Contents

Data Storage Protocol UvA Psychology, updated March 2020

This is a 9 rule protocol.

Each researcher needs to comply with the following storage protocol. For questions, contact the psychology data stewards: Jasper Wijnen, Tinka Beemsterboer (datasteward-psy@uva.nl)

1. *Data storage before publication*. During the research process, all research data and related materials collected should be properly stored, on university servers (which have automatic backup).

2. *Published data.* When a paper is accepted for publication, it is obligatory to register your publication in Pure (https://pure.uva.nl/admin). Data should be placed on UvA servers in the published data section (see where-to-store section). Uploading your data should be finalized maximally 1 month after (online) publication. An audit will check this regularly. Pseudonymized data may be stored on an on-line repository. *Figshare* (http://uvaauas.figshare.com) or the Open Science Framework (OSF).

3. Data sharing. If the data are suitable (see *Sensitive data*, Appendix A), we recommend that you share your data with other researchers. *Figshare* and OSF facilitate public access of research data. The researcher can control the level of access to his or her data (e.g. open, under embargo, only with explicit consent, non-public). If anonymity cannot be guaranteed or if data are highly sensitive (see below), you should keep the data non-public. If you are uncertain about this choice, please contact the data steward.

4. ***Data belonging to other parties***. In case you make use of data that are owned by organisations, other universities, or persons, who would object to storage on our repository, make a Figshare metadata entry without uploading the actual data.

5. *Who stores?* All staff members are lead scientists, and responsible for appropriate data storage. For a given project responsibility applies to researchers that are first author on a paper, or collected the data. Co-authors should check if their co-authored publications are registered in Pure and add if necessary (especially if the first author is from a different university).

6. PhD students and Bachelor and Master students. PhD students and post docs are responsible for their own data storage (in case of PhD students, this should be checked by the promotor). BA and Master students are required to submit a data storage folder to their supervisor, containing their thesis and the data and research materials upon completion of their project.

7. Unpublished data. Unpublished data (e.g., collected by students) from finalized projects can be stored on the faculty fileserver. You can use your personal folder or your students folder.

8. ***Storage duration***. All (raw, unprocessed) data should be stored for at least 15 years after the publication date. The University library will check this period.

9. ***Folder structure and contents.*** We recommend a specific *structure* (detailed on the next page of this protocol) for research data folders. A read-me text can be added if the folder structure deviates from the recommended structure or is otherwise not obvious.

STRUCTURE AND CONTENTS OF YOUR PUBLISHED DATA FOLDER

Data and other files associated with a published paper should be placed in the published data storage folder on the fmgstorage faculty fileserver. Please request such a folder via https://top.socsci.uva.nl/?page_id=294. Use the structure below or include a readme file to describe your subfolders.

PUBLISHED DATA FOLDER: ***Folder name should contain name of lead scientist.***
- Each folder has subfolders, containing data from different research projects.
- Names of subfolders should clearly reflect the contents (e.g., year of data collection, name of research project, or name of journal in which these data were published).

CONTENTS OF EACH FOLDER

a. **ETHICS**
- Ethics Protocol and Approval (PDF exported from EC site)
- Information brochure, materials and debriefing brochure (as uploaded in the EC submission).

b. **METHOD, MEASURES, MATERIALS**
- The **experiment scripts** (e.g., the Presentation or PsychoPy task code) and stimuli, accompanied by instructions on software dependencies and/or hardware requirements that have to be met to get the task to run.
- PDF of paper-pencil questionnaires.
- If online survey software is used (e.g., Qualtrics), it is often possible to export the questions in PDF. Always include a description of the *exact version* and *platform* of the software used.

For custom software (written expressly for an experiment), a zip file with the software itself may be uploaded with installation instructions and a clear mention of the platform (e.g., Windows 10 or higher).

c. **DATA COLLECTION**
- Appropriate configuration parameters (e.g., EEG configuration files, VSRRP driver files)
- File with *lab log* with entries identified by date and experimenter; subjects identified by participant number, any subject-related materials or comments
- *Raw* data files (e.g., output files)

d. DATA ANALYSIS
- All scripts and syntax-files used to transform and/or analyze the data (e.g., Excel files, SPSS.sav and syntax files, R scripts, EEG and MRI analysis scripts).
- A list of dropped subjects plus reason for exclusion
- A *code book*: description of all variable names and labels with sufficient detail to understand both the raw and processed data
- The resulting transformed **data** that formed the basis for statistical analyses in the published paper.

- ***e.*** **PAPERS OR REPORTS**
- The final, submitted manuscript (PDF, refer to as final publication).

