Data management protocol, research institute Psychology

last updated: Nov 2022

Index

Background

This is the research data management protocol of the Psychology Research Institute. This document aims to provide practical information about all matters concerning research data, such as storage, publication and archival, and about RDM-related infrastructure at the research institute.

The Faculty of Social and Behavioral Sciences (*FMG*) of the University of Amsterdam, and therefore the Psychology Research Institute, is bound by The Netherlands Code of Conduct for Research Integrity[1] which requires researchers to practice good Research Data Management (RDM). Most importantly, good RDM is in the interest of the researchers themselves.[2,3] Proper data storage and regular back-ups ensure that researchers always have quick access to their data, which warrants the quality of their scientific practice. Moreover, it stimulates collaboration with others, who will find it easier to understand and re-use data, making RDM cost- and time-efficient.

The sections of this document are organized alphabetically rather than thematically or chronologically. It is under continuous review; if you think something is missing, please get in touch with the data stewards Psychology – Tinka Beemsterboer, Jasper Wijnen and Klaas Seinhorst (room G2.23, datasteward-psy@uva.nl). Also feel free to approach us if you have any further RDM-related questions after reading this document and/or if you need RDM support.

[1] *Netherlands Code of Conduct for Research Integrity*. KNAW, NFU, NWO, TO2, VH, & VSNU. 2018.
[2] The Turing Way: A Handbook for Reproducible Data Science, version 0.0.4. The Turing Way Community, March 2019: http://doi.org/10.5281/zenodo.3233986
[3] If researchers abide by the faculty's RDM guidelines, the university will always take full responsibility in case of (legal) dispute on RDM issues.

Data archival

When a paper has been accepted for publication, all data and other files associated with the paper should be archived within one month after publication, that is, placed in the published data storage folder on Research Drive via the data stewards. The responsibility for archival lies with the research institute with which the first author is affiliated; the first author takes care of the archival procedure. The archived data should be stored for at least 10 years after the publication date. An audit will check this archive regularly. RMS (Research Management Services) provides the researcher with the necessary information to create such a data package for archival. For projects that were requested in the *old* ethics portal, please use the protocol in this document for archiving data packages. Please note that while all data needs to be archived, not all data can necessarily be published open-access: it is not allowed to publish personal data (see "Data publication" below).

If you

Data exit meetings

If you leave the UvA, the research institute still needs to be able to access to your research data, for reasons of scientific transparency and integrity: the law requires that research data are stored for at least ten years, and this obligation rests with the research institute. Because PhD projects are inherently short-term, the data stewards invite all graduating PhD students for a so-called "data exit meeting" in which they assess how to comply with the GDPR and the FMG guidelines. All other staff members who leave the UvA are encouraged to reach out to the data stewards to make an appointment for a data exit meeting as well. If a researcher wants to take their data with them when they move to a different research institute, they should contact the data stewards (datasteward-psy@uva.nl) to check if this is possible. In most cases a Data Transfer Agreement can be drafted, to be signed by the researcher and the research director.

Data publication

If your data set is suitable for publication and if you want to make your data (openly) available to the general public, you can consider publishing it in a public data repository. When doing so, make sure your data set is FAIR (**f**indable, **a**ccessible, **i**nteroperable, **r**eusable), citable and compliant with funder requirements.

Important: the principles of Open Science are not necessarily compatible with the GDPR (General Data Protection Regulation; in Dutch Algemene Verordening Gegevensbescherming or AVG). The GDPR prohibits the publication of personal data, so make sure to anonymize and/or pseudonymize your data. If your data contains personal information that cannot be anonymized and/or pseudonymized before publication, you are still recommended to publish only a description of the data.

In choosing a data repository, pay attention to the following requirements to publish your data FAIR:

- A persistent identifier is given, creating a permanent link that points to the data, making your data findable and citable;
- A license (or contract with use conditions) is given or can be chosen, creating clarity and certainty about user conditions for potential users of your data. The UvA recommends publishing data under a CC0 (public domain: not strictly a license, but a waiver of rights, allows you to give up your copyright and put your work into the worldwide public domain) or CC-BY (others are allowed to use your work in any medium or format, as long as they recognize you as the original author of the work) license. For a list of the different licenses go here;
- A description is given of the data, by which it is clear when, how, and what was collected. This way, data are usable for others (and your future self);
- Metadata is provided; to ensure that your dataset appears as a search result, a dataset must be provided with data that can be read by search engines.

FAIR repositories

You are free to choose the repository of your own preference as long as it meets the requirements mentioned above. Examples of such repositories are:

- UvA/HvA Figshare
- OSF
- DANS EASY

Data storage

Temporary storage of data, for collecting and analyzing data, is done at a location that is exclusively accessible to the researcher(s) (e.g., Teams, OneDrive, ResearchDrive, a lab PC or UvA laptop). For integrity purposes, researchers must store their data on one of the storage solutions below, so not only on their laptop.

