Organising your research data

Webinar

Organising your research data

Zoom Webinar

November 5, 2020 - 09.00–10.30

University of Oslo Library, Research Data Management Group

Host: Annika Rockenberger

Facilitators: Elin Stangeland, Ivana Malovic, Matthew Good

Contact: research-data@uio.no

Using Zoom for interactive participation

Zoom for interactive participation

Open the menu "Participants" and click on "Yes" in the bottom of the menu. Then click on "No".

Designing a data file structure

Designing a data file structure

- In an early stage of your research, you are faced with the question of what form your data files should take. Your initial decision about the structure of your data files should be considered thoroughly.

- The data file structure has a huge impact on the possible ways your files can be processed and analysed and once your structure has been filled with data, any changes to it are usually laborious and time-consuming.

Designing a data file structure

- Data files may have **different internal structures** and a research study may encompass **several different data files in different relations** to one another. The structure of the data files is also determined by the **formatting of its content** (e.g. types and organisation of variables). It provides information on the relationship among different elements and parts of its content.

- An important part of the **metadata** is often **embedded into the data file** (e.g. in the form of variable names and variable and value labels, different kinds of notes and content of supplementary variables). So, the **structure** of your data also **contributes to the clarity of** your **data documentation**.

Question 1

Have you designed a data file structure for a new/ongoing research project?

yes   no

Question 2

During a research project, have you had to change the data file structure significantly?

yes   no

File naming and folder structure

File naming and folder structure

- Structuring your data files in folders is important for making it easier to locate and organise files and versions. A proper folder structure is especially needed when collaborating with others.

- The decision on how to organise your data files depends on the plan and organisation of the study. All material relevant to the data should be entered into the data folders, including detailed information on the data collection and data processing procedures.

File naming and folder structure

Name	Date Modified	Size	Kind	Date Added
▼ 📁 Norwegian_Correspondences	Today at 20:56	--	Folder	Today at 21:00
📄 Icon?	29 Nov 2019 at 18:34	341 KB	TextEdi...cument	Today at 21:00
▶ 📁 .ipynb_checkpoints	29 Nov 2019 at 18:34	--	Folder	Today at 21:00
📄 README_NorKorr_GoogleDriveFolder.gdoc	27 Nov 2019 at 16:19	181 bytes	Google...cument	Today at 21:00
📄 Thesis Proposal-_ED-AR.docx	24 Jun 2019 at 19:00	36 KB	Word 2...cument	Today at 21:00
📄 DKNVS brevregistrant.pdf	3 Jun 2019 at 16:21	16,1 MB	PDF Document	Today at 21:00
📄 DKNVS brevregistrant med rettinger_AR-test.xlsx	3 Jun 2019 at 16:20	764 KB	3rd par...ormats	Today at 21:00
📄 DKNVS brevregistrant med rettinger.xlsx	3 Jun 2019 at 16:20	563 KB	3rd par...ormats	Today at 21:00
📄 Untitled spreadsheet.gsheet	21 Jan 2019 at 12:00	181 bytes	Google Sheets	Today at 21:00
📄 parse_html_beautifulsoup.ipynb	21 Jan 2019 at 11:44	28 KB	Document	Today at 21:00
📄 submission_NorKorr_DHN2019 (1).pdf	23 Nov 2018 at 13:45	3,1 MB	PDF Document	Today at 21:00
📄 rockenberger_NorKorr-LOD_manuscript.pdf	23 Nov 2018 at 13:44	118 KB	PDF Document	Today at 21:00
📄 AR_talk_norkorr_trondheim_2018-11-02.gdoc	23 Nov 2018 at 11:41	181 bytes	Google...cument	Today at 21:00
📄 Norgesbrev1840-1960.xls	29 Oct 2018 at 11:32	1,3 MB	3rd par...ormats	Today at 21:00
📄 HANSKE_NorKorr_collett_network.gsheet	26 Oct 2018 at 11:03	181 bytes	Google Sheets	Today at 21:00
📄 NorKorr_HANSKE_collett (1).csv	26 Oct 2018 at 10:49	12 KB	Comm...d values	Today at 21:00
📄 collett_letters.gtable	25 Oct 2018 at 14:41	175 bytes	Google table	Today at 21:00
📄 NorKorr_HANSKE_collett.gsheet	25 Oct 2018 at 14:40	181 bytes	Google Sheets	Today at 21:00
📄 NorKorr_hanske_signatur_brev.gsheet	25 Oct 2018 at 14:36	181 bytes	Google Sheets	Today at 21:00
📄 Project description.gdoc	3 Sep 2018 at 15:13	181 bytes	Google...cument	Today at 21:00
📄 parse_html_beautifulsoup.html	14 Aug 2018 at 13:59	280 KB	HTML text	Today at 21:00

Example of a research project **without** a data file structure – Annika Rockenberger ©

Question 3

Given the example before, do you think it is easy for new collaborators to navigate the project folder?

yes   no

Question 4

Name one thing you would want the project lead to improve in regards to the folder structure.