Data Leak

When your laptop / external harddrive / phone / usb device is stolen and it contains research data, you have to deal with a data leak. A failure to report can result in fines for the university. *In case of a data leak you have to act quickly*.
Inform the UvA-cert team cert@uva.nl (020 525 3322) about the loss of data. Do this immediately, even during evenings or weekends. They will respond and tell you what to do. Inform the data steward as well via datasteward-psy@uva.nl.

Collecting and storing research data

Ethical approval

All research projects where data is collected directly from a subject need ethical approval. You can find information on ethical approval procedures on (www.lab.uva.nl/lab/ethics). The EC will evaluate whether adequate security precautions are in place given a project's potential privacy risks. Research projects where data is not collected directly from human subjects, using existing databases for example, can skip ethical approval but should still fill out the data-management and privacy-impact questions.

The faculty fileserver FMGSTORAGE

It is obligatory to store published research data on the faculty fileserver. Different types of accounts exist for this server:

- The published data folder: obligatory for all staff, used to store data and other files for **published work** according to the recommended file structure on page 3. Request access via https://top.socsci.uva.nl/?page_id=294 (since this is now obligatory we'll look into automatically creating a published data folder for all staff ASAP)
- The personal folder: used to store data and other files for **ongoing research**, optionally a 'students' folder associated with your personal folder can be requested if you have students collecting research data. An account can be requested via https://top.socsci.uva.nl/?page_id=235
- The project folder: used to store data and other files for **ongoing research**, for collaborative projects involving multiple UvA staff members. An account can be requested via https://top.socsci.uva.nl/?page_id=273

You can add or remove users to students or project folders using the form at https://top.socsci.uva.nl/?page_id=278. Be sure to remove student access after a project has finished.

Information on how to access these folders on your machine (after your account request has been processed) can be found here: https://top.socsci.uva.nl/?page_id=399.

Acquiring informed consent

Use the informed consent form approved by the ethical review board. For the correct procedure with regard to accepting consent we make a distinction between 1) research projects that are purely on-line, 2) research projects where there is face-to-face contact between subject and experimenter.

For research projects that are exclusively on-line informed consent is presented on-screen with an accept and decline button. In this case, we do not ask for the subject name.

When a subject is physically present in the same room as the experimenter at any time during the research project informed consents are signed by name. **Do not place a subjectID on a consent form.** It should not be possible to tie research data to a name using the filled in informed consent form.

Collecting data in the lab or in the field

Personal or sensitive data that is collected on a desktop or laptop harddrive should be removed and placed on a UvA file server after each experimental session, or after each full day of data collection if the device is not transported or left unattended during the day. Only use devices which have proper security measures in place (software updates, anti-virus, password protection, full disk encryption a.k.a. bitlocker etc.). If you use a laptop for collecting data, download the UvA manual for security measures at https://medewerker.uva.nl/en/content-secured/az/security/manuals-security-measures/manuals-security-measures.html and follow the steps in the manual.

Collecting survey data on-line

Survey data collected on-line should use the Qualtrics framework. Please create your survey using the UvA portal for Qualtrics at https://nlpsych.eu.qualtrics.com/login . Please make sure to toggle of the automatic collection of ip addresses in Qualtrics settings. Like any experiment, subjects should sign informed consent prior to taking part in the survey. This notably includes screenings used to include/exclude participants (these often collect data with high leak-impact).

Where to store *ongoing* research data

Data should be saved in a location with backup. At the FMG, the following options for researchers are available without costs:

1. We recommend saving research data on the faculty file server (FMGstorage). Access can be requested via https://top.socsci.uva.nl/?page_id=235.
2. You can store your data on the cloudserver surfdrive (surfdrive.surf.nl). You can share data with other staff when appropriate via surfdrive. Students do not have access to surfdrive.
3. File servers provided by the central IT department (ICTS):
    a. Every employee has a home directory (Windows users know this as the ‘h-disk’, Mac users can connect to the ‘personal home’ folder). This allows you to save a limited amount of data (up to 5 GB).

b. Each employee can work in a shared folder on the 'p-disk' (for Mac users 'public folder'). There is no access for researchers from outside the UvA. Requests for such a folder can be done by the ICT contact person.

