GUIDELINES FOR DATA MANAGEMENT (COLLECTION, STORAGE AND PROCESSING OF DATA) IN STUDENT RESEARCH

In their collection, processing and storage of the data pertaining to their research, students will have to comply with laws and regulations. This applies especially when these data include subjects' personal details. Below, we will briefly describe the three most common situations, i.e.

a) new data collection,
b) the use of existing data, and
c) naturalistic observation,

and include some guidelines for the responsible use of the data in all the stages of the data collection.

Situation A. New data collection

Step 1: Asking for consent

Make sure to ask the participants you want to involve in your study for permission to collect their data. Active consent (based on an informed consent form with signature; see appendix) is required when the collection includes *(particular) personal data* (as explained in the appendix), video recordings, audio recordings or invasive questionnaires.
To compose an adequate informed consent form and accompanying information letter, see our instruction document [link].

Step 2: Collecting data

Make sure that questionnaires are completed anonymously: there should be no name, date of birth or student number printed on the cover page or in the (online) questionnaires, nor should you ask for it. If you use online questionnaires, make sure that you uncheck the IP addresses when you print out the answers to the questionnaire (Qualtrics, for instance, has this functionality).
When collecting data from the web, make sure that the collection of the data provided by each subject online does not conflict with their reasonable expectations with regard to their privacy.
- For video data, see our instruction documents [link]. For video recordings, use secure devices available for collection from the Techsupport desk (https://Techsupport.fss.uu.nl/) whenever you can and make sure you save the video data as soon as possible by uploading them to the secured domain of the faculty server and then remove them from the device. (For advice and access to the faculty server, contact the Techsupport desk.)
- Note for your supervisor: if the data is collected from an external body such as a school, institution or organisation, check with the faculty's privacy officer (privacy-fsw@uu.nl) whether it is necessary to enter into a **data processing agreement** with this external body.

Step 3: Processing and storing the collected data

Anonymisation

Anonymising means that you strip the data of all the information that can be traced back to a person (see also appendix). You should do so in consultation with your supervisor. Your supervisor can ask Techsupport for further advice, if necessary:

https://Techsupport.fss.uu.nl/ or via privacy-fsw@uu.nl.
- Save the anonymised data as soon as possible in the appropriate folder on the faculty server. You can have a folder created by Techsupport staff.
- Next, delete the data from the device on which they were originally stored; also delete the documents from the "recycle bin" of your device. Never store the data in the cloud! If it is not possible to anonymise the data (because you need to be able to approach participants again for follow-up study, for instance), pseudonymise them.

Pseudonymisation
Pseudonymising data means that you assign a unique code to each person in the data collection. You then make two data sets: one with the code and the identifying information, the so-called *key*, and one with just the code without the identifying information, the so-called pseudonymised data (see appendix). The two datasets are stored separately. You will use the pseudonymised dataset in your analyses. The unique code allows you to link the key to the pseudonymised data at a later date to enable you to contact the participants again if necessary. Always consult your supervisor when you need to pseudonymise. Your supervisor can ask Techsupport [link] for further advice, if necessary.

- Save the pseudonymised data as soon as possible in the appropriate folder on the faculty server. You can have a folder created by Techsupport staff.
- Save the key on the secured faculty server as soon as possible.
- Next, delete the pseudonymised data as well as the key from the device on which they were originally stored; also delete the documents from the "recycle bin" of your device. Never store the data in the cloud!

Linking newly collected data to already existing data
Sometimes you need to link newly collected data to existing data of the same persons. For example, you have collected new data from students and want to link it to their previously collected exam scores. The connection between these data will most likely be made through their personal information. Make sure that, whenever possible, you use this personal data only at the location where the new data collection takes place.
Preferably, you anonymise the data once the link has been established; otherwise, pseudonymise them.

Storage periods
The principle is that the data collected by a student for the purpose of writing an assignment or thesis is stored. The storage period is similar to the periods for storing study results (2 years for papers, 7 years for theses). If the data is used for a scientific publication, the storage periods as referred to in the Guidelines for the archiving of academic research for faculties of Behavioural and Social Sciences in the Netherlands, Version 2, July 2017, apply.

Step 4: Data transport
If you take the data from the school/institution to the university to be able to continue to work on your project, do this

- via a secure connection, such as Surffile sender: https://www.surf.nl/surffilesender-veilig-en-versleuteld-grote-bestanden-versturen.

- by borrowing a secured device from the faculty through the Techsupport desk:https://Techsupport.fss.uu.nl/.
- by storing the data on Surfdrive: www.surfdrive.nl (and removing it from the device). Make sure that only you and your supervisors (can) have access to the data.
- See the data storage guidelines: https://la0171.its.uu.nl/.

