D-Psy-FAIR Workshop

16 & 17 March, 2022

Leibniz Institute for Psychology (ZPID)

sy-FAIR

How to create a sustainable & high-quality data documentation of your entire research data process with the help of the D-Psy-FAIR standard

leibniz-psychology.org

Leibniz Gemeinschaft

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Let’s talk about data documentation…

- Which aspects of data documentation play a role in your daily research practice?

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Introducing the subject matter

1. Research data documentation and management in psychology - Why?

   - ➢ Transparency in scientific research
   - ➢ Replicability and traceability of research results
   - ➢ Quality-oriented (re)use of the data

   ⬇

   - ➢ Good data documentation is always a seal of quality for your data
   - ➢ Research data management as a fundamental requirement for the sustainability of research data

Introducing the subject matter

- Archiving data for reuse within own professorship / institution

- External archiving / publication of the data for public or scientific use (discipline unspecific platforms like OSF vs. discipline specific repositories like PsychArchives)

Introducing the subject matter

2.) The BMBF-funded project PsyCuraDat

(…) the goal of preserving the long-term interpretability of research data require substantial effort from researchers. This is because psychology is a complex field with multiple assignments to the social sciences, humanities, and natural sciences with correspondingly diverse, hardly standardized qualitative and quantitative methods. Common standards for the documentation of psychological research data that can meet these discipline- and method-specific requirements and thus guarantee the long-term interpretability and re-usability of these data are largely lacking at the moment. (…)
To this end, the project PsyCuraDat aims at the development of user-oriented curation criteria considering the needs of researchers in their role as contributors and users of research data. Thus, the overarching goal is to enable a more effective and efficient documentation and re-use of psychological research data."

https://leibniz-psychology.org/en/institute/third-party-funded-projects/psycuradat/

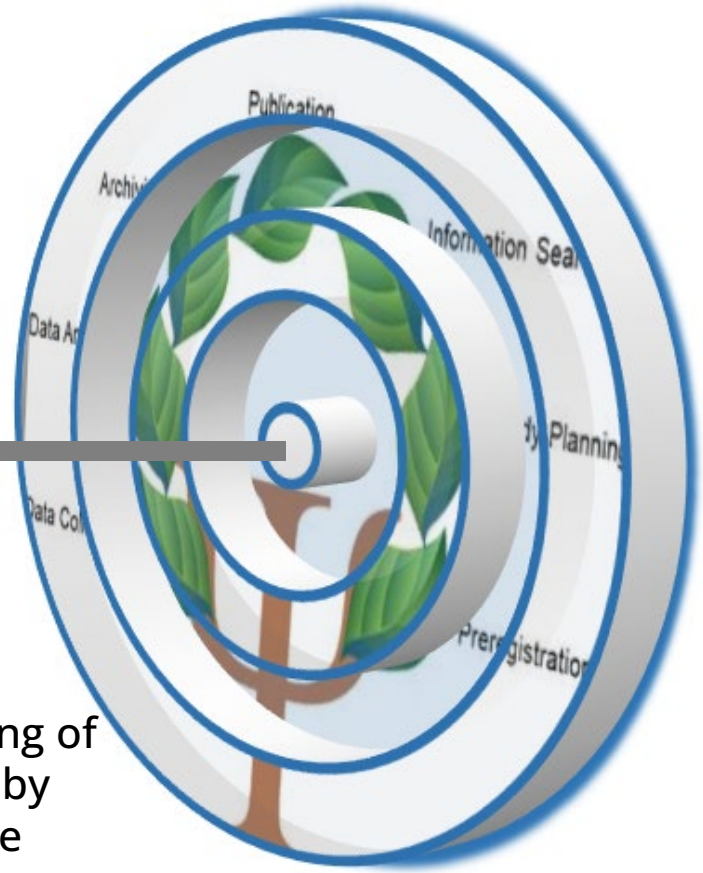
Goals of the PsyCuraDat project

Curation criteria

Development of user-oriented curation criteria for psychological research data.

Re-usability

Improving and facilitating re-usability of the research data.



