

Data management protocol Institute for Criminal Law and Criminology

December 2018

This document describes the steps that researchers should take to ensure storage of data¹ and supporting material. The goal is to ensure that our way to store data is in line with the regulations of the University and the protocol of the Faculty of Law. Point of departure is that data and supporting materials will remain available for all publications for at least 10 years after publication. The focus of the protocol is on storage, not on sharing data. The Faculty of Law does encourage its staff members to share data (see chapter *Public repository*, below).

In this protocol, the focus is on the storage of data from publications, including dissertations. The storage of data + supplementary materials is mandatory for all publications². However, this protocol is only applicable for scientific publications.³ The protocol is only applicable for members of the Institute for Criminal Law and Criminology. Research committed by students for the purpose of writing a bachelor- or master thesis, is excluded from this protocol.

A distinction should be made between different research stages: 1) before the research, 2) during the research, 3) after the research project is completed, and 4) after publication. The latter distinction is necessary because not all research culminates in a publication. However, it is then also important that the research data is preserved and is findable. Externally funded research projects will generally formally conclude upon their financial-administrative completion.

Before the research

A researcher or a research group must draw up a data management plan (DMP) before data collection for a research project⁴ begins. The DMP elaborates upon the data management policy of the faculty/institute for the specific research project in question. It describes what data will be collected and how the project will store the data during the research and after the research project has been completed.

A DMP is a written document that describes the data that is expected to be acquired or generated during the course of a research project, how it will be managed, described, analyzed, and stored, and what mechanisms will be used at the end of your project to share and preserve the data. Until another storage tool is selected, the DMP should be stored on the J-disk of the Faculty of Law/Leiden University, and should be available for 20 years.

¹ Please see the section on 'definitions' (page 6) for a description of how data are conceived of in this protocol.

² Please see the section on 'definitions' (page 6) for how publications are defined of in this protocol.

³ Please see the section on 'definitions' (page 6) for how scientific publications are defined in this protocol.

⁴ Please see the section on 'definitions' (page 6) for how research project is defined in this protocol.

The DMP for a research project must record which data-management responsibilities are assigned to the various members of staff who are working on the project and which responsibilities have been assigned to whom. Projects in which researchers from multiple institutes and/or external parties are involved require special attention. Every researcher should store a copy of the DMP at his or her institute (J-disk). It is also important to record what will happen to the data and the responsibility for it if one of the researchers leaves.

The following aspects at least must be covered in the DMP of a research project:

- How data-management responsibilities are assigned within the project and what will happen if the researcher, or one of the researchers, leaves;
- The types of data that will be generated and collected;
- The collection method(s) or origin of the data;
- The chosen metadata⁵ standard for the documentation of the data ;
- Where the data will be preserved during the research and how security and access will be arranged. The measures that will be taken for the long-time preservation of and access to the data;
- Who will have access to the data at which point;
- How sensitive or otherwise confidential data is dealt with;
- If needed, the DMP should be updated during the research process.

We propose that from 1 January 2019 onwards, for every new research project, a DMP will be drawn up.

The preferred standard format for a DMP can be found at the J-disk in the folder on data management. Further information on how to store information on the J-disk and a data management cost guide and costing tool will be placed in that folder as well. A list of courses on data management that can be followed will also be provided.

During the research

- There are no mandatory guidelines for how to preserve data during the research process. However, all researchers are required to follow the University's academic integrity regulations.⁶ This implies that during the research, research data must be securely preserved and – if required – the confidentiality of the data must be guaranteed. Please note: once the 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 (see below). We therefore strongly advise researchers that they already start storing the data and preparing the metadata documentation during the research project. It will, for instance, generally be difficult to provide satisfactory metadata once the research has been completed as it can be

⁵ Please see the section on 'definitions' for explanation of what metadata entail.

⁶ Please see <https://www.universiteitleiden.nl/en/research/quality-and-integrity/academic-integrity>

hard to remember all the details. This is why we recommend a procedure that safeguards that you can indeed comply with the protocol when a study is finished and/or published. In addition, during the data collection phase, electronic data will preferably be stored at the J disk as soon as reasonably possible, instead of, for example, on a (personal) external hard disk or USB device.

After the research project is completed

Research data must be managed in such a way that it is preserved so that it is at least findable, accessible, comprehensible and reusable in the long term.

The minimum retention term for research data is ten years. In case of multiple publications arising from a single research project, this amounts to ten years starting from the data of the last publication.

Research data must be stored together with the metadata, other documentation and possibly the software and version of the software required for its potential reuse. Data and metadata should be stored at the J-disk. In addition, if possible, researchers are encouraged to store the data and metadata in a public repository (e.g. DANS archive).

Data that are publicly accessible (e.g. jurisprudence, literature, policy documents) do not have to be stored. The DMP and metadata should include a description of which documents were used and why (sampling methods), and instructions on how the data can be accessed. Researchers should bear in mind that the data also need to be publicly accessible within 10 years.

J-disk

Each unit has a separate destination folder. Within this folder, each employee or research project can be assigned a unique folder. All data of a research project, and all supplementary materials (including metadata) should be stored here. The folder with archived data and supplementary materials (including meta data) will be “frozen” immediately after archiving; no changes can be made and no materials added after this point. Therefore, it is important to check before archiving that all the to-be-stored materials are included. In case of an erratum, a new folder will be created. Archiving and freezing data and metadata does not imply that the data cannot be used for writing other publications.

Please find here a checklist of the material that should be stored:

- The original “read only” raw data file(s). These can be questionnaires, transcripts of interviews, observations, field notes, statistical material, audio- or visual material. The raw research data must be stored in such a way that it is independent of the underlying equipment/hardware, such as microscopes, scanners, or recording equipment. Long-term data formats that are supported by data archives should be used if possible. Retention of hardware can be considered in certain cases, for

instance in the case of software that is only compatible with an obsolete computer operating system. Close attention must be paid to the cost/benefit ratio here. Please consult the data manager.