The UvA offers three cloud storage solutions for workspace storage; additionally, there is also the possibility to use an on-premise server for video and MRI data.

1. Microsoft OneDrive: offers personal storage. In order to use Microsoft apps (including Teams) on their computer, the researcher needs to install O365, which is available for every UvA employee.
2. Microsoft Teams/SharePoint: offers group storage. Collaboration is especially easy when working with Microsoft office documents.
3. Research Drive: offers group storage. To request a Research Drive account, click here.
4. FMG storage: in addition to the cloud storage, the faculty also has its own storage, the so-called “FMG storage”. FMG storage is intended for large data sets whose analysis requires fast connectivity. Access can be requested via your data steward. Please note: the current server will be replaced by a new server in the coming months; on this new server, only video and MRI data can be stored.

Table 1: Overview data storage facilities

		Storage type	Location	Speed[4]	Sharing	Security
Microsoft OneDrive	Personal file storage	Personal storage	Cloud	Medium	Internal and external	2FA
Microsoft Teams/ SharePoint[5]	Simple (student) projects	Project storage	Cloud	Medium	Internal and external	2FA
Research Drive	Projects with different access rights	Project storage	Cloud	Medium	Internal and external	2FA
FMG storage: research folder		Personal storage, project storage	On-premise	Fast	Internal	UvAnetID

For projects that require extra security, researchers can use a so-called "virtual research environment" (VRE), an online secured system. Features usually include document hosting and some discipline-specific tools, such as data analysis, visualization, or simulation management. A VRE is well suited to share sensitive data with others (e.g. students), as it prevents users from downloading files.

Data welcome meetings

The data stewards and the Behavioral Sciences Lab (BSL) organize regular meetings to introduce new researchers to the facilities at the research institute: this includes a tour of the labs and an introduction to the GDPR.

Encryption

Encryption entails the conversion of information into a format that can only be accessed with a so-called "key"; without this key, the information is meaningless. This restriction is particularly useful when working with sensitive personal data. Encryption must be used when processing the following types of research data:

- Use of special-category personal data, that is, data concerning:
  - racial or ethnic origin,
  - political opinions,
  - religion or belief,
  - membership of a trade union,
  - genetic or biometric data for the purpose of unique identification,
  - medical data (including mental health),
  - sexual life,
  - criminal record.

Encryption for such data is mandatory until the data have been anonymized (if that is at all possible).

[4] Storage solutions with fast speed are recommended when working with very large data sets (>10GB)

[5] Part of MS-Teams, available since June 2020, comes in place of P-drive, P-drive will be discontinued in summer 2021

- Videos / photos with participants recognizable on screen
- Sound clips of participants who are recognizable (e.g., by voice or video)
- Contact details of participants. Collect these separately from the rest of the research data, or separate them as soon as possible by replacing the identifying information with pseudonyms or subject-IDs.
    - If it is necessary to keep the identifying information over a longer period, place the contact info and pseudonyms in an encrypted key file.
- Anonymized or pseudonymized research data where the risks of indirect identification are non-negligible.
- Information that participants have shared with the researcher, where the disclosure of that information may place the participant at risk e.g., interview data in authoritarian regimes

If your data belongs to one of the above categories, we classify the data as sensitive, and it should be encrypted. To do so, you can use the program VeraCrypt. Please share your password with your supervisor. We are working on a service to share passwords with the faculty information security officer (fiso-rec@uva.nl), which we expect to be ready before the end of 2022. Furthermore we expect a roll-out of more advanced encryption tools that do not require researchers to handle passwords themselves by the end of 2023.

Research Management Services

Whenever a researcher starts a new project, they register it within Research Management Services (RMS). RMS also allows researchers to take care of all corresponding administrative procedures, and request support such as guidance and approval by the Ethics Review Board and/or the data stewards. All collaborators of the requested project, both within and outside of the UvA, should get access to the project in RMS. External researchers can request an account using eduID.

For Psychology researchers, the following procedures are available in RMS:

Before data collection:

- Ethics Review: specifies the treatment of participants in the research project; evaluated by the ethics member of the programme group, or in the plenary meeting of the Ethics Review Board.
- Data Management Plan: specifies RDM aspects of the research project; evaluated by one of the data stewards.
- Informed Consent: includes templates for different situations to help the researcher create a GDPR-compliant Informed Consent form; evaluated by both the Ethics Review Board as the data stewards.
- Privacy Agreement (if necessary): arranges a legal agreement if personal data are shared outside the UvA; guided by the data steward.