Funding: BMBF (Federal Ministry of Education and Research)

Duration: 3 years (2019-2022)

Project manager: Dr. Katarina Blask

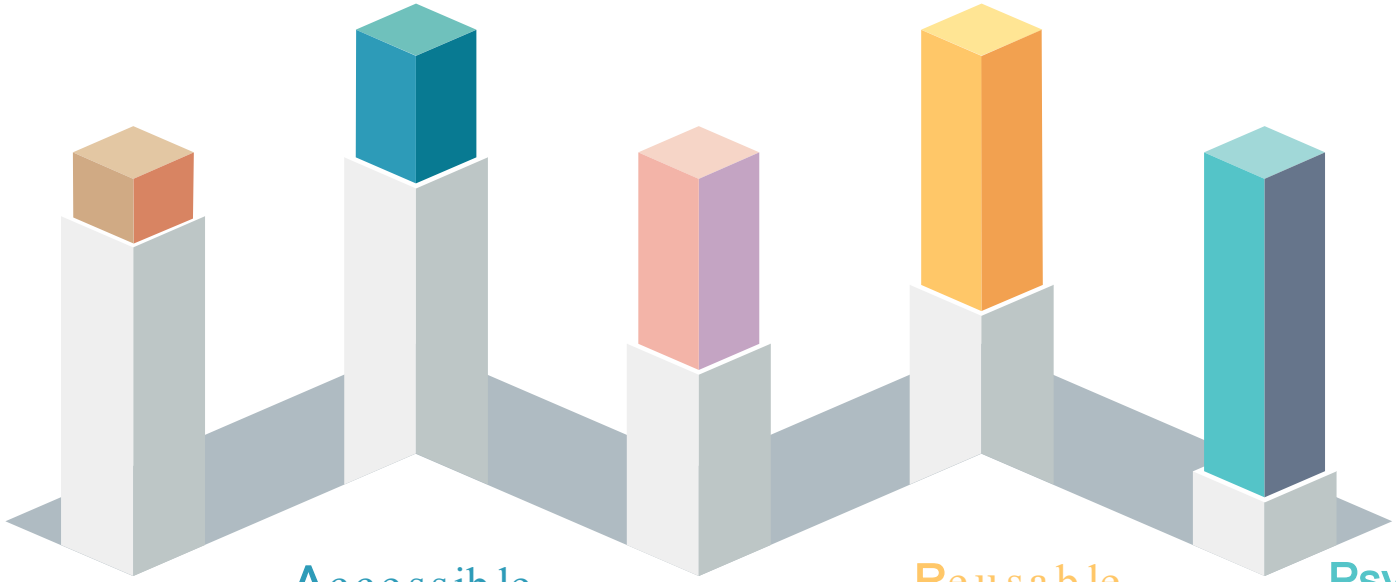
Standard

Establishing a coherent standard for sharing psychological research data.

Data Sharing

Increase sharing of research data by introducing the standard.

User-oriented curation criteria



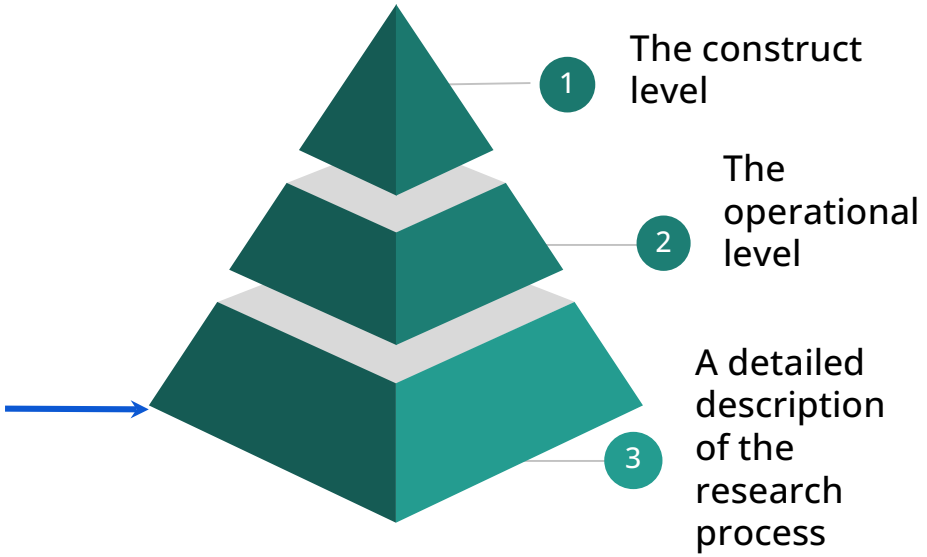
Findable	Accessible	Interoperable	Reusable	Psychological Research Process
• Persistent Identifier • Metadata	Retrievable & accessible (Meta)data	(Meta)data use a formal language for knowledge representation	(Meta)data are richly described with relevant attributes	Findings from PsyCuraDat