- A document containing the metadata.
- Supplementary information (if not already included in document on metadata). This includes all the information necessary to understand the data. For example:
 - the sample (e.g., number and type of participants), sampling methods used, information on non-response;
 - the names of all researchers involved (including the names of students, if applicable) period during which the study was conducted, organizations and contacts involved studies with which the current study was combined.
- If applicable, the protocol and supporting materials for the ethics committee, including the code/number assigned by the ethics committee. The approval of the ethics committee, including the code/number.
- To safeguard confidential data (such as informed consent forms or original audio or video material), these data will be encrypted and stored on a separate location on the J-disk.

Public repository

We encourage metadata to be stored in a public repository when it concerns data that are potentially interesting to other researchers, for example a code book or questionnaires. The data must be findable for other researchers and involved parties. Its findability improves if it is deposited in a data archive or repository (e.g. DANS). The information about the data (metadata) is then registered in a standardized manner and is findable, also for search engines. Future findability is guaranteed by assigning a persistent identifier to the data. A commonly used identifier for data is DOI (digital object identifier). A DOI is assigned by a data archive or data publisher and is a unique number for a digital object, in this case a dataset. The DOI remains the same even if the location of the dataset (URL) changes. A DOI or other persistent identifier is used in citations or references to the dataset.

Depositing your metadata in a public repository does not necessarily mean that the data must be made fully available to other researchers. The data will have to be made accessible in some instances, for example for verification by a grant provider or journal, but will remain inaccessible to the wider public. With sensitive data, full publication will never be an option. Research data can be retained under embargo; then only those who have deposited the data have access to it. The duration of the embargo is determined in consultation with the data

archive. A further option is only to grant access if a request for access is approved by the researcher. The researcher then knows who is consulting the data and can reach agreements about its use and reuse. NB: if the decision is made to limit access to the data, it is essential that the metadata (description of the data) is findable.

After publication of research

Every academic/scholarly publication should at least have a DMP and metadata document. If no data have been used in the publication, this can be indicated in the DMP. In that case, a metadata document is not necessary.

Please indicate what kind of information is used in this publication:

- a. Literature that has been collected in a non-systematic way
- b. Case law that has been collected in a non-systematic way
- c. Empirical data such as information from interviews, questionnaires, experiments, databanks, observations
- d. Literature that has been collected in a systematic way
- e. Case law that has been collected in a systematic way

In case you use information from category c to e, further completion of this DMP is required.

If a publication has made use of data, apart from the DMP the metadata should be stored in such a way that it is findable, accessible, comprehensible and reusable.

If a publication uses data that are already stored in another place it is not necessary to duplicate them but to indicate where the data can be found. In addition, employees should include:

- the published manuscript;
- the datamanagement plan (DMP) in which it is indicated where the data can be found;
- the metadata document;
- a copy of the used materials needed for a replication (e.g., (SPSS) syntax,(Altas.ti) coding, E-prime script, or in case of a paper-and-pencil study a pdf, lab log).

Responsibilities

The Data management Protocol applies to all researchers at the Institute. Employees are responsible for storing their own data; both after completion of the study and after publication. For studies conducted by students or PhD students, the responsibility for the mandatory storage of data after publication lies with the supervisor. Researchers can call in

the assistance of a data manager who handles the storage of data for members of the institute.

The data manager will practically assist employees of the institute fulfil their data management responsibilities.

The data manager will check if new publications have a folder on the J-disk where all the data (if applicable) and necessary information is stored according to the protocol.

In the R&O form, a sentence will be included that states that the employee adheres to the institute's regulations concerning data management.

Apart from the data manager of the unit, the "Data Storage after Publication" can be monitored by the coordinators of the Criminal Justice research program.

Definitions

Data

Research data is understood to mean the following: all data that is gathered and generated during academic research. This also includes the data that is acquired by processing and analyzing (raw) research data. We follow NWO here, which states, 'we understand "data" to mean both gathered, unprocessed data and analyzed, generated data.' This includes jurisprudence, literature, interviews, questionnaires, observations, policy documents etc. This does not mean that all data must also be preserved. The decision concerning which data to preserve is recorded in this protocol. These regulations apply to all digital and non-digital research data. We follow NWO policy here. This stipulation does not mean that all non-digital data must be digitalized per se, but it does mean that the same requirements apply to non-digital data concerning findability, reusability, long-term preservation, etc.

Research project

A research project is defined here as a data collection for specific research purposes. Research projects can vary from very small scale (e.g. data collection connected to student's master thesis), to large scale (e.g. a Vidi project). A project can result in one publication, or several publications can stem from a project. A project has a specific starting and ending date.

Publication

A publication refers to all scholarly writings that are published and therefore accessible to other researchers. These include journal articles, book chapters, monographs, dissertations etc. These do not include professional publications. This data management protocol

applies to all publications that are entered in LUCRIS under the heading of scholarly publications.

Scholarly publications

Scholarly publications are those publications that increase the body of academic knowledge. They are the result of academic research and are aimed at the forum of peers. The publication shall fulfil the basic requirements of scholarly rigour generally accepted in the relevant field.

Metadata

The term metadata refers to descriptive information about data, which renders the contents of a dataset intelligible to other users. Metadata is often defined as data about data. It is information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage the data. Whereas there are many standards about for cataloging metadata, these are very discipline-specific. Please see https://en.wikipedia.org/wiki/Metadata_standard for more information, or take a look at the examples of metadata files for different types of research (quantitative and qualitative) on the J-disk.