DE HAAGSE HOGESCHOOL

Research Drive

Guidelines folder structure and data documentation

The Hague University of Applied Sciences and its researchers have a shared responsibility to store research data in a correct and secure manner, to retain the data for the required duration and to make it available for review considering the applicable (legal) conditions.

Research data are in themselves of great value and useful for reuse. The original research can be reproduced, which contributes to its validation. In addition, funding bodies and publishers may impose the condition that the data are archived and shared.

This guide contains principles for storing, archiving and sharing research data. It is useful for the researchers themselves during the research process and for the archiving process after the research.

Folder structure

In Research Drive, the following standard folders are offered within the main project folder:

- *Project documents*: research proposal, data management plan, collaboration / consortium agreement, authorisation matrix, etc.
  - with subfolder *Consent forms*
    [Be careful when allowing access to this folder.]
- *Raw data*
  [Be careful when allowing access to this folder. Ensure that identifiable personal data is stored separately from associated research data, linked by a unique reference.]
- *Anonymised data*
  [Ensure that personal data in the *Raw Data* folder are destroyed when they are no longer needed to achieve the project's objectives.]
- *Analyses* • *Publications and Products*

The names of the folders are preceded by the project code.

Within these standard folders you can set up your own folder structure. Make sure you have a solid naming convention for both your folders as well as files. This will make it easier to identify, locate and retrieve the research data. Use logical and clear file and folder names and be consistent in your naming. The fields in your spreadsheet also deserve the same attention. Also use a consistent date format.

Example of naming conventions

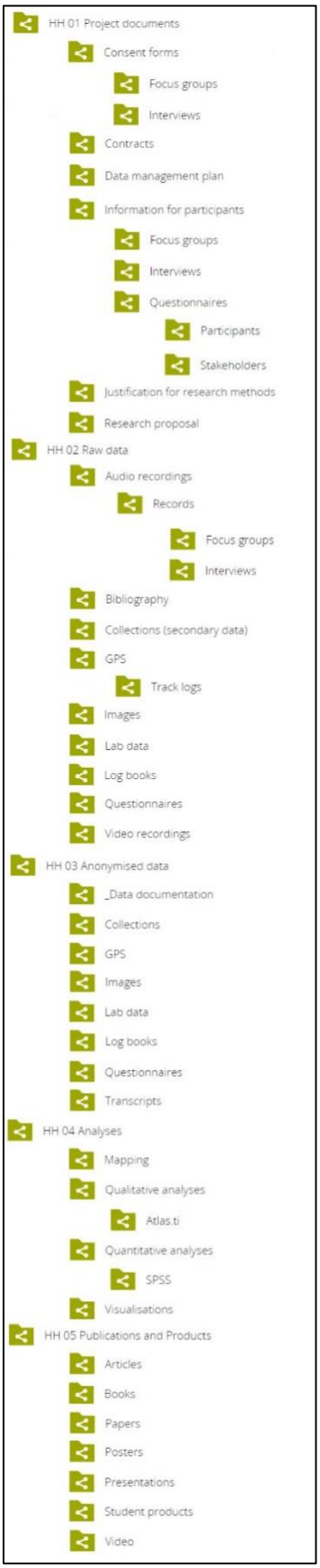
File names, no naming strategy applied	File names, naming strategy applied
• Test_data_2013.xlsx • Project_Data.docx • Design for project.docx • Lab_work_Eric.docx • Second_test.xlsx • Meeting Notes Oct 23.docx	• 20130503_DOEProject_DesignDocument_Smith_v2-01.docx • 20130709_DOEProject_MasterData_Jones_v1-00.xlsx • 20130825_DOEProject_Ex1Test1_Data_Gonzalez_v3-03.xlsx • 20130825_DOEProject_Ex1Test1_Documentation_Gonzalez_v3-03.xlsx • 20131002_DOEProject_Ex1Test2_Data_Gonzalez_v1-01.xlsx • 20141023_DOEProject_ProjectMeetingNotes_Kramer_v1-00.docx

Example of folder structure in Research Drive with base folders and subfolders

Version 1.0, February 2021, Leen Liefsoens, University of Applied Sciences Library

Working with a standardised list of keywords, a thesaurus, which is used by everyone within the research field, can also be advantageous. Standardised ontology and terminology also make it possible to combine and analyse different datasets. On bartoc.org you will find an overview of thesauri, ontologies and classifications. In Research Drive you can also work with keywords to make the folders and files easier to find. You can read how this works in the Research Drive manual.

