This document describes the steps that researchers should take to ensure an adequate storage of data\textsuperscript{1} and supporting material. The goal is to ensure that our way to store data is in line with the broad regulations of Leiden University and the policy of Leiden Law School. 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. Nevertheless, Leiden Law School does encourage its staff members to share data (see chapter \textit{Public repository}, below).

The protocol focuses on the storage of data that are used in publications\textsuperscript{2}, which is mandatory. The protocol is relevant for all stages in a research project: 1) before the research, 2) during the research, 3) after the research project is completed, and 4) after publication. The distinction between stages (3) and (4) is necessary because not all research culminates in a publication. In these cases, it is also important that the research data is preserved and is findable. Externally funded research projects will generally formally conclude upon their financial-administrative completion.

\section*{Responsibilities}

The Data management Protocol applies to all researchers at the department. In principle, each individual researcher (each co-author) is equally responsible for storing his/her own data; both after completion of the research project and after publication. Within research projects, alternative divisions of responsibilities can be described in a data management plan (see below).

\section*{Before the research}

A researcher or a research group must draw up a data management plan (DMP) before the actual data collection for a research project\textsuperscript{3} begins. The DMP elaborates upon the data management policy of the faculty 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.

\textsuperscript{1} Please see the section on ‘definitions’ (page 6) for a description of how data are conceived of in this protocol.

\textsuperscript{2} Please see the section on ‘definitions’ (page 6) for how publications are defined of in this protocol.

\textsuperscript{3} Please see the section on ‘definitions’ (page 6) for how research project is defined in this protocol.
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 department, and should be available for 20 years.

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 project or of one of the institutes no longer participates in the project.

Summarizing, the following aspects at least must thus be covered in the DMP of a research project that is initiated:

- How data-management responsibilities are assigned to researchers 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 metadata\(^4\) 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 of the research project;
- How sensitive or otherwise confidential data is dealt with;
- If relevant, include a link to a document that describes how ethical issues are addressed.
- If needed, the DMP should be updated during the research process.

The preferred standard format for a DMP can be found at the J-disk in the folder on data management.

**During the research**

In principle, there are no mandatory guidelines on the specific way of preserving data during the research process. However, all researchers are required to follow Leiden University’s academic integrity regulations.\(^5\) This implies that during the research, research data must be securely preserved and – if required – the confidentiality of the data must be guaranteed. At this point, it should already be stressed that, upon completion of the project, the research

---

\(^4\) Please see the section on ‘definitions’ for explanation of what metadata entail.

\(^5\) Please see [https://www.universiteitleiden.nl/en/research/quality-and-integrity/academic-integrity](https://www.universiteitleiden.nl/en/research/quality-and-integrity/academic-integrity)
data must be preserved for the long term together with the metadata, software and other documentation required for re-use. 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 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:

During the data collection phase, the electronic data will preferably be stored at the J disk. The institute encourages the use of the network server instead of, for example, a (personal) external hard disk.

**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.

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 in 10 years.

**J-disk**

Within the folder on the J-disk of the Department of Economics, each 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.

Summarizing the above, we thus can formulate the following checklist of the materials that should be stored:
• The original “read only” raw data file(s). These can be statistical datasets, questionnaires, transcripts of interviews, observations, field notes, 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.

• Materials needed for processing the data and for a replication of the analysis (e.g., Stata do-file; SPSS syntax, Excel files, etc.).

• 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:
  o the sample (e.g., number and type of participants), sampling methods used, information on non-response;
  o the names of all researchers involved (including the names of students, if applicable) period during which the study was conducted, organizations and contacts involved.

• The protocol and supporting materials for the ethics committee, including the code/number assigned by the ethics committee (if applicable). The approval of the ethics committee, including the code/number.

Public repository
We strongly 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 immediately 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 submitted to 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 publication should at least have a DMP and metadata document. If no data has been used in the publication, this can be indicated in the DMP. In that case, a metadata document is not necessary.

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, researchers should include:

<table>
<thead>
<tr>
<th>Dataset managed by external organization</th>
<th>Before the research</th>
<th>During the research</th>
<th>After the research project is completed</th>
<th>After publication of research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datamanagement plan (DMP)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Original “read-only” raw data files</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Dataset</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Metadata document</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Materials needed for replication (e.g. do-file)</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Published manuscript</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Researchers are increasingly using datasets which are managed by an external organization (e.g. Statistics Netherlands (CBS)). Because of that, researchers are not able to store datasets themselves. This is not necessarily problematic, as these organizations might manage the data equally well or even better. Researchers are responsible for checking the data management policies of these external organizations. Furthermore, they should very carefully store the version number of the dataset. Finally, they should store a copy of the used materials needed for processing the data and for a replication of the analysis (e.g., Stata do-file; SPSS syntax, etc.) on the J-disk of the Department of Economics.
Definitions

Data
Research data is understood to mean the following: all data that are gathered and generated during academic research. This also includes the data that are 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. Also here we follow NWO policy. 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. This data management protocol applies to all publications that are entered in LUCRIS.

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.