Use the chat to write your answer

File naming and folder structure

```
project_name/
├── README.md          # overview of the project
├── data/              # data files used in the project
│   ├── README.md      # describes where data came from
│   └── sub-folder/    # may contain subdirectories
├── processed_data/    # intermediate files from the analysis
├── manuscript/        # manuscript describing the results
├── results/           # results of the analysis (data, tables, figures)
├── src/               # contains all code in the project
│   ├── LICENSE        # license for your code
│   ├── requirements.txt  # software requirements and dependencies
│   └── ...
└── doc/               # documentation for your project
    ├── index.rst
    └── ...
```

Example of a research project **with** a proper data file structure. Image taken from CodeRefinery, Lesson on Reproducible Research. https://coderefinery.github.io/reproducible-research/02-organizing-projects/. Shared under CC-BY 4.0.

File naming and folder structure

Avoid using the following characters in directories (folders) and filenames

# pound	**%** percent	**&** ampersand	**** back slash
{ left curly bracket	**}** right curly bracket	***** asterisk	**?** question mark
< left angle bracket	**>** right angle bracket	**/** forward slash	blank spaces
$ dollar sign	**!** exclamation point	**'** single quotes	**"** double quotes
: colon	**@** at sign	**+** plus sign	**`** backtick
| pipe	**=** equal sign		

File naming and folder structure

Also, keep these rules in mind:

- Don't start or end your filename with a space, period, hyphen, or underline

- Keep your filenames to a reasonable length

- Most operating systems are case sensitive; always use lowercase

Question 5

Improve the following file names:

- Tech & Admin Costs.xls
- Image.3.png
- Article-revised-by-NN-Thursday.docx

Use the chat to write your answer

File naming and folder structure

- When you are actively working with research data and other project data, it is often recommended that you follow the conventions for use of software, analytical tools, programs, and data types

- When your project is finished and you plan on archiving your data, it is widely recommended to transfer data and accompanying files into a more sustainable format

- Trusted data archives often have guidelines for this, e.g. the recommendations of Data Archiving a Networked Services (DANS)
https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depo   rmats

Documentation and Metadata

Question 6

Do you know what a README-file is?

yes   no

Question 7

Have you written a README-file for data files?

yes   no

Documentation and Metadata

- Systematically documented research data is key to making the data publishable, discoverably, citable, and reusable (and FAIR)

- Clear and detailed documentation improve the overall data quality

- It is vital to document both the study for which the data has been collected and the data itself. These two levels of documentation are called project-level and data-level documentation

Documentation and Metadata

- The **project-level documentation** explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc.

- **Data-level or object-level documentation** provides information at the level of individual objects such as images or variables in a database/table or transcripts, etc.

- It's become a convention to create various **README-files**, both for project-level documentation and for data-level documentation

Question 8

Are you aware of metadata standards in your field of research?

yes   no

Documentation and Metadata

- Metadata or "data about data" are descriptions that facilitate cataloguing data and data discovery.

- Metadata are intended for machine-reading. When data is submitted to a trusted data repository, the archive generates machine-readable metadata.

- Machine-readable metadata help to explain the purpose, origin, time, location, creator(s), term of use, and access conditions of research data

- Your discipline very likely has standards for metadata!

Data Organisation – in practice

Data organisation – in practice

Carpentry@UiO

https://www.ub.uio.no/english/courses-events/courses/other/

We teach **Software Carpentry** and **Data Carpentry** workshops

Subscribe to announcements@carpentry.uio.no

Q&A

You can use the chat - or raise a hand!

Organising your research data

Slides by **Annika Rockenberger**
Licensed under Creative Commons Attribution 4.0 International License
CC-BY 4.0

The material for this webinar is partly based on:

CESSDA Training Resources, Ch. 2: Organise & Document. Created by CESSDA ERIC and licensed under a Creative Commons Attribution ShareAlike 4.0 License **CC-BY-SA 4.0**