

Data management plan for the Freudenthal Institute

Wouter van Joolingen, scientific director Freudenthal Institute
Version: 14 September 2018

This document describes the way research data can and should be handled for projects carried out under the responsibility of the Freudenthal Institute (FI). The plan is written in the form of a practical manual, providing guidance for its users. Typical users who should heed this document are:

- UU Employees working for the FI.
- PhD students working under supervision of a FI supervisor, including external PhD students who are not employees.
- Postdocs working on FI-based research projects – with or without employee status.
- Students (bachelor and master) carrying out research projects under supervision of a FI supervisor.

Basic principles of handling research data

Much of the data collected in FI research concerns data obtained from people: interviews, test scores, video observations etc. This automatically implies that privacy is a major issue in collecting and handling data. Therefore, before deciding on storing and collecting data privacy and other ethical issues need to be resolved. Furthermore, requirements by funding bodies and publication agencies (conferences, journals) need to be taken into account. The principles in this section provide the major considerations.

1. Ethical considerations

- The design of your study should be approved by the ethical commission.** In their regulations, a data management plan for the project is also part of the procedure, including the way you obtain consent from the participants.
- Ensure you have consent.** Never collect personal data without prior consent of the subject, or in case of minors younger than 16, their parents. The way consent is given may depend on circumstances. A written consent by the participants or parents, a spoken consent at the start of an interview can suffice. Sometimes schools include a generic consent in the contract with parents. Check which way of consent applies to your project. Store written consents with your data (signed PDFs of the consent forms). The ethical commission will need to approve your study – including the way you plan to obtain consent.
- Inform participants.** Participants have a right to know which data you store about them and about the purpose of the study. If telling them everything before the data collection would hinder your research, debrief them afterwards. Also share your contact information so participants can inquire about the data you store about them. They have a legal right to (1) inspect

the data you have stored about them and (2) to be forgotten, i.e. that you delete their personal data. Structure your data in such a way that inspection and deletion of personal data are possible.

- d. **Be careful and minimal.** Collect what you need, but do not ask more of participants' time and effort than needed for answering your research question. Also, do not store more than you need. If there is no need to store personal details then don't.
- e. **Share your results.** Make your results available to your participants, for instance by mailing them a link to a publication you write about the study. Even better: make a laymen's summary and mail it to the participants or publish it online.
- f. **Be kind and polite at all times.**

2. Ownership

- a. Whereas the university will be the formal owner of the data produced and collected in UU projects, each data set will have a specific owner.
- b. The standard case is that the owner is the main researcher of the specific research project under which the data is produced or collected
- c. In the case of a master student, the owner is the supervisor of that student
- d. In the case of a project that involves multiple persons *within* the FI, the project leader is the owner of the data, unless the project team decides otherwise.
- e. In the case of a project that is a collaboration of FI with other institutes, data ownership will be settled in the data management plan of the project. The same plan will contain agreements on data use by all partners in the project.

3. Storage, transport and analysis

- a. Electronic data are volatile. Hard discs may crash, USB sticks may be lost and laptops may be stolen. So, always have a backup at the data repository. Details on how to do that are elaborated below.
- b. Data is especially in danger when transported. The preferred way is to transfer data electronically over an encrypted connection. The following means are possible:
 - i. Directly over a VPN connection to the Betastor server (see below).
 - ii. Using SURF file sender (<https://www.surffilesender.nl/>) with encryption on. The password should be transferred over a different medium, e.g. a text message.
 - iii. Using SURF drive.

If electronic transfer is not possible, data may be transferred on a portable medium, *if that medium is encrypted*. Make sure it is stored on Betastor on the first possible occasion. This only applies when transporting data from the source of collection (e.g. at a school) to the university.

- c. Treat results of data analysis in the same way as original data. The SPSS file or typed and scored manuscript has probably taken a great amount of time. Also these data may still be privacy-sensitive.
- d. Paper material, such as filled in questionnaires: Scan them and treat the resulting files as original data. Store the originals in a safe place.

4. Be privacy aware

- a. Ensure that only people who should work with your data have access. If for instance a student will code your data, he or she needs access. They should sign a contract stating that they will use the data only for the purpose of the work agreed. A standard contract is available at the FI website.
- b. The owner of the data is responsible for the integrity of research and data management!
- c. People working on your data should keep to all privacy and storage rules stated in this and other relevant documents, indicated by the ethical commission.
- d. The above specifically applies to sharing data with others. Never do that using portable media such as USB sticks or portable hard discs are not acceptable. Always use protected electronic transfer using <https://www.surffilesender.nl/>. This can handle files up to 500 Gb.
- e. If you want to use data that can be traced back to one of your participants, outside the original use, for instance a video in a presentation or a picture in your thesis, ask for renewed specific permission.