B. USING EXISTING DATA

The student uses data available at UU (supervisor's project).

Check your access to the data

- Check whether and how your access to the data has been arranged.
- Make sure that you only have access to data that you need for your study, which is to say: only to anonymous data, or, if pseudonymisation was justified, the pseudonymised data.
- If you need access to the data from home, use a secure connection. For advice, contact the Techsupport desk.
- Make sure that you do not download any data on your own laptop.

C. NATURALISTIC OBSERVATION

In naturalistic observation, the student collects data in the street, in the field or online by observing participants or other matters. It is important that you do not write down any personal data or make any video or audio recordings. You will store the data in the same way as described above for anonymous data.

Appendices:

1. Example of an Informed Consent Form
2. Definitions of the basic concepts relating to data management and privacy

Definitions of privacy, data management and ethics terminology

Version: 28 June 2019

Definition of personal data

Any information about a person that may lead to the identification of that person, such as their name, address, telephone number, email address, date of birth, identifiers and location data (IP addresses, smartphone IMEI number, cookies, MAC address) and fingerprint. Separately or combined, all these data can lead to identification. The exception is their first name only, unless this name is very rare.

Definition of particular personal data

Particular personal data is extra sensitive personal data; its processing may have an especially negative impact on a person. Examples of such particular personal data are their health data, genetic data, biometric data, ethnicity, religion, political opinions, trade union membership, criminal past or sex life. Particular personal data only exist if they can be traced back to individuals, which is the case if, for instance, they are requested in combination with personal data. A combination of particular personal data can also be (indirectly) identifying in a certain context.

The definition of 'processing' within the meaning of the GDPR (broadly speaking)

The legal term 'processing' covers any action involving the personal data – when it is consulted, collected, recorded, organised, stored, retrieved, used, distributed or destroyed, for instance. If, for instance, a student completes an internship at a school and they are shown a class list, they process personal data and are expected to keep it confidential, which is to say they are not to copy the list and only use it for the purpose at that specific moment. The same applies to (patient) records.

The legal basis for the processing personal data

This concerns the question whether the participants have given their permission to process their personal data. Participants can give their consent through a so-called Informed Consent Form (see below). If they have not given their consent, the researcher might invoke the "justified interest" principle or one of the other principles of the GDPR[1].

Informed consent

This term refers to both information and consent. It means that participants are provided with adequate information about the study, after which they can agree to participate and give their permission to use their data.

Data minimisation

This means requesting only the personal details necessary for the purpose of the study. Think carefully about the personal data you need. Are all these details really necessary for the purpose of your study? For example: if the age of the participant is necessary to answer

[1] https://www.autoriteitpersoonsgegevens.nl/nl/onderwerpen/avg-europese-privacywetgeving/mag-u-persoonsgegevens-verwerken - hoe-weet-u-of-u-persoonsgegevens-mag-verwerken-6310

your research, their year of birth or age would suffice, so why ask their date of birth? Do you really need their addresses to contact them again, or would it be possible contact them by email?

Anonymisation/pseudonymisation

Anonymity exists when personal data has been fully deleted, in which case the GDPR no longer applies. As a rule, we work with *anonymous* data. If they are not anonymous as yet, we anonymise them. Anonymising is more than leaving out names and contact details. The point is to make it entirely impossible to identify a person through any reasonable means. If someone in the town of Appingedam provides their profession as a dentist, chances are that this detail can be traced back to certain individuals.

Sometimes it is necessary to store personal data with a view to contacting participants again at a later stage (in longitudinal studies, for instance). In such cases, the data *are pseudonymised* (by assigning each participant in the database a code that is used as a key to replace all their identifying information, and storing the identifying information, together with the key, separately from the original data). The research data with the code are the *pseudonymised data*. The *key* is the personal data with the code. Pseudonymisation does not mean that the data are no longer personal, for the data can still be traced back, albeit with more difficulty, since it would require the key as well as the pseudonymised data. The key is stored on a secure server.

Storage of the data

Personal data are preferably deleted, but if they are saved, they should be stored separately from the research data. See also the definitions for anonymisation and pseudonymisation. Describe how the data are stored. If you intend to make the data publicly accessible (open access) in the future, state so.

Storage period

Personal data is preferably deleted as soon as it is no longer required for the study. The raw data is stored for at least 10 years (and at least 15 years if the study is, or was, subject to the WMO Medical Research (Human Subject) Act.

Access

The people who will have access to the data should be named, and it should be stated whether the eventually anonymised data will be made available for open access.
In principle, participants have the right to inspect their personal data for as long as they are stored and to request removal of their personal data. Such requests can only be made if the data has not been destroyed as yet.