Digital Data Storage Protocol - concept
Updated March 2019

Introduction

This protocol describes the steps that researchers and institutes should take to ensure suitable and adequate storage of *digital data* and supporting material.[1] As such, it is an appendix to the University-wide regulations for Data Management ([available from the website](#)).

After your research has been completed, the research data must be preserved for the long term together with the metadata, software and other documentation required for reuse. The aim is to store all data from any kind of research projects for at least 10 years after the publication of the results of the project concerned.

Data storage in an archiving system *during* the project is not mandatory but nevertheless highly advisable.

The focus of this protocol is on storage of data, not on sharing it, although sharing of data is encouraged by the Faculty of Humanities (e.g. through Open Access channels. As a default, the university has an arrangement with Dataverse; see section 5 below).

The goal of this document is to answer the following questions:
1. Why? Why do we need to store collected and analyzed data?
2. What? Which types of digital data should be stored?
3. Who? Who is responsible for appropriate storage? Who can monitor the data storage?
4. When? When should you store your data?
5. Where? Where can I store my data?
6. Whom? Whom can you ask for help?

1. Why? Why do we need to store collected and analysed data?

In recent years, many stakeholders within the global research community and organisations like NWO have initiated actions to develop [open science](#). Introduction of a more professional level of research data management is expected to make the research more accountable and verifiable, and to make the research data FAIR: 'Findable, Accessible, Interoperable and Re-usable'.

Systematic storage of research data has various benefits for the researcher. It ensures that the data are secure and available several years after the research is finished. Furthermore, some funders and journals also require that original datasets should be made open wherever viable (given constraints of privacy and security; see GDPR in section 2 below). The reviewers for a journal may want to double-check the accuracy of the authors' analyses and/or carry out further analyses. In a long-term

---

[1] Note that this is a guideline. *Everyone* should adhere to proper data management, but some researchers may have obligations (enforced by external grant organizations) to archive abroad, for instance. Of course that is all right, as long as you store and archive it does not matter where.

perspective, open science practices may become part of the evaluation of the research quality (see, for instance, the guidance in [the Leiden Manifesto)](.).

2. What: which (types of) digital data should be stored?

Digital data covers *every piece of digital data* collected during your research. Examples are:
- text files (pdf, html, word, etc.)
- spreadsheets
- statistical files: both the original code and the analysis
- databases
- (field) notes
- image materials
- audio-visual files
- cartographical data
- spectrograms
- fMRI data

It is important to be aware that not all data should be stored. For instance, the storage costs should be taken into account. Another important aspect of data storage is that the stored data comply with GDPR/AVG.

- The GDPR concerns: Data on the basis of which a living individual can specifically be identified

It does not concern:

- Information about living individuals that one may find in sources that have already been published

To comply with the GDPR in your Data Management practice, you need to adhere to only a few simple basic rules.[2] If you use personal data in your research:

- make sure you have the proper legal base for it
- contact Privacy Support to register your research
- collect *only* the personal data that you really need for your research, *nothing more*
- be aware of Internet Security issues

These rules form the basis for a few practical guidelines that we list below:

- If you plan to work with personal data, use consent forms. Subjects can withdraw their consent, but *not after* you have anonymized your data

---

[2] AVG compliance entails more than the points mentioned here. We list those aspects that we consider to be paramount for compliance with respect to data Management. If you have any remaining questions you can turn to Privacy Support.

- If you do not plan a follow-up study with the same subjects, anonymize as soon as possible
- Do not keep the personal data longer than you need to; if you have anonymized, then you can keep the data for as long as you like.
- If you do plan a follow-up study, pseudonymize the data and keep the key file separate
- Whenever in doubt about keeping sensitive personal data, encrypt the dataset (with AX Crypt)
- Raw data should be stored on University Servers, never outside, unless officially licensed by the University.
- If you need to store data locally, use data encryption.
- Do not share raw data with collaborators outside the university; do not use cloud services (apart from SURFDrive) to share data.
- Do not email datasets, use SURFDrive or SURF Filesender instead.
- If you need assistance with Internet Security issues, or have questions about the GDPR in relation to your research, please contact Privacy Support.

3. Who: who is responsible for appropriate storage? Who can monitor the data storage?

Researchers are responsible for storing their own data, both during data collection and at the end of the project. The storage of data should be an integral part of the project planning. It is important that the researcher responsible for a research project keeps track of all responsibilities with respect to data management.

During the planning phase, the researcher should already decide on which metadata are needed for the research (see section 5).

For studies conducted by students or PhDs, the responsibility for the storage of data lies with the supervisor. The storage of data should be discussed at least annually during the ROGs (Performance and Development Interviews).

Apart from the lead researcher of a project team and/or supervisor of a PhD student, proper data storage can be monitored by the head of the research group (sectievoorzitter) to which the researcher belongs, the Academic Director, and the Lab Manager or local data steward. The director of the Institute should also have access to all data files *uploaded to Dataverse.*
Academic Directors of the institutes are responsible for dissemination and enforcement of the Data Management and Storage protocol in their institutes.

4. When: when should you store your data?

A distinction should be made between storage of data *during the project* (*data collection*) and storage of data *at the end of a project (Archiving)*. We distinguish these two phases, A and B, in a project, and note that phase A can apply cyclically during any project:

A. *Collection*: Materials that become available during or immediately after completion of a project phase (experiment, study, data gathering period, fieldwork period, etc.). Each cycle of phase A can be followed by analysis of data and publication of articles based on these data.

B. *Archiving*: Materials and information that become available at the end of the project, after final analyses and publications. Publishing of the whole dataset in an archiving system.