4. If you have large data sets that you don't use anymore, the Technical Support can help you archiving it. There are several options:
    a. an external hard disk (anonymized non-sensitive data only) in combination with a tape backup
    b. Personal and sensitive data that you don't expect to use for a longer time, but cannot be deleted, can be send to the data steward for encryption and storage. Send this via Surf Filesender (https://www.surffilesender.nl/), which has an encrypt during transfer feature.
    c. the archive centre of Surf Sara

How to handle *sensitive* research data

For a detailed description of what constitutes 'sensitive' data see Appendix A, summarized, such data either contains directly indentifying information or information on certain sensitive topics (i.e. health status). The first rule of handling such data is: avoid having it. Only collect and process sensitive data if doing so is *critical* for your project.
To minimize processing of sensitive data:

- Replace directly identifying information with pseudonym subjectIDs. Split-off identifying and sensitive data from the rest of the research data. Personal data gathered solely for administrative purposes should be deleted as soon as no longer needed.
- Do not ask unnecessarily detailed questions that may be used to re-identify subjects even after pseudonymization (e.g. ask for year of birth rather than date of birth).

Where and how to store sensitive data:

- Identifying data that cannot be deleted should be encrypted or stored on a UvA file server. Further information on encryption can be found in Appendix C.
- Never store unencrypted sensitive data on local disks, or removable hardware. Avoid dropbox, google drive or any non-UvA cloud server.
- Never send personal data by email.
- You can share sensitive data with project collaborators via a project folder on fmgstorage or you can use Surf Filesender (with encryption option flagged) to share information.

Where to store *published* research data

All research data collected for publications where the first author is a researcher affiliated with UvA Psychology should be placed on the faculty fmgstorage server using the structure described on page 3. See the FMGSTORAGE section above for information on how to access this server.

Paper data. If your research is on paper the department 'DIV' of the university can archive it. Please see https://top.socsci.uva.nl/?page_id=227.

Informed consent (written). You can leave Informed consent papers of finished studies at the TOP desk. They will archive it. Be sure that the research identifying document is attached. Please see https://top.socsci.uva.nl/?page_id=227.

Informed consent (digital): See Appendix B

Sharing data with third parties

If your project involves sharing personal data with an external organisation a processing agreement might be required. You can find more information on these agreements at https://medewerker.uva.nl/en/content-secured/az/privacy/processing-agreements/processing-agreements.html. Data stewards at psychology can provide advice concerning such agreements. Please contact datasteward-psy@uva.nl for more information.

Publically sharing data

Research data that was collected for a published research project can be made public and shareable with other researchers. Only data that does not contain identifying information nor sensitive data can be shared with the scientific community this way. The default sharing platform is Figshare, but other similar platforms, notably the OSF, are of course also fine. The publication should be registered in Pure.

Figshare can be found on http://uvaauas.figshare.com. Everyone with an uvanetID can login. Since *Figshare* is protected by *two-factor* authentication, you will have to register your mobile phone number and identify yourself (with passport) at the library.

- Login in *Figshare* with your uvanetID.
- Arrange the data on your computer in the way you would like to upload them, we recommend the example folder structure given in this protocol. Zip those files. Create a figshare 'fileset item', fill in the metadata and upload the zip.
- Extensive instructions on the use of figshare can be found via: https://knowledge.figshare.com/
- A researcher can opt not to upload certain data on *Figshare*, for example because the data is highly sensitive. Check if it is possible to split the data and upload non-sensitive parts. Please still make a metadata entry on *Figshare* for data that was not placed on *Figshare*. Encrypt the sensitive data and store it on a UvA file server.
- Data can be shared in *Figshare* under a specific license (i.e. non-commercial use only see: https://support.figshare.com/support/solutions/articles/6000089895-what-licences-are-available-on-figshare-)
- Sharing data can be done with limitations:
    - within a specific time-period and/or,
    - requests for the data are granted only after the researchers consent and/or,
    - requests for the data are granted only after consent from RDM support
- If you plan to share your data in any form, please announce this in your informed consent.
- Only pseudonimized or anonimized data can be publically shared.
- If another researcher or institute works with you on a research project and they collect your data (including personal data), then they have to sign a processing agreement with you. This also applies when you hire a company to collect your data.

Appendix A - Different levels of sensitivity

We can categorize data on the basis of the impact that a potential leak of these data would have:

- Severe impact: Sensitive data with direct identifiers
- Moderate impact: Sensitive data where direct identifiers have been removed, risk of inferential identification remains present *or* Non-sensitive data with direct identifiers
- No impact: Non-sensitive anonymized data. These data can be publically shared.

To prevent data leaks and/or to limit the impact of a potential leak all information leading to persons should be separated from the rest of the data. Identifying data should not be uploaded to figshare, but placed on UvA servers instead and/or be encrypted in storage. Contact information that was gathered with the sole objective of making an appointment for data collection may be deleted all together. This also applies to especially sensitive data that was collected as part of a screening and for inclusion/exclusion purposes.