On the library website (in Dutch, under the heading Bestandsnamen en mappenstructuur) you will find useful links (in English) with information on a good strategy for organising your research data. If available, follow the standard procedures and workflows in the relevant field. Communicate the strategy to all those who will be supplying data or working with the data.



Data documentation

To keep your research data useful for yourself and others, it is a good idea to pay attention to documenting your data regularly during your research project, and to keep this information in a text file that you save with your data. In your documentation, include information about data collection, data entry, data storage and data processing. Lab journals or code books also function as documentation of a dataset and are stored with the data if possible.

README.txt file

An overview of all the files of the dataset with a description of the content per file can be recorded in a Readme.txt file. The 4TU Centre for Research Data offers guidelines for writing such a README.txt file. With these guidelines you can also find an example of a good README file.

Metadata

Metadata is a special form of standardised data documentation or 'data about data'. Not only humans but also computers can read, interpret and combine metadata. The assignment of metadata helps, for example, to identify the time of collection of the data, the collection location, the creator(s) and the terms of use of the research data (licence).

The most common types of metadata are:

Type of metadata	**Objective**	**Example**
Descriptive metadata	These are the minimum metadata required to find a digital object. If contextual metadata is also present, users get more insight into how they can use the data themselves	Author, title, abstract, date Contextual metadata are for example location, time, methods of data collection (tools)
Structural metadata	These metadata establish the relationship between individual objects that together form a unit	Links to related digital objects, (e.g. the article written based on the linked research data)
Technical metadata	Information about the technical aspects of the dataset	Data format, hardware/software used, calibration, version, authentication, encryption, metadata standard
Administrative metadata	Metadata focusing on user (rights) and management of digital objects	Licence, possible grounds for embargo, waivers Search logs, user tracking

In addition, metadata can be classified on file level and on data level:

- Metadata on file level:
    - The context of data collection: project history, aim, objectives and hypotheses ○ Data collection methods: sampling, data collection process, instruments used, hardware and software used, geographical coverage and secondary data sources used
    - Dataset structure: relationships between files
    - Data validation: the procedures conducted in relation to checking, cleaning and quality assurance of the data
    - Changes made to data over time since their original creation and identification of different versions of data files
    - The legal implications of the research including privacy ○ Information on access and (re-)use or data confidentiality ○ The retention (periods) and disposal of (parts of) the dataset
- Metadata on data level:
    - Names, labels and descriptions for variables, records and their values ○ Explanations or definitions of codes and classification schemes used ○ Definitions of acronyms used ○ Codes of, and reasons for, missing values
    - Derived data created after collection, with code, algorithm or command file

Metadata on data level are preferably embedded within a data file itself. Should this not be possible a separate file with the required information should be linked to the data file. Many data analysis software packages have facilities for data annotation and description, as variable attributes (labels, codes, data type, missing values), data type definitions, table relationships, etc.

Assigning metadata is an activity that is mainly part of archiving the data after research. But as the saying goes: well begun is half done. It is therefore advisable to start this activity during the research. Metadata can also prove their usefulness at an earlier stage, for example when new researchers are added to the study. However, it is useful to take the later archiving process into account when assigning metadata during the research. The metadata fields that are obligatory or desirable differ per data archive and research discipline. It is therefore desirable to adjust your metadata to the standard of the storage location where you will be archiving your data after the research. DANS Easy, a Dutch online archiving system for the deposit and reuse of research data, uses the Dublin Core metadata standard. This is a generic metadata standard which is quite common internationally and includes the following fields:

Metadata field	**Meaning**
Creator	The key researchers involved in producing the data
Title	Name or title of dataset
Date created	Date in which the data were created
Description	Content description of the data
Audience	Audience for whom the dataset is of interest, described in terms of research domains
Rights holder	The person or organisation holding the copyright or intellectual property rights
Access Rights	A basic choice between Open Access or Restricted Access and an obligatory choice of the type of licence if Open Access is chosen (CC0-1.0; CC-BY-4.0 etc)

In addition to a generic metadata standard, you can also look at the prevailing metadata standard in your own field. On the library website (in Dutch, under the heading Metadatastandaarden) you will find useful links (in English) that provide an overview of metadata standards per subject area.

Sources and more info:

- File naming conventions: https://guides.lib.purdue.edu/c.php?g=353013&p=2378293
- File naming and folder structure: https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-andfolder-structure
- Data documentation: https://libguides.graduateinstitute.ch/rdm/documentation
- Metadata on file level and data level: https://srs.saxion.nl/wpcontent/uploads/2018/01/RDM_Guidelines2018.pdf
- Metadata standards: https://www.dcc.ac.uk/guidance/briefing-papers/standards-watchpapers/what-are-metadata-standards