

Instructions for archiving publication packages at the Institutes of Psychology and Education & Child Studies¹

Version 4, May 2023

What is a publication package?

A publication package is an anonymized/pseudonymized bundle of data, that is created with the [FAIR principles](#) in mind. This package contains the materials necessary to replicate the study, reproduce the results and reuse the data. We advise researchers to start creating the publication package *during* the project as it is much harder to do this afterwards.

For all published empirical studies and empirical studies appearing in (unpublished) PhD thesis chapters and research master theses, a publication package has to be stored on the platform DataverseNL for at least ten years. These guidelines are applicable to data that are collected and/or stored at Leiden University. The principle “apply or explain” applies here: please inform the data stewards if your project cannot be shared in an online repository.

What should a publication package contain?









(Not all elements will be applicable to each type of study. For instance, a systematic review may not have raw data or approval from the Ethics Committee.)

- 1. Manuscript or publication:** for publications either the Original Author Manuscript as submitted to the publisher (preprint) or the Author's Accepted Manuscript (postprint). You may use the published manuscript only if the publication was originally open access on the publisher's website.
- 2. The instructions, procedures, the design of the experiment and stimulus materials** (blank consent form, interview guide, questionnaires, surveys, tests) that are necessary for an exact replication of the research. Please take care not to include any personal signatures, e-mail addresses or phone numbers.
- 3. The (de-identified) raw data files.** Raw data files provide the most direct registration of the behaviour or reactions of participants (e.g., unfiltered export file of online surveys, raw time series for an EEG measurement, e-dat files for an E-Prime behaviour experiment, recordings or transcripts of interviews, descriptions of observations, archive and other source or media material). Please make sure that the data does not contain names, phone numbers, email addresses, etc. of participants. If the raw data files have been accessibly stored in an external archive (e.g., at DANS), making reference to the files in this archive will suffice.
- 4. Computer code and/or logbook** (e.g., SPSS syntax file, MATLAB analysis scripts, R code) describing all steps taken to de-identify the data to make it suitable for sharing, the steps taken to process and convert the raw data into the processed data and the steps taken to analyze the processed data into the results in the manuscript. Please give a brief explanation of the steps in English.
- 5. The processed data files** that were eventually analysed when preparing the article (e.g. an SPSS data file after transforming variables, after applying selections, etc.).
- 6. The data management plan** if this was created for the study.
- 7. A readme file** (metadata) describing how data files are organized and internally structured, as well as any other information that a researcher in a similar field would need to reproduce the published results or carry out new analyses based on the components of the publication package.
Please use this template to write the README (or use a different metadata standard like [Dublin Core](#)): <https://www.universiteitleiden.nl/binaries/content/assets/sociale-wetenschappen/psychologie/instituutsbureau/readme-template-for-publication-packages-v1.0.docx>
- 8. Documents related to the ethical approval** or a reference to such documents
- 9. Preregistration documents**, if they were created for the study.

¹ This instruction is based on the Guideline for the archiving of academic research for Faculties of Behavioural and Social Sciences in The Netherlands. Version March 2022.

Met opmerkingen [P(1)]: In this version:

- Preregistration documents were added as an element of the publication package
- It is no longer necessary to zip all folders separately
- More information on how to safely share sensitive data
- It is now possible to reserve a DOI on DataverseNL after an update

-  1. PUBLICATION
-  2. MATERIALS
-  3. RAW DATA FILES
-  4. COMPUTER CODE
-  5. PROCESSED DATA FILES
-  6. DMP
-  7. README FILE
-  8. ETHICAL APPROVAL

Example of a Publication Package folder

How to publish your publication package?

A publication package must be stored as soon as possible after the definitive publication of the manuscript. The data stewards are there to support the researchers in publishing their publication packages on DataverseNL. In order to publish a publication package, researchers are asked to:

1. Make a folder per component and number and name these folders, e.g. "3. Raw Data Files".
2. Move the publication package to J:\ResearchData\FSW\Archivering DataverseNL\ (choose folder of department). If the package is not meant to be openly available, encrypt it with a unique password. Encryption can be done via 7-ZIP (can be installed via the Software Center). Right-click the publication package folder, via 7-Zip, select 'Add to archive...'. Select '.zip' as 'Archive format' and enter a password. If you can't access this folder, you can also share your pub pack via OneDrive. Please share your folder with w.p.plomp@fsw.leidenuniv.nl and j.a.b.mink@fsw.leidenuniv.nl.
3. Fill in this [Formdesk questionnaire](#) with relevant metadata about the publication package.
4. Inform the Behavioural Sciences data stewards that a publication package is ready to be checked via datastewards_psy_ped@fsw.leidenuniv.nl. If the publication package is encrypted, don't forget to add the password.

The data stewards will check if your publication package is complete, and whether it doesn't contain too much sensitive or identifiable information. The data stewards will contact the researchers if they have any questions or concerns. When the publication package is uploaded to DataverseNL the researchers will receive a DOI (permanent link).

Reserve a DOI

A DOI (Digital Object Identifier) is a unique identifier that provides a permanent link. Sometimes journals or funders want a DOI to your publication package before they publish your article. The data stewards can reserve a DOI for you in this case. You do not need to have a publication package yet when you reserve the DOI; you can make this later.

Temporary link

It is also possible to have a publication package uploaded in draft, but not published yet. Sometimes funders want proof of open data sharing. This link expires when the publication package is published.

Maximum Size

DataverseNL has a maximum size of 9.3 GB per file, and up to 2000 files per dataset. If any of your files exceed these sizes, please inform the data stewards and they can help you select a different repository. This is often the case for studies with MRI data. For more information on sharing MRI data please consult the [Brain MRI data sharing guide](#) | [Zenodo](#).

Adding restrictions ('As open as possible, as closed as necessary')

Privacy is an essential component of Open Science. We should aim to be "as open as possible, as closed as necessary". If there is a valid reason, the researcher could choose between 'restricted/upon request' (not directly available, but request for access can be sent to the contact person via mail) or 'closed' (never available for anyone and is only added for archiving purposes). This choice can be made *per file*. For instance, raw data may be restricted while the processed data can be open. You can indicate which files you want to set as open/upon request/closed in the formdesk questionnaire (mentioned above).

Sensitive Data

In most cases data can be shared Open or Upon Request. If your data is sensitive, you could use the following strategies:

- Making categories less specific. Use age rather than date-of-birth, provinces instead of cities, etc.
- Upload transcripts instead of video/audio files
- Deface your MRI images
- Remove variables that contain too much personal information (especially if they're not necessary for reproducing your results or reusing your data)

Some types of data, such as video and audio recordings, can usually not be de-identified and will not be uploaded to DataverseNL. In some cases, you could invite interested researchers to Leiden to view this data under supervision. If you believe your data is too sensitive to share, please contact the data stewards and they'll discuss options with you. Even if your data cannot be shared, it is still valuable to make a publication package with your materials, instructions and procedures, as this can help other researchers who want to set up a similar study.

Changes to the uploaded publication package

For archiving and transparency purposes, publication packages on DataverseNL should be left unchanged for ten years. It is possible to add data or change details in the metadata (e.g., change email addresses). This will be listed as a second version. If it is necessary to remove data (e.g., due to an error), please contact the data stewards.

Contact

Willemijn Plomp & Jaap-Willem Mink, Data stewards

Mail: datastewards_psy_ped@fsw.leidenuniv.nl

Open Office Hours: Every Tuesday 13:00-15:00 at 4A22