Note that European privacy laws make a distinction between anonymous and pseudonomynous data. Anonymous data is only those data where the link between direct identifiers and the data has been *irreversibly* severed, *and* the risk of indirect identification is negligible. Such anonymous data, where the researchers nor the research institute is able to re-identify subjects, is free from all restrictions and can be shared at will.

When data cannot be made fully anonymous separate and secure data belonging to the following categories:

- Direct identifiers: i.e. full name, address, telephone number, e-mail, social security number, passport or drivers-license number, facial images, fingerprints, hand writing, credit card/bank account nr, date of birth combined with birthplace, any information uniquely associated with one individual.

- Indirect identifiers: a combination of variables that can be used to form a profile that allows identification of an individual. Researchers should analyse their data for these issues and take steps to prevent inferential identification. High risk factors include: small and special populations, individual known to be in study/self-disclosure, longitudinal data, variables with high level of detail (e.g. date of birth rather than year of birth).

- Very sensitive data (e.g., health information, drug-use, criminal records, school records, political or religious affiliation, sex-life/sexual orientation, information that can cause harm, legal jeopardy or damage reputations, any information about minors etc.) or video data require *encrypted storage*. Ethical approval procedures will inquire into the sensitivity of data that will be collected in a project. Ethical approval may depend on the commitment of a researcher to follow enhanced security procedures with regard to the storage of sensitive data.

TIPS to limit sensitivity/identification risks

- Do not use participant names in filenames but use an anonymous identifier
- Avoid use of unnecessary detailed questions (e.g. year of birth rather than year of birth).
- Avoid use of open questions into sensitive areas (e.g. if it's important to know about anti-depressant use in your participant group, ask about that instead of making general inquiries into a subjects' medical background).
- You can only ask questions relating to particularly sensitive topics (i.e. health status) if you can argue that asking these questions is *critical* for your project.

- fMRI research: Use of the BIDS (http://bids.neuroimaging.io/) file structure comes highly recommended. Researchers planning to make MRI raw data sets publicly available should use tools to prevent identification of subjects on the basis of facial features (i.e. mri_deface, pydeface (https://openfmri.org/de-identification/)).
- Surveys:
    - if you ask for personal data in your survey, anonymize your data as soon as possible
    - delete contact information as soon as possible
    - only share anonymized files with colleagues. Only share with colleagues directly involved in the project.
    - delete your data from the survey system.

Appendix B - Digital informed consent forms

The following preconditions must be applied to on screen informed consent forms

- The participant must have access to a (virtual) keyboard or pointing device during the procedure (in order to *actively* consent).
- The informed consent text must be presented on a single page (scrollable if necessary), or on multiple pages with the ability to flip back and forth between pages.
- The presentation duration must be infinite or until the participant responds.
- The participant must be prompted to respond with one of two buttons; a decline and an accept button (mapped to keys or using a mouse/touchscreen interface).
- The participant must confirm his choice or be able to press cancel to return to the previous screen.
- [**only applies to digital informed consent in the lab**] If the subject clicks or presses accept and confirms, the subject should be given the opportunity to enter his first and last name. When this information has been entered, the experiment can start.
- [**only applies to informed consent for projects that are exclusively on-line**] If the subject clicks or presses accept and confirms the experiment can start.
- If the participant clicks or presses decline and confirms, a screen should appear informing the participants that they have declined the informed consent and that they should call the experimenter. This will terminate the experiment. If the experiment is online, a screen should appear explaining the subject that the experiment has been terminated and why.
- [**only applies to digital informed consent in the lab**] The following output should exist as a separate text file or record for each participant accepting informed consent.
    - The full text of the informed consent as presented to the subject
    - The name of the participant

Appendix C - Encryption

Research data that contains direct identifiers should be placed on UvA servers or be encrypted in storage. This also applies to sensitive data where the risk of inferential identification is non-negligible. Personal/sensitive data that is collected on a desktop or laptop harddrive should be encrypted and placed on a UvA file server after each experimental session, or after each full day of data collection if the device is not transported or left unattended during the day. Note that

encryption for data placed on UvA servers is not mandatory except for data with extreme leak-impact.

We are looking at various software solutions that combine features related to ease-of-use (folder encryption, encrypted drive creation), ability to collaborate without sharing personal keys, multi-platform use, and possibility for the research institute to retain a master key (in case of key-loss, and for scientific integrity considerations). This appendix will be updated as soon as new encryption methods become available.

Meanwhile researchers can use encryption features of programs like 7-zip (folder-encryption), MS Office (document-encryption), or Veracrypt (drive-encryption). Since these programs lack collaboration options and in keeping with the research institute's four-eyes policy (scientific integrity protocol, 2015) encryption keys should be shared with data stewards when papers based on the encrypted data are published.