

# JYU organisational instructions to General Finnish DMP template [version: Tuuli working group. (2021, April 30) Finnish DMP evaluation guidance. Zenodo. <http://doi.org/10.5281/zenodo.4729832>]

## 1 General description of the 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.

#### Instructions:

- Briefly categorize and describe your data in different types (e.g. already-existing data that you reuse, raw data you collect yourself, analysis data you process from it). You can use a table or list (see [example](#)).
- 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, you might 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 are 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 consists 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 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 source material used in the project consists mainly of mmth-nnth century archival documents and contemporary literature that is preserved in archives and libraries (mainly X and Y archives). The material is available either in digitized form or in microfilm and the use of it does not require permits. I will also use published registers that are available either online or as printed publications. To make the microfilmed material usable, I will take photographs of the documents in archives and libraries or scan

the microfilmed pages to my own USB drive. Other relevant material, like contemporary literature, can be retrieved from library databases or ordered directly from libraries (.pdf).”

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

### 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](#)

### Instructions:

#### 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)?**

Where applicable, 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?**

Note: In projects that are consortiums with Finnish Academy and JYU, rights to all results (including) are transferred to university. Even in such cases it might be worth to note that permit to use the data is agreed on and proves no problem.

IN ALL CASES please emphasize that ownership, 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!

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.”

See 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 contact:

[Lakipalvelut JYUssa](#)

[Legal Services at JYU](#)

### 3 Documentation and metadata

#### 3.1 How will you document your data in order to make the data findable, accessible, interoperable and reusable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to 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 /DOCUMENTATION subfolder where you collect all the documentation files you create, storing it 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 conventions 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 the data be backed up?

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 100 GB.

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 NextCloud cloud storage service provided and managed by the Digital Services of the University of Jyväskylä. Cloud service is also ideal solution for sharing research data between collaborators.”

[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 task.”

“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.”

“Non-sensitive video data will be stored in the University's secure and backed-up Researchvideo.jyu.fi storage system.”

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

#### Instructions:

Please see: in Finnish: [Tietoturvaohjeet henkilöstölle](#); in English: [Data security principles for personnel](#)

#### Model clauses:

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

"[Non-sensitive] data will be available to all researchers of the project via NextCloud cloud storage system."

[In case of sensitive data] "PI acts as administrator of 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."



## 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 its metadata, be made available?

#### Instructions:

Considerations: Are you using data formats and software that enable that ensure long-term validity of data? For how long will and must different parts of the data be stored? Are there any embargo times for the data?

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.

Note: If your project shall not make the data public, describe here the justification for it. According to the data policy of the University of Jyväskylä, all relevant data should be made publicly available at the end of the project except in case of a specific, justifiable reason (such as data privacy or IPR) to keep the data secret. If the data themselves cannot be made available, publishing the metadata suffices.

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 repository [JYX](#), if a more suitable home cannot be found.

#### Model clauses:

"Metadata entries will be published immediately in the University's JYX publication repository when they are considered sufficiently complete, even if the data themselves are not yet public. (For description of creation and curation of metadata entries please see Section 3 of this plan.) The metadata will then be indexed in the national Metax system and become searchable in e.g. Etsin and Research.fi metadata catalogues.

Dataset(s) themselves, complete with full description of methods, will be published in [JYX / a certified, field-specific repository / Zenodo]."

"Sensitive parts of the data will be anonymized [describe how] and published along with other data."

"Sensitive parts of the data cannot be anonymized and thus cannot be made openly available."

"Sensitive data will be stored 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 made available 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?

### Instructions:

For how long the data will be available depends on where the data is archived. If you use an archive other than JYX, check their specific terms and describe those here. The University recommends the use of certified, field or data type-specific data repository that is widely used in your field.

### Model clauses:

"University of Jyväskylä will store all data archived in JYX indefinitely (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 (for example role, position, and institution) will be responsible for data management (i.e., the data steward?)

#### Instructions:

Considerations: Here you are expected to show that you're aware that data management needs people responsible for it. It might also need specific resources (time, money, computational facilities, support from data steward etc.).

#### Model clauses:

"Final preparation of the data and publishing it will require *[best estimate of time and effort]*. This will be supported by Open Science Centre of University of Jyväskylä."

"Anonymization of sensitive data will require *[best estimate of time and effort]*."

### 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)?

Do you need to pay to reuse existing data? Have you estimated costs in time and effort to prepare the data for preservation and sharing? Have you specified your data management costs in your research budget? Who is responsible for each data management activity? How are responsibilities split across partner sites in collaborative research projects?

If you already cover this in the earlier sections, e.g. "See earlier sections in this plan."