Folder structure scheme for   
experimental research

In this document we provide best practices for digital file organization within the context of an experimental research project. While the suggested structure can serve as a backbone, personal and contextual adaptations should be obvious and easily applied. We assume that an experimental research project (e.g. a PhD) typically consists of one or more subprojects (i.e. research lines), and that subprojects consist of one or more experiments. In Figure 1, a tree view representation is shown of the folder structure discussed. What follows is a description of the folders, subfolders and their content.

Researcher\_YYYY\_MainProjectName\_SubProjectName/

├── 1\_Background

│ ├── 1\_Literature

│ │ └── Authors\_year\_title.pdf

│ ├── 2\_Documentation

│ ├── 3\_EthicsGovernance

│ │ ├── 1\_EthicsApproval

│ │ └── 2\_ConsentForms

│ └── 4\_Reports

├── 2\_Experiments

│ ├── 1\_Experiment1

│ │ ├── 1\_Materials

│ │ ├── 2\_DataRaw

│ │ ├── 3\_DataDerived

│ │ └── 4\_Analysis

│ └── 2\_Experiment2

│ ├── 1\_Materials

│ ├── 2\_DataRaw

│ ├── 3\_DataDerived

│ └── 4\_Analysis

└── 3\_Dissemination

├── 1\_Publications

│ ├── 1\_Draft

│ └── 2\_Submission

├── 2\_Presentations

└── 3\_Posters

Figure 1 – Tree view of the template folder structure

# Top Level folder

The top level folder name describes the main project and subproject conducted. Preferably it contains the researcher's initials, the year of conduction, the main project’s name and a subproject identifier. It is good practice to give both the main project and subproject unique and informative names.

Typically for a PhD-student the main project will be the PhD, although side projects can exist. A main project consists of one or more research lines (subprojects).

Naming template: Researcher\_YYYY\_MainProjectName\_SubProjectName/

Example: PVD\_2014\_PHD\_ContingencyAwareness

# Content subfolders

In experimental research a research line (i.e. subproject) typically consists of one or more experiments which are grouped conceptually. Therefore each subproject folder can have one or more “Experiment” folders. Besides the Experiment folders, background information can be stored in the “Background” folder, and all publicity material is placed in the “Dissemination” folder.

### Background folder

The Background level contains:

* Scientific “Literature”, for instance research articles of related studies.
* All “documentation” that helps to describe the current study. For instance data fact sheets, field notes, ...
* All things related to “Ethics governance”. For instance, application file to the ethics committee. Important: to avoid privacy issues, the consent forms are also stored at this level to separate them from the actual data stored at the Experiment-level.
* A data management plan.

### Experiment folder(s)

Multiple experiments can be stored within a subproject. To make sure these experiments are sorted in a sensible way, add a sequence tag to the name. This can be a simple number (e.g. Exp01, Exp02,...) or a date if that makes more sense (e.g. Exp\_20160605,...).

The Experiment folder contains the following subfolders: Materials, DataRaw, DataDerived, and Analysis.

#### Materials

The materials used to run the experiment. This can be computer programs, experiment code, stimuli, surveys, informed consent template,...

#### DataRaw

This folder contains the raw data. This refers to the first registration of what was meant to be measured. We recommend that the raw data is accompanied by adequate documentation (e.g. a codebook describing the variables). Raw data should remain unaltered. To avoid accidental changes to the raw data it is placed in a separate folder from derived data.

#### DataDerived

The derived data folder contains all data sets which are derivations of the raw data. To explain the transition from raw to derived data it is recommended to provide ample documentation. This also includes documentation describing the relation between different data sets. The documentation can be provided in this folder, either in a separate document or by means of annotated data processing scripts.

#### Analysis

The analysis folder contains all data analysis plans, scripts and output files. Also the documentation explaining analysis design, choices and set-up can be stored here (either in a separate document or by means of annotated data analysis scripts).

### Dissemination folder

The dissemination folder contains publications, presentations, posters and other publicity material.

# Some guidelines

* Avoid storing copies of identical files in different folders. For instance, it is tempting to copy files from the “DataDerived” folder to the “Analysis” folder. However, if you change one copy of these files you have to manually update all other copies. This is often a source of error and confusion. It is better to use scripts and make relative reference to data files in the analysis/processing scripts.
* Avoid confusion, try to be informative:
  + do not give identical files different names (but see also the previous point)
  + do not give different files identical names (e.g. data.sav).
  + Do not call a data file “data.dat” or an analysis script “analysis.R”. Be informative in you choice of file names. For instance, if you use a script called “exp1\_cfa\_preparation.R” and this script writes out a .csv data file, call the output file “exp1\_cfa\_preparation\_output.csv”. This way the filename describes the content and the link between both files is clear immediately.
* If you have different data files with similar data (e.g. numerical data matrix) use the same data formats (e.g. csv).
* Use appropriate extensions for plain text data files. A csv-file is in essence a plain text file, so a .txt extension is not wrong, but a .csv extension is informative about the type of plain text (i.e. it is a plain text file containing comma separated values).