After data collection:

- Data Archiving: a guide to data archival in the closed archive of the Psychology Research Institute.

Sharing data and collaborating with external parties

Collaborating with third parties or sharing data with parties sometimes means that, according to the GDPR, the UvA has to make sure that the rights of the data subjects are protected. The following situations are examples of collaborations:

- A UvA researcher sets up and executes a research project with a researcher from Leiden University.
- A UvA researcher wants to consult a colleague from another research institute when it comes to data analyses and therefore sends a data file to that other researcher.
- For a UvA research project, the researchers want to use a mobile app to gather research data.
- The UvA researcher wants to send (part of) their research data to a database.
- **Note:** In case a researcher obtains data from or via a third party, it is important to inquire about their security measures.

The researcher should indicate such collaborations when they register their project in RMS; the data stewards will provide support if any legal agreements are required.

Students as experimenters

Students often work with personal data, either as student-assistants or for courses (including their thesis).

Students conducting their own research

Students who conduct their own research as part of a course or their thesis are subjected to many of the same rules as UvA staff:

1. Anyone who uses UvA ICT facilities is required to adhere to its security regulations, including the encryption of personal data: an overview of these regulations can be found on the UvA student website.
2. Students need to apply for approval from the Ethics Review Board, aided by their lecturer or supervisor.
3. Students must comply with the GDPR/AVG.
4. Supervisors must have access to the same data storage solution as the students working on the project. In case of integrity issues, the supervisor must make these files accessible.

Students as research assistants

Student assistants carry out research for UvA researchers: the researcher, rather than the student, determines how and where the research takes place. Research assistants must use UvA equipment, and additional administrative steps should be taken if students process personal data. For any administrative and/or practical support, researchers can reach out to the data stewards.

General guidance

Protected laptops. See Computer device requirements. Projects where **especially sensitive categories of data** are processed (e.g., identifying data combined with ethnicity, religion, political opinions, (mental) health, sex-life) must be performed on secured UvA laptops or a Virtual Research Environment. Personal

laptops of students should not be used for these purposes, as it cannot be checked whether they meet all security requirements. The Behavioral Science Lab has a limited number of laptops available that can be borrowed for data collection and/or analysis. Students/researchers can reach out to and request a laptop as needed.

Non-disclosure agreements (NDAs).

To ensure appropriate protection of personal data, contractual arrangements must be made when personal data are accessed by student assistants at UvA (or at other universities) as part of the data collection and/or analysis. NDAs must be signed by all student assistants processing personal data. Contact the data stewards to arrange such an agreement.

The GDPR (AVG) and how to secure your data

In the Psychology Research Institute almost all researchers work with data from human subjects. The processing of personal data is subject to the General Data Protection Regulation (GDPR; in Dutch "Algemene Verordening Persoonsgegevens" or AVG). The FAQ GDPR, created by the legal team of the UvA, provides a lot of information when it comes to research and the GDPR. Researchers can also consult the data stewards about these issues.

Working with personal data of participants requires researchers to take certain security measurements when storing their data.

Personal data

The GDPR defines personal data as any information that relates to an identified or identifiable living individual. The law aims to safeguard individuals' privacy, which has implications for researchers working with personal data. The safest security measurement when working with personal data is anonymization. If this is not possible, pseudonymization and encryption are good alternatives.

Anonymization

When directly identifying information is permanently removed, we speak of anonymization. For a dataset to be anonymous it shouldn't be possible for researchers in the project, or anyone else, to re-identify subjects by any means at their disposal. Truly anonymous data are not considered personal data, and the GDPR does not apply to it (provided that the risk of indirect re-identification is negligible, see the section below).

Pseudonymization

Pseudonymization means replacing directly identifying information with pseudonyms (e.g., subject IDs) and storing the link between the identifying data and the pseudonym elsewhere (i.e., in an encrypted key file). Pseudonymization is used instead of anonymization when the identifying data are needed at a later stage (e.g., for longitudinal research). Most data sets can be pseudonymized in this way, however in some cases directly identifying data is a core part of the research data (e.g., studies using video data) and pseudonymization is not feasible. Data sets that cannot be pseudonymized always require extra security measures, such as encryption.

Indirect re-identification

Both anonymization and pseudonymization require evaluating whether indirect identifiers (a combination of variables that can be used to form a profile of an individual) in the dataset allow re-identification of a data-subject and taking measures to prevent that. If risks of re-identification remain, data cannot be said to be truly anonymous and the GDPR remains applicable to the dataset. The extent to which pseudonymization efforts allow for sharing of data or unencrypted storage is dependent on whether indirect identifiers remain present. Risks of re-identification depend in large part on the characteristics of your sample. For example, if you ask your department's student pool about their gender and age, although seemingly general and anonymous, the one 54-year-old male student could still easily be identified. In contrast, for a nationwide sample, these variables cannot lead to identification.