Data storage in phase B, at the project, is mandatory. To comply with the regulations (for example, store the original, raw data file) you should already start collecting and storing the available information immediately after completion of each data collection phase (phase A). This timely collecting and storing is important because there might be quite some time between conducting the research and the end of the project or producing a publication. Also, you may not always be certain if (and when) a completed phase will result in a publication. This is why we recommend a procedure (see section 5) to safeguard that you will indeed comply with the protocol when publishing about your research.

5. Where: what storage archives are there and what should be included as metadata?

Part A: archive building during data collection

Phase A, Collection Materials become available during or immediately after completion of a project phase (experiment, study, data gathering period, fieldwork period, etc.).

During the collection phase, data can be stored either on the university network drives, SurfDrive, or any other such storage platform. In any case, be sure to save regularly and make backups. Immediately **after** completion of a data collection phase, it is advisable to transfer the electronic data to an online archive directly. As noted above, this will save you the trouble of having to do *everything* in phase B. The Humanities faculty provides a DANS Dataverse account https://dataverse.nl/. Each institute will have their own portal on this platform, with individual storage space for all staff members. For access rights, please contact your local labmanager /data steward. The researcher should make a preliminary estimation of the required storage and inform the Lab Manager / data steward about it. If you publish an article based on a phase A dataset, it is even more important, and sometimes even compulsory, to move your data to an archive like Dataverse.

Alternatively, institutes offer mobile storage for researchers in the field without access to cloud solutions. For available alternatives, IFZ should be contacted. The Faculty strongly advises against the use of (personal) external hard drives for data storage. Whenever you can, transfer every dataset you collect to network drives (with guaranteed backups) as soon as possible. If you have to use local storage, for whatever reason, the rule should be to transfer the data to a cloud solution (e.g., Surfdrive) as soon as a connection to internet is available to you. For audio-visual data, a useful platform could be Databrary https://nyu.databrary.org/.

If you have to use local storage, make **backups** on Solid State Drives (especially when doing fieldwork without access to internet).

If you transfer a dataset to an archive after phase A. The following information should be included in the storage archive, if applicable:

(1)
- original raw data file(s): experimental results/measurements, scans, AV-recordings etc.;
- Ethics declaration form. Once the Ethics Committee is formed, the researchers should include the Ethics code/number and the approval of the Ethics Committee

4

- informed consent [forms](#)
- a copy of the experimental tools and stimulus materials used, if applicable (e.g. the E-prime program, "praat"-scripts, description of equipment used in experimental set-up, audio or video stimuli, database, or in case of a paper and pencil study in pdf; for online studies the materials in pdf). If you want researchers who may use your scripts to cite or acknowledge you as an author, you can include "conditions of use" text with the scripts.
- the lab log (if available)
- Metadata (part A). This includes:
    1. Information on the sample (e.g. number and type of participants)
    2. Dates of data collection
    3. The names of all researchers involved (including the names of students, if applicable)
    4. The names of all assistants involved (for example, if someone transcribed or annotated the primary data)
    5. (information about) the period during which the study was conducted
    6. (information about) the organizations/schools and contacts involved
    7. the studies with which the current study was combined
    8. which lab was used
    9. location of fieldwork
    10. source library for documents (to trace back the documents used)

Phase B: Archiving

Materials and information that become available after the final analyses and publication(s).

At the **end of the project**, the researcher will *publish* the original data files and all relevant additional materials in an archive. This archive may be the local Dataverse, offered by the Faculty of Humanities (see phase A for details), national (like DANS: [Digital Archiving and Networking Services](#)) or international, depending on needs, project requirements, international agreements etc. The use of DANS is being supported at the institute and faculty level. If you have not archived datasets after phases A, this phase is where you add all the metadata and other relevant materials to your datasets in an organized way. If you have used Dataverse for archiving *during* the project, but do not wish to (or cannot) use it for final archiving, contact your lab manager / data steward who will be able to assist in migration of the data (especially a move from Dataverse to DANS can be executed without any problems).

The archived version of the dataset should contain all relevant information for all data gathered and used during the research. This automatically covers data archived (for publication purposes) in phase(s) A. The set includes all data from the collection phase (A) of the project, as well as the archiving phase (B).

Next to the information listed in (1) the following information should be included in the storage archive:

(2)
- the processed data files used for publications, if applicable
- syntax files and/or any custom-made (e.g. Text- or visual-analysis/Praat/R) scripts/software, annotated PRAAT/ELAN/toolbox files used for the analyses

- statistical output as reported in the manuscripts (as pdf)
- Metadata (part B). This includes:
    1. Numbers of excluded participants
    2. Inclusion/exclusion criteria
    3. Publication details
    4. PDF of the published manuscript (but check if the journal has copyright)

Once you have collected all information mentioned under A and B, ensure that the combined package is stored safely in one of the archives mentioned above. The [Lab Manager/ the faculty GDPR-team /LUCDH/CDS] can assist you in this.

Note: The archived data and accompanying metadata will be "frozen" immediately after archiving in phase B (at Dataverse, for instance); no changes can be made and no materials can be added after this point. Therefore, it is important to check before archiving that all the necessary materials are included. When archiving in Dataverse, the Lab Manager, Data Steward or the Centre for Digital Scholarship will act as administrators to make sure the necessary materials are included and that materials violating the GDPR or copyright are not included.

Whom can you ask for help?

For help with issues on data management, please look at the website of the Centre for Digital Scholarship https://www.bibliotheek.universiteitleiden.nl/onderzoek-en-publiceren/centre-for-digital-scholarship. If you cannot find what you are looking for there, please **contact your Lab Manager / data steward.**