

CWTS research data management guidelines

September 15, 2020

Prepared by Josephine Bergmans, Thomas Franssen, Nees Jan van Eck (coordinator), and Thed van Leeuwen

Introduction

This document presents the CWTS research data management guidelines. These guidelines can be seen as part of a broader effort to build trust in scholarship, by increasing reliability, validity, and reproducibility, as well as to promote academic integrity. Good data management is important for:¹

- Making research data findable and accessible, also in the long term.
- Guaranteeing the safety and confidentiality of data.
- Ensuring the quality of research.
- Data re-use specifically and progression of research in general.
- Increasing the visibility and impact of research.
- Compliance with requirements from funders, institutions, and publishers.

In line with the CWTS open science policy, the CWTS research data management guidelines enable researchers to make data openly available as much as possible. The guidelines also provide room for keeping data closed whenever necessary, for instance because of privacy or ownership issues.

General requirements and guidelines

CWTS follows the *Research Data Management Regulation of Leiden University*² (Annex 2) and the *Guidelines for the archiving of academic research for faculties of Behavioural and Social Sciences*³ (Annex 3). The most important requirements and guidelines are:

- All research projects must have a data management plan before the start of the project.
- During a research project, data must be stored securely, which means that the integrity, availability and, if required, confidentiality of the data must be guaranteed.
- After a research project has been finished, data must be managed in such a way that data is findable, accessible, assessable, re-usable, and sustainable.
- Data used in publications must be archived for at least 10 years.

These requirements and guidelines apply to everyone at CWTS who conducts research that results in a publication. In addition to CWTS staff members, this includes PhD candidates, retired colleagues, and visiting researchers. If a CWTS staff member supervises a research master student, the guidelines apply

¹<https://www.library.universiteitleiden.nl/researchers/data-management/rdm-checklist>

²https://www.library.universiteitleiden.nl/binaries/content/assets/ul2ub/research--publish/research-data-management-regulations-leiden-university_def.pdf

³<https://www.universiteitleiden.nl/binaries/content/assets/sociale-wetenschappen/pedagogische-wetenschappen/ethiek-commissie/gedragcodes/guidelines-for-the-archiving-of-academic-research-for-faculties-of-bss.pdf>

as well. The guidelines do not apply to bachelor and one-year master students, unless their research results in a publication.

The CWTS research data management guidelines specify how the above-mentioned requirements and guidelines are implemented in the specific context of CWTS. In the CWTS research data management guidelines, a distinction is made between three phases of a research project: prior to the research, during the research, and after the research. Special attention is paid to the third phase. Guidelines are provided for archiving data, materials, and information that form the basis for publications. Archiving is done in so-called publication packages.

Prior to the research

Before starting a research project, or in the proposal writing phase, you need to think carefully about the data that you are going to collect and use in your project and how you are going to manage, store, and, if applicable, share this data. If your project will generate large amounts of data, this usually requires specific data management facilities. This is also the case if your project makes use of personal data or data that is otherwise sensitive in nature. Writing a data management plan (DMP) before your project starts will help you to think carefully about all relevant data management aspects. It will also help you to comply with the data management requirements of Leiden University, research funders (e.g., NWO and H2020), and publishers. Most externally funded research projects and PhD projects require writing a DMP before the start of the project. In the DMP, you need to discuss the data that will be used in a project and the way in which the data will be stored. You also need to discuss the extent to which you will be able to comply with the FAIR principles (findable, accessible, inter-operable, re-usable). The Leiden University template for DMPs can be found in Annex 4. In Annex 5, an example is provided of a data management plan for qualitative research. This example was obtained from the department of Cultural Anthropology and Developmental Sociology at Leiden University. The Centre for Digital Scholarship at Leiden University offers support in writing a DMP.

During the research

During a research project, research data must be safely preserved. This means that the integrity, availability, and confidentiality (if required) of the data must be guaranteed. During ethnographic fieldwork, additional measures may be necessary to protect data using encryption software and passwords on laptops and external hard disks. You should make sure that all digital data you collect is systematically backed up. Non-digital data (notebooks, sensitive paper files) should be stored behind lock and key when not in use and should not be left unattended when in use. If these precautions are not possible during ethnographic fieldwork, you should explain this in your DMP. Research participants may sometimes rightfully claim both raw and processed data as their own. Decisions about data protection and sharing differ depending on the specific professional and ethical standards upheld at a particular moment in the evolving relationship between researcher, research participants, and audience. How to deal with data during ethnographic fieldwork should be discussed in your DMP.

Platforms and services for storing and sharing digital data in a safe way during a research project include:

- J-drive (research data): University network storage drive maintained by ISSC. Every night a backup is made.
- DataVerseNL: Online platform provided by DANS for storage, sharing, and registration of research data.
- SURFFileSender: Service provided by SURF for sending (large) files in a secure and encrypted way.

After the research

After a research project has been finished, the main results are usually disseminated in publications such as journal articles, conference papers, and book chapters. Following the guidelines of the faculties of Behavioural and Social Sciences in the Netherlands, CWTS researchers are requested to create a so-called *publication package* for each publication. Such a publication package must contain all information needed to validate the results presented in a publication, such as the raw data, the processed data, the documentation of qualitative data production processes, or the computer codes used to process and analyze the data. Elements that must be stored in a publication package are described in Annex 1. Publication packages are not made publicly accessible.

Furthermore, following the CWTS open science policy, CWTS researchers are expected to make their data openly available as much as possible. An exception is made for data that cannot be made openly available, for instance because of privacy or ownership issues.

When and where must a publication package be stored?

Within one month after the publication of a manuscript, a publication package must be created. CWTS researchers are responsible for supplying the publication package or providing the necessary elements of the publication package to the CWTS data manager. The CWTS data manager is responsible for managing and archiving publication packages in a protected folder on the J-drive. Publication packages will be stored for at least 10 years after the appearance of a publication.

Who is responsible for creating a publication package?

If the first author of a publication is affiliated with CWTS, this author, together with the CWTS data manager, is responsible for creating a publication package. If the first author is not affiliated with CWTS but other authors do have an affiliation with CWTS, these authors must make sure that the data is properly stored. If the first author is affiliated with a faculty of Behavioural and Social Sciences in the Netherlands, the other authors may assume that the first author will follow the guidelines of his or her institution and will create a publication package. The other authors then do not need to create a publication package. If the first author is not affiliated with a faculty of Behavioural and Social Sciences in the Netherlands, the authors affiliated with CWTS must make sure that a publication package is created, either by creating it themselves or by making sure that it is created by one of the other authors.

In the case of PhD candidates and research master students, supervisors are responsible for creating publication packages. Supervisors may delegate the execution of this task, but they carry the final responsibility.

Who has access to a publication package?

Only the CWTS data manager responsible for archiving and managing publication packages has access to the protected folder on the J-drive where publication packages are stored. To facilitate audits of publication packages, the CWTS Board of Directors and the FSW Faculty Board may grant reading access to specific individuals. They may for instance grant access to members of an academic integrity committee.

How to make data openly available?

Publication packages are not made publicly available. To make data openly available, the use of one of the following platforms and services is recommended:

- DataVerseNL: Online platform provided by DANS for storage, sharing, and registration of research data.
- EASY: Online archiving system provided by DANS for depositing and reusing research data. Example: <https://doi.org/10.17026/dans-xpt-2ues>.
- FigShare: Online open access repository supported by Digital Science where researchers can preserve and share their research outputs, including figures, datasets, images, and videos.
- Mendeley Data: Online data repository owned by Elsevier. Example: <https://doi.org/10.17632/5bxw69mzht.1>.
- Zenodo: Open access data repository developed under the European OpenAIRE program and operated by CERN. Example: <https://doi.org/10.5281/zenodo.3339177>.

Training, education, and support

At Leiden University, PhD candidates must write a DMP and must attend a course on how to write such a plan. Other researchers are encouraged to follow a data management training or workshop, for instance at the Centre for Digital Scholarship at Leiden University. Relevant workshops offered by the Centre for Digital Scholarship include:

- How to write a data management plan
- Bring your own data
- How to publish your data

Annexes

Annex 1. What must be stored in a publication package?

A publication package is intended to validate or replicate research. It is written in English. A publication package is not meant to share data for re-use. The following elements must be included in a publication package:

- The published (or accepted) manuscript, which includes a description of the problem definition, research design, conceptual framework, data collection, and the used methods.
- The instructions, procedures, design of the experiment, and stimulus materials (e.g. topic list, interview guide, questionnaires) that can be reasonably deemed necessary for validating or replicating the research. These materials must be available in the language in which the research was conducted.
- When using primary data, the raw data files (e.g., scientometric databases, recordings or transcripts of interviews, descriptions of observations, archive, and other source or media material). If the raw data files have been stored in an external archive (e.g., an archived version of a scientometric database), making a reference to the files in this archive is sufficient. Stored raw data must be anonymized so that it cannot be directly traced back to people or groups of people.
- The data files (either raw or processed) that were eventually analyzed when preparing the publication. This is not necessary if the raw data files were directly analyzed.
- Computer codes (e.g., SQL scripts , Atlas.ti, SPSS syntax, R codes) and brief descriptions of the steps taken to process the raw data into analysis data.
- Ethical approval from the Ethical Committee (if applicable).
- A readme file explaining which documents and files are included in the publication package and how they should be interpreted. The following information needs to be provided in the readme file:
 - For each document or file, the name of the person by whom the document or the file was created
 - Division of roles among authors, indicating at least who analyzed the data
 - Names of the people who collected the data
 - Date/period of data collection
 - If relevant: addresses of field locations where data was collected and of contact persons
 - Date on which the manuscript was accepted for publication, including a reference
 - Whether or not an ethical assessment was performed

Using the readme file, a fellow researcher must be able to validate or replicate the results reported in the publication.