5. Long-term storage

- a. In order to make research transparent and traceable, original and processed data should be kept until five years after the most recent publication based on those data has been released. The storage place for these data will be Betastor. Betastor is an external server, dedicated to the Faculty of Science ([https://ict.science.uu.nl/index.php/Storage_research_data_\(Betastor\)](https://ict.science.uu.nl/index.php/Storage_research_data_(Betastor))).
- b. After a researcher or student has left the institute where the data were collected, the institute's data manager will move data to the Archive area of Betastor, which is maintained by the data manager of FI.
- c. Data should be documented by or under responsibility of the owner: when was it collected, which publications are related and when they were published.
- d. Owners remain responsible for updating the data documentation even after they have left the FI. If the owner wants to keep using the data in a new position, the data will remain at FI, but the owner can keep using it – provided an agreement is made on the conditions of further use.
- e. In anonymized form, data sets may be published on open databases or as online supplement to a publication, under the condition that the scientific director provides consent.

Storing research data

This section contains the technical and procedural information needed to store research data on Betastor, the science faculty's storage facility for research data that is safe and privacy compliant.

Connect to the data storage

For FI, the Betastor data storage consist of three partitions:

- FI-Researchers. This one is for projects of individual researchers, mainly PhD students and postdocs.
- FI-Students. For master students collecting data for their thesis project.
- FI-TLL. For use specifically for guests in the teaching and learning lab.

The following steps lead to obtaining access to the storage.

For employees and PhD students:

1. ask the FI data manager, currently Nathalie Kuijpers (N.Kuijpers@uu.nl) to give you access to FI-Researchers. In your mail mention your solis id (the one you use to log in on the intranet, e.g. jooli001). It will take a few hours between granting access and actually being able to access the partition.
2. Mount the partition on your computer as an external drive. On [https://ict.science.uu.nl/index.php/Storage_research_data_\(Betastor\)](https://ict.science.uu.nl/index.php/Storage_research_data_(Betastor)) you find instructions on how to do so. Wherever you see projectname in the instructions replace that by FI-Researchers.
3. Open Explorer/Finder. You will see a folder with the name of your solis id. This is the folder where you should start storing your data.

For students, the *supervisor* should ask the data manager to provide access to FI-Students. Subsequently, a similar procedure applies, of course with FI-Researchers replaced by FI-Students. Users of the TLL should contact the TLL manager, currently Frans van Dam, for instructions.

What to store in the data storage?

Basically, you have to store all data you collect. This may include but is not limited to:

- Raw video and audio data collected from interviews and observations
- Scans of filled in questionnaires
- Data tables collected from online questionnaires
- Log files from software used by participants in your studies
- Written notes from observations
- Eye-tracking data
- etc

Also the data storage is meant to store processed data such as:

- Annotations of videos
- Tables with scores from analysing think-aloud protocols, interviews or other observations
- Tables with scores from questionnaires
- SPSS files used to process data
- Result tables of analyses

So, basically use the research data storage as a place to store all data, raw and processed that is necessary to trace what you have done in collecting and processing your data for research. Finally, a data folder should contain a readMe file documenting the data that is stored in that folder. If there is a need to maintain a link from anonymized data to the

original person from whom those data was obtained, the file containing these links should be encrypted with a password to prevent accidental reading by others. The passwords for such files should be kept by the data manager. As soon as the need for maintaining the link – usually at the moment the data set is complete, such files should be deleted.

Collaborative projects generating research data

If you work in a research project with multiple researchers from FI, you can use the “projects” folder in FI-Researchers. Make a subfolder with the project name and store the data there, using the same guidelines as for individual projects. Please note that the project folder is accessible to *all FI Researchers*. This means that you should not use it for data that is too sensitive to be exposed to their eyes. This is subject to change in the near future.

Hygiene in structuring and storing data

The data storage is meant for long term data storage, as to be able to reconstruct and underpin published results. The general rule that data should be stored five years after publication date. This implies that the data needs to be stored possibly years after (many of) you have left the university. Therefore, it is important that future colleagues will know what to do with the data without having to consult you. For this, the following rules should be taken into account about data hygiene:

- **All data and data only:** following the description above, store all data in the storage so when needed, it can be retrieved on request without anything missing. *Do not store anything else.* Do not use the data storage as backup for your holiday pictures, documents you write, or anything that is not research data.
- **One folder per study:** below your personal root, create folders for each study you perform, e.g. leading to an article and or thesis chapter. Provide meaningful names to the study and start the name with the month data collection started in the name. Example: 201805 - Thesis Study 1, meaning that the data collection of this study started in May 2018.
- **Include a ReadMe:** in each folder for a study write a ReadMe file in Word format or plain text. On the top of the readme file include a status, which can be: “Data in use”, “Submitted – date”, “Published – date - link”. Data in use means you are still working on it. The other two status descriptions indicate the publications based on the data in the current folder. There may be more publications linked to one data set of course. Include status lines for each of them. A PhD. or master thesis is also a publication. The status will be used to decide whether data can be deleted at a certain point in time, so update for each publication! Furthermore, the ReadMe should contain a description of the data. If there are subfolders describe what is in there. Register dates for changes to the data set. An example ReadMe is provided as an appendix to this document.
- **Be transparent:** Look at your data folder through the eyes of someone who does not know your work. Will he or she understand what the files mean? Will he or she be able to decide whether the data needs to be retained or can be deleted? Would another researcher be capable of using the data?

Sticking to these rules would ensure that the data is in good hands, even after projects have ended. If in any doubt, discuss with others (colleague, supervisor), the data manager or the research director.

Appendix 1: Example of ReadMe file

all information in this file is just for illustration and does not apply to any real existing data.

Metadata

Owner of the data in this folder: Wouter van Joolingen

Supervisor (if applicable): <none>

Name of the study: NRO Kortlopend Blended in Beta

Time period in which data were collected: October-December 2017

Kind of data: Videos and Annotations

Status

[Choose between: In Use, Submitted – Date, Published – Date – Link to online article]

Multiple lines possible.

Published NRO Final report – 19-10-2017

In Use

Notes

Folder “videos” contains the raw video material from six cameras in the teaching and learning lab.

Folder “scores” consists of excel tables with time-stamped fragments that were scored

Folder “logs” contains log files of the software used. Software to read is available in the same folder.

....

Log

Date	Event
19-09-2017	Added video files session 1. Named Session_1_<camera number>.mp4
26-09-2017	Added video files session 1. Named Session_2_<camera number>.mp4
28-09-2017	Added Session_1.xls with codes.
	...

Appendix 2 - Example of letter for consent form for parents of under 18 participants (Dutch)

Betreft: Onderzoek '...'

Beste ouder/verzorger van,

Mijn naam is [naam onderzoeker] en ik ben [functie]. Ik verricht onderzoek op het gebied van [onderzoeksveld] aan de Universiteit Utrecht. Dit onderzoek draagt bij aan de verdere ontwikkeling van [...]. Voor dit onderzoek heb ik ook leerlingen nodig om [reden].

Uw zoon of dochter is uitgenodigd om deel te nemen aan [de eerste/tweede/etc] fase van het onderzoek. Daarin richten we ons op [...]. Het onderzoek duurt ongeveer [xx] minuten. Deelname of stoppen met deelname staat los van elke vorm van schoolbeoordelingen.

Tijdens het onderzoek zal uw kind ook dingen opschrijven en tekenen. Het hele onderzoek wordt gefilmd. Uw kind is [wel/niet] herkenbaar in beeld. De film-, geluids- en geschreven materialen worden geanonimiseerd opgeslagen en alleen gebruikt voor wetenschappelijke publicaties en presentaties.

Uw kind heeft een proefpersoon formulier meegekregen. Op dit formulier kunt u tekenen wanneer u akkoord gaat met deelname van uw kind aan het onderzoek. Voor vragen aangaande het onderzoek kunt u via onderstaand email-adres bij mij terecht.

Vriendelijke groeten,
[naam onderzoeker],
[emailadres onderzoeker]

Appendix 3 – Example of consent form (Dutch)

Proefpersoon formulier

Datum: _____

Naam: _____

Geboortedatum: _____

Leeftijd: _____

School: _____

Klas: _____

Email-adres: _____

Handtekening voor akkoord deelname: _____

Als de leerling ten tijde van het onderzoek jonger is dan 16 jaar:

Naam ouder/verzorger: _____

Handtekening voor akkoord deelname: _____

Appendix 4 – Example of a contract on processing data by a person not directly involved with the project.

The undersigned:

[name of owner of the data], hereafter called the owner of the data

and

[name of person working with the data], hereafter called the [assistant, use other term if more appropriate]

Agree that assistant:

1. is assigned a task involving access to research data collected within the research project [name], performed under the responsibility of owner, within the context of the data management plan of the Freudenthal institute,
2. gets access to the data for the sole purpose of executing this task,
3. will ensure that the data cannot be accessed by any other person other than agreed with the owner,
4. will destroy any copy of the data in his or her possession on completion of the task

Signed,

Owner

Name:

Place, date:

Assistant

Name:

Place, date:
