a data protection impact assessment (DPIA)

- research has a principal investigator and an appropriate research plan
- limitation of the retention of personal data (data will be destroyed or anonymised upon completion of the study)
- data subjects cannot be identified from the research results
- agreements required by the GDPR have been reached.

Model clause:

"The data will be pseudonymised in the analysis phase. The data will be archived in a pseudonymised manner *[define where].* It is not possible to archive data in anonymous form because it is a follow-up survey. The personal data safeguards include…."

g. When to make a DPIA (Data Privacy Impact Assessment)?

Note: The researcher must identify a possible need to carry out a DPIA.

A DPIA is used to evaluate the impacts of processing on the protection of personal data. According to the Office of the Data Protection Ombudsman, a DPIA has to be made every time the processing causes a high risk for the data subject's rights and freedoms.

High risk factors for the data subject include the following:

- a large number of data subjects
- a large amount of personal data per data subject
- processing of sensitive data (e.g. health data)
- processing personal data of vulnerable natural persons (e.g. children)
- automatic decision-making (e.g. credit decision)
- systematic monitoring of the data subject.

Furthermore, a DPIA is required in the following cases especially:

- new technologies are used for the processing of personal data
- processing on a large scale of personal data relating to criminal convictions and offences or special categories such as data concerning health, ethnic origin, political opinions, religious belief or sexual orientation
- a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person
- large-scale systematic monitoring of a publicly accessible area.

Whenever there is an exception from the right to be informed, a DPIA must be made. If the DPIA indicates that a high risk is still being encountered, the evaluation will be submitted to the supervisory authority prior to the processing of personal data.

Model clause:

"The project poses a high risk to the data subject's rights and freedoms because the research focuses on the personal data of vulnerable natural persons (children) and a large amount of personal data is collected per data subject. Before the start of processing the personal data, a data protection impact assessment must be made."

2.2 How will you manage the rights of the data you use, produce and share?

Instructions:

In projects that are consortions with Finnish Academy and JYU, rights to all results (including) are transferred to the University. Even in such cases it might be worth to write in the plan that permit to use the data is agreed on and will prove no problem. IN ALL CASES emphasize that the ownership and rights to distribute and use data are always agreed on so that the FAIRness of data is not endangered due to vagueness of ownership issues!

See JYU policy decision:

[Research data policy at the University of Jyväskylä](#)

Other resources:

[IPR–neuvonta JYUssa](#)

[IPR counseling at JYU](#)

[IPR ja tutkimustulosten kaupallistaminen -opas](#) (Intranet Unossa)

[The guide to IPR and commercialization of research results](#) (in JYU Intranet Uno)

For further information of research agreements, please contact:

[Lakipalvelut JYUssa](#)

[Legal Services at JYU](#)

Model clauses:

"The ownership and access rights of data shall be agreed on when creating the data management plan, prior to the start of actual research."

"The ownership and access rights of data was agreed on when the data management plan was created, prior to the start of actual research."

3. Documentation and metadata

How will you document your data to make them findable, accessible, interoperable and re-usable for you and others? What kinds of metadata standards, README files or other documentation will you use to help others understand and use your data?

Instructions:

Describe the types of documentation that will accompany the data to provide re-users with any necessary details to prevent misuse, misinterpretation or confusion. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data.

Tips for best practices:

- Identify the types of information that should be captured to enable other researchers to discover, access, interpret, use and cite your data.
- Create a subfolder titled /DOCUMENTATION and collect there all the documentation files you create. Store the DOCUMENTATON folder alongside the data for the duration of your project.
- Describe all the types of documentation (your /DOCUMENTATION subfolder, README files, code books, vocabularies, lab notebooks, etc.) you will provide to help secondary users to find, understand and reuse your data.
- Name your files in a uniform manner that helps you and others to make sense of their contents and create a logical folder structure for them.
- Check whether your chosen repository requires specific metadata standards and prepare to make use of them in the data publication phase.

Model clauses:

"The project shall name a team member responsible for monitoring and enforcing these data documentation rules. These documentation and metadata and their publication will ensure FAIRness of the data.

When anyone saves data on the project's shared network storage space (provided and technically administered by University of Jyväskylä, including backups, access control and security), that person shall also update and maintain this documentation:

The storage root as well as every sub-folder shall include a README.txt file. Each README.txt file describes the current naming conventions and purpose of all folders and files in that level of storage organization. Any additions, deletions or changes to naming con- ventions or purposes are immediately updated to README.txt.

[About "master data files": if you use version control software, or otherwise use a method to consistently distinguish work copies from master data files, mention and describe that here briefly.]

Each master data file [or folder depending on type of data] will be clearly named as such [you can enter your naming convention here], and each will always be accompanied, in the same folder, by two files:

1) ABOUT_filename.txt/ABOUT_foldername.txt which contains link to an initialized metadata entry in the University's research information system Converis. *[The preliminary entry can remain private and unpublished, just for your own use. If, and when, you want to publish your data, it is easy because the metadata entry is already there.]*

2) METHODS_filename.txt which/METHODS_foldername.txt contains link to, or text of, description of methods used to obtain, process and document the data, in detail sufficient to allow replication and understanding and usage of the data for other purposes.

Both the ABOUT and the METHODS files, and the preliminary metadata entry in Converis, will be created immediately as the data gathering begins, and amended and updated throughout the research project as data is building up. These will then be readily available to convert to machine-readable formats upon archival or publication of the metadata in the University's repository JYX, and the eventual publication of the data in [*name chosen repository*]."

4. Storage and backup during the research project

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

Guidance: During the research project, your data (and other files) can be stored on secure storage space of 100GB in the Nextcloud cloud storage system managed by University of Jyväskylä, with automatic daily back-up. The university can also provide even larger storage space, but in that case please budget 0.15€/GB/year for the space exceeding 100GB.

In addition, your data can be stored in the university's version control environment GitLab, which also features automatic back-up, and/or encrypted CollabRoom environment developed for sensitive data.

Model clauses:

"Research data (except sensitive data) of the project will always be stored in a secure cloud environment Nextcloud, provided and managed by the Digital Services of University of Jyväskylä. These network drives are also used to share files safely between participants."

"[IF project handles sensitive data:] Sensitive data is stored and operated in university's highly secure CollabRoom environment that has been tailored especially for this purpose."

"Systems will take automatic backups of the data to prevent catastrophic loss of data. In addition manual backups of master data files will be taken regularly *[insert suitable time interval]* and always before any major file-format or data conversions."

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

Model clauses:

"The PI will have the main responsibility of data management, monitoring it and granting access to the data."

"Paper questionnaires will be stored in offices at both consortium sites that cannot be accessed without permission and that are locked with doors that opens with a registered card key card. Paper questionnaires will be scanned and digitized into electric form. This data as well as all other data (e.g., web-surveys, eye-tracking data, audio and video data files, interview data) will be saved and stored on a data protected network drive that is password protected. The files record an automatic access log."

"Non-sensitive data will be available to all researchers of the project. For sensitive data, PI will define who will need and get access to the data. Access to the data will be documented and PI will be at any point able to tell who has access to what data."

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

Guidance:

Please note: If your project cannot make the data public, e.g. for data privacy or IPR reasons, describe here the justification for it. Both University of Jyväskylä and the Academy of Finland have explicit policy that all data should be published unless there are specific justifiable reasons to keep the data secret.

Ideally, eventually your data should be deposited to international repositories where other researchers in your field typically find it and can cite it. Many fields of research have well-established repositories that "everyone" in that field uses - look for them first, by asking colleagues or exploring what repositories

publications in your field are listing. These resources may help finding a repository: https://www.nature.com/sdata/policies/repositories, https://fairsharing.org/databases/.

There are also widely recognized international general-purpose repositories, both non-profit such as Zenodo and EUDAT and Dryad, or commercial such as figshare.

Organisations in Finland maintain both general-purpose data repositories, such as IDA, and especially in social sciences and humanities subject-field specific repositories such as FSD and FIN-CLARIN. Finally, you can always deposit your data in University of Jyväskylä's own repository JYX, if a more suitable home cannot be found.

Model clauses:

"Basic project-level metadata will be published in the University of Jyväskylä's institutional JYX archive immediately when they can be considered sufficiently complete, even where the data itself is not yet public. (For description of creation and curation of metadata entries see Section 3 of this plan.) Metadata can then be found via ETSIN and OpenAIRE metadata search engines. The JYX archive provides the metadata with a persistent identifier, thus promoting their findability and interoperability.

Datasets themselves, complete with full description of methods, will be published as datasets are completed in *[JYX / FSD / Langaueg Bank of Finland / IDA / Dryad / Zenodo / some other repository]*.

"Sensitive portions of the data will be anonymized and published along with other data."

"Sensitive portions of the data cannot be anonymized and thus cannot be openly published. Sensitive data will be archived in JYX, only the metadata will be public, and access to the data can be requested and granted on certain conditions: *[describe the procedure and terms on which access and right to use the data can be granted]*."

"Sensitive parts of the data cannot be stored and will be disposed of after the project is finished."

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

You can categorise your datasets according to the anticipated preservation period:

1. data to be destroyed upon the end of the project
2. data to be archived for a verification period, which varies across disciplines (e.g., 5–15 years)
3. data to be archived for potential re-use (e.g., for 25 years)
4. data with long-term value to be archived by a curated facility for future generations for tens or hundreds of years.

Guidance: For how long "long-term" is, depends on where the data is archived. If you use archive other than JYX, you should look for their terms and describe those here.

Model clauses:

"The University of Jyväskylä will store all data archived in JYX for minimum of 10 years, in the format originally deposited in, but does no special packaging or continuous curation to guarantee long-term integrity and usability."

"Data with long-term value will be proposed to forthcoming true long-term storage solutions provided by Finnish Ministry of Education and European Open Science Cloud, and University of Jyväskylä provides support for enabling this archival where possible. National and university-level policies for determining eligibility for Tutkimus-PAS are currently being developed, but not available yet."

6. Data management responsibilities and resources

6.1 Who (e.g., role and institution) will be responsible for data management?

For JYU's instructions to roles and responsibilities in data security, please see:

in Finnish: Tietoturvaohjeet henkilöstölle, in English: Data security principles

Model clauses:

"PI will be responsible for data management. Responsibilities for specific issues have been described in earlier sections."

"Right to access the data is controlled by *[PI or the project, or someone specially dedicated to this]*, and technical access control is provided by the Digital Services of the University of Jyväskylä."

"Non-sensitive data will be available to all researchers of the project via an institutional cloud storage space, shared via secured remote connection in case of external partners)."

"For sensitive data, PI acts as the administrator of the project's CollabRoom space and will define who will need and get access to the sensitive data in that workspace."

"Access to the data will be documented and PI will be at any point able to tell who has access to what data."

"The data that is not archived will be pceserved in the Nextcloud cloud storage service of the University of Jyväskylä for at least ten years."

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

Model clauses:

"Final preparation of the metadata and publishing it will require 1 to 2 weeks of work. Publishing the metadata in a standard form and with a persistent identifier will be supported by the Open Science Centre of the University of Jyväskylä. The Finnish Social Science Data Archive (FSD) will support the publishing of the data. The FSD will preserve the data in a file format suitable for reuse."

"No additional resources are required due to the nature and size of the data (see section 1.1)."