

# Organize Data Projects Guidelines

*Reinhard Simon*

*November 17, 2016*

Based on a guide by K. Bromann and an article in PLOS.

1. Be organized
2. Automate
3. One directory for each project
4. Separate final data from raw data
5. Separate code from data
6. Use relative paths
7. Choose file names carefully
8. Write ReadMe files

## Basic outline:

### *Project name*

#### **README.md**

Describes the files and the process. Document changes.

#### **src**

A directory for raw or source data. May have further sub-directories as needed. Different data version should be in sub-directories named by date.

#### **data**

Final data for analysis. Should contain a sub-directory ‘code’ that has scripts transforming source data into final data.

#### **analysis**

A directory containing specific scripts for analysis. May contain sub-directories named by date corresponding to different data versions.

#### **docs**

For final docs with sub-directories for each derived document like a paper or presentation.