The key steps to secure research data storage

Computer device requirements

Laptops used for research need to meet the following conditions:

- Laptops are encrypted (Bitlocker or FileVault is on)
- Laptops are password-protected at start-up and from sleep/standby mode
- Laptops automatically log off after a period of inactivity of up to 15 minutes
- Firewall is enabled, anti-malware is enabled (i.e., MS Defender)
- System software (e.g., MS-Windows, Linux or Mac OSX) is up-to-date.

Managed UvA laptops meet these requirements. Self-support UvA laptops do not automatically meet these requirements, employees must set this up themselves. The above measures also apply when using your own/ other non-UvA laptops. More information for students and employees.

In general, for all research data:

- Secure your laptop
- Use UvA infrastructure for storage (e.g. Teams, Research Drive, fmgstorage)
- Anonymize the data (see "The GDPR (AVG) and how to secure your data"); if anonymization is not possible, pseudonymize the data. Encrypt the data in storage if anonymization is not possible
- Restrict access to the data to researchers who are directly involved
- If pseudonymization applies, encrypt the keyfile with a unique password, preferably only accessible by the PI and FISO.

Specific for different types of data

Paper and pencil:

- Paper and pencil should be stored in a closet locked with a key.
- After data collection is finished, ask the *Documentaire Informatievoorziening (DIV)* to archive the data by sending an email to archiefbeheer-div@uva.nl.

Screening data

- Do not keep the data longer than needed

- Store separate from research data

Physiological data: EEG, MEG, ECG, EOG, GSR, EMG, Eye tracking

- These types of physiological data are not considered identifying by themselves. However, when linked to other identifying data they are considered (sensitive) personal data.

Physiological data: MRI

- The raw anatomical data are not anonymous and need to be stored with encryption.
- The anatomical scans must be defaced when shared or published.

Audio and/or video data, images

- When data are especially sensitive and multiple people need to work on the data, analyze your video or audio data in a virtual research environment or use the on-premise server fmgstorage
- If possible, transcribe your data as soon as possible, and encrypt the original data.
- Make sure to carefully pseudonymize the transcripts (if possible) removing any references that may help identify individuals.
- Archive the videos and audio-files with encryption.

Appendix A: How to prepare your data package for archival

After a paper has been accepted for publication, data and other files associated with the research must be placed in a data package. This data package will be stored in the published data section of the Psychology archive. This has to be done one month after publication at the latest. An audit will check this regularly.

Before archiving your data package, you need to prepare this in your own workspace (e.g., Research Drive or MS OneDrive or Sharepoint). Next, you contact your data steward who will archive the data package at a secure location (GDPR compliant). This archive is closed and only accessible in emergencies, after permission from the Dean, via the Data Steward.

Important: make sure that your data package has the following name structure: surname_RMSProjectID (e.g., Bakker_FMG-123) or surname_LABERBnumber (e.g. Bakker_2021-BC-12345)

The data package contains the following subfolders:

1. **Ethics**:
   - Ethics project number from the *old* ethics portal or RMS (in the latter case use the number from the project overview)
   - Ethics protocol and approval
   - Information brochure
   - informed consent form
   - other relevant files like a debriefing procedure
2. **Setup** (Method, measures, materials)

- The experiment scripts (e.g. the Presentation or PsychoPy task code) and stimuli, accompanied by instructions and software dependencies and/or hardware requirements that have to be met to run the task.
- PDF of paper-pencil questionnaires.
- If online surveys are used, a (PDF) download of the survey.
- Description of the exact version and platform if software has been used.
- When custom software is used (written for the current experiment), upload a zip file with the software and other relevant specifications for use.

3. Data collection
   - Raw data files without identifiable information of the participants. If your data contains identifiable information of participants, only keep the identifiable data if it is part of your research data. Contact information has to be permanently destroyed right after data acquisition.
   - Configuration parameters (e.g., EEG configuration files, VSRRP driver files).
   - Lab log with entries identified by date and experimenter; subjects identified by participant number, any subject related materials or comments.
4. Data analyses
   - 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 *code book*: description of all variable names and labels with sufficient detail to understand both the raw and processed data
   - A list of dropped subjects plus reason for exclusion
   - The resulting transformed data that formed the basis for statistical analyses in the published paper.
5. Papers and report
   - The final manuscript or DOI of the article.
   - If applicable also the DOI of the data set.
6. Declaration of the retention period·
   - If a data set contains personal information, then the data package should also include a statement indicating when the original data set should be removed from the archive. Declaring a retention period is required by law and should conform to any agreements made with research participants or providers.