D-Psy-FAIR- Providing a key to your data

The D-Psy-FAIR Standard - a user-friendly curation standard enabling the sustainable reuse of psychological research data

Method-specific

- based on method-specific curation criteria considering the needs of researchers in their roles as data providers and data users
- enables a detailed documentation of all central phases of the psychological research process by providing a three-level-structure



D-Psy-FAIR - Providing a key to your data

The D-Psy-FAIR Standard - a user-friendly curation standard enabling the sustainable reuse of psychological research data

Effective & efficient

- enables an effective and efficient documentation and organization of psychological research data for the individual research process
- guarantees the data´s long term interpretability & reusability

User-friendly

- provides an information architecture with a maximum of usability: an easy-to-use and easy-to-learn data documentation structure

How to use D-Psy-FAIR - A 4 step data documentation

1. Write a study description

This requires information on different aspects especially important for an increased findability and reusability of your data. Including information on the creators of the dataset as well as on the study conducted to produce these data.

2. Create a codebook

A codebook is essential for the reusability of your data and should allow other researchers to understand all the variables collected in your study.

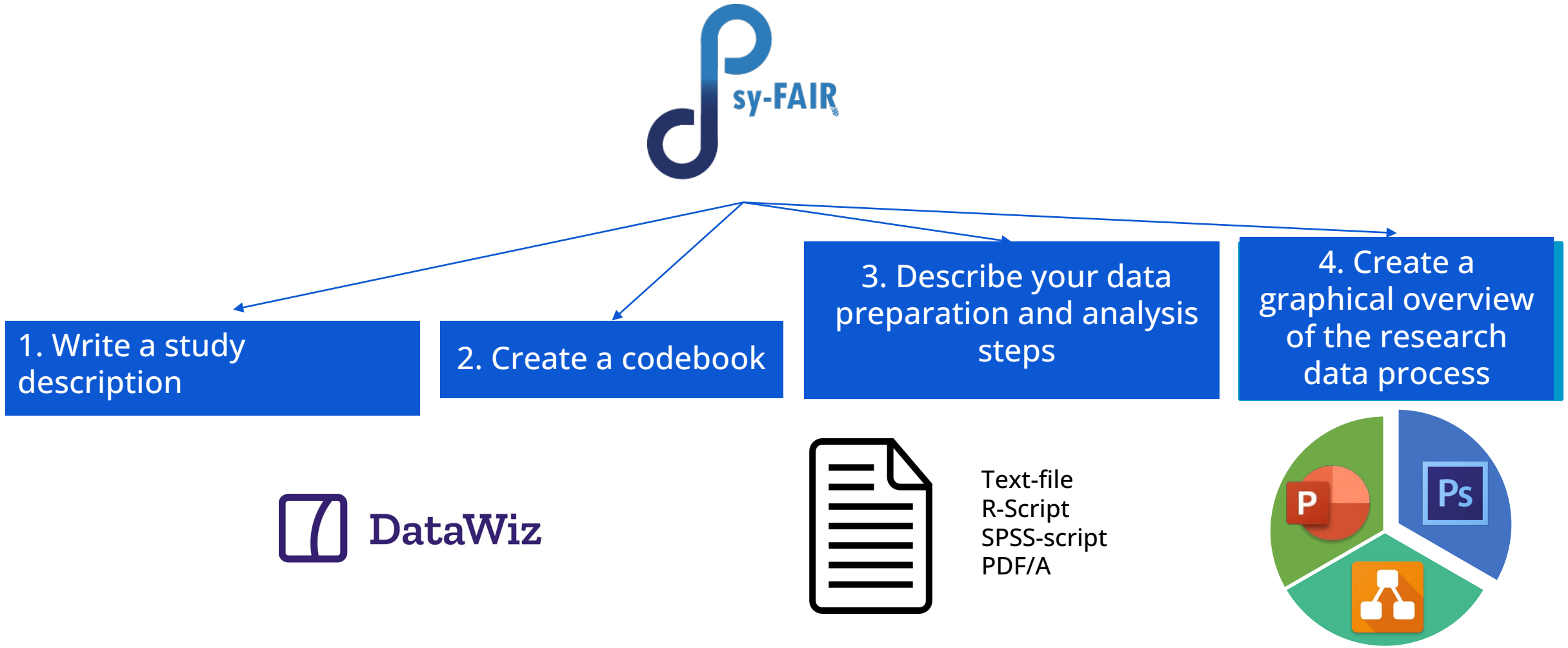
3. Describe your data preparation and analysis steps

Illustrate every single step you have taken to prepare your data so that it can be used for analysis (i.e. re-coding or transformation). Give a short description of each analysis that you have carried out (e.g. by adding appropriate comments).

4. Create a graphical overview of the research data process

Describe all the important steps in the research data process on which your data is based (e.g. in a flow chart).

Just 4 steps- Easy to integrate in your research process

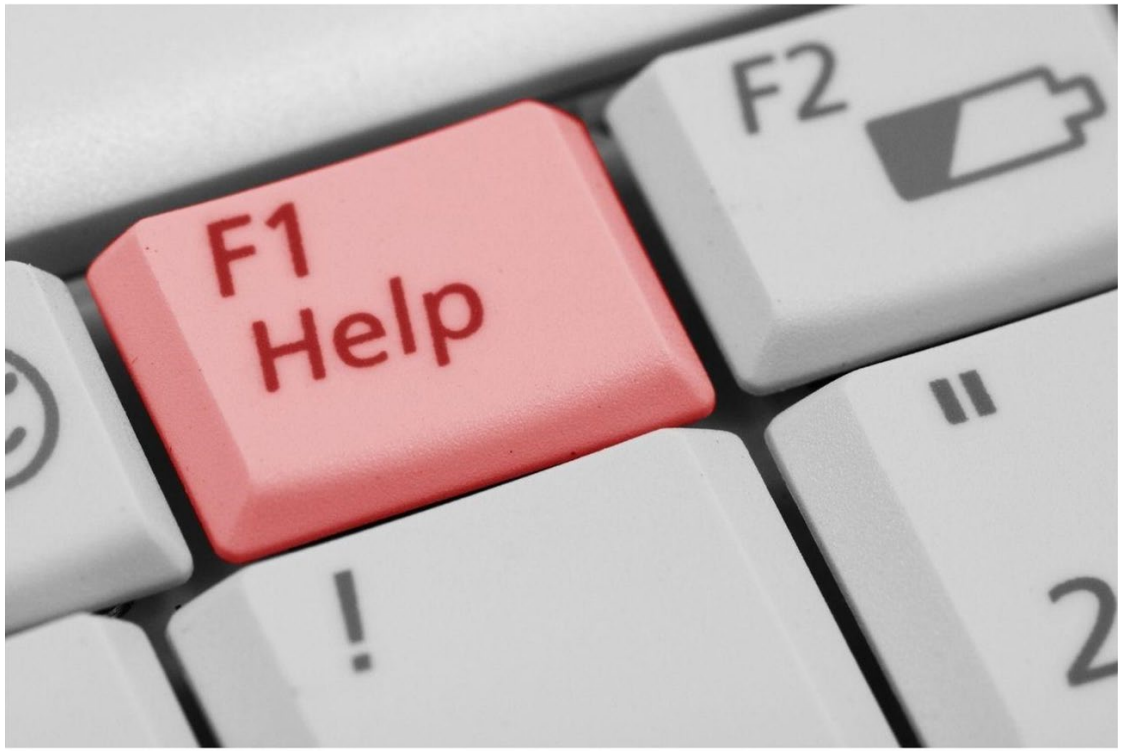


Agenda

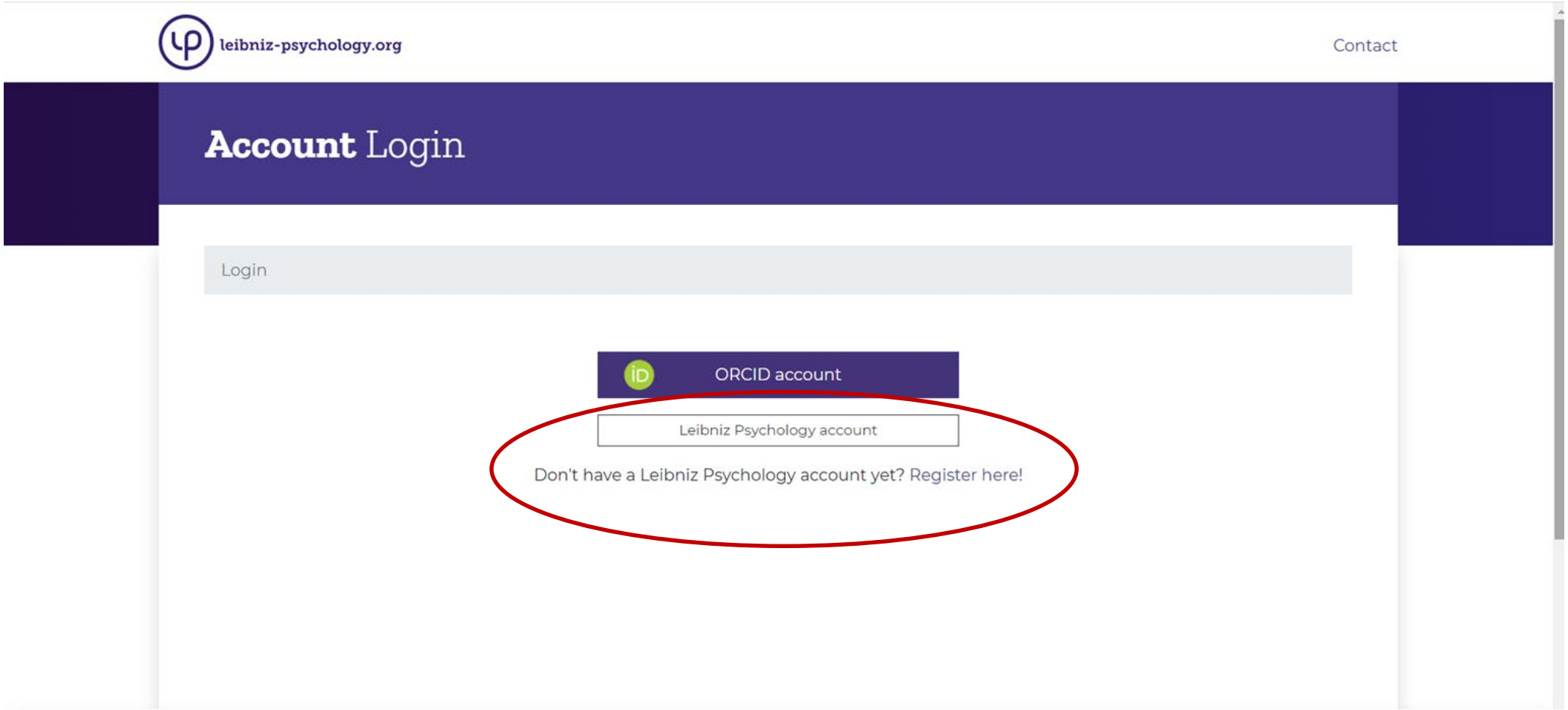
Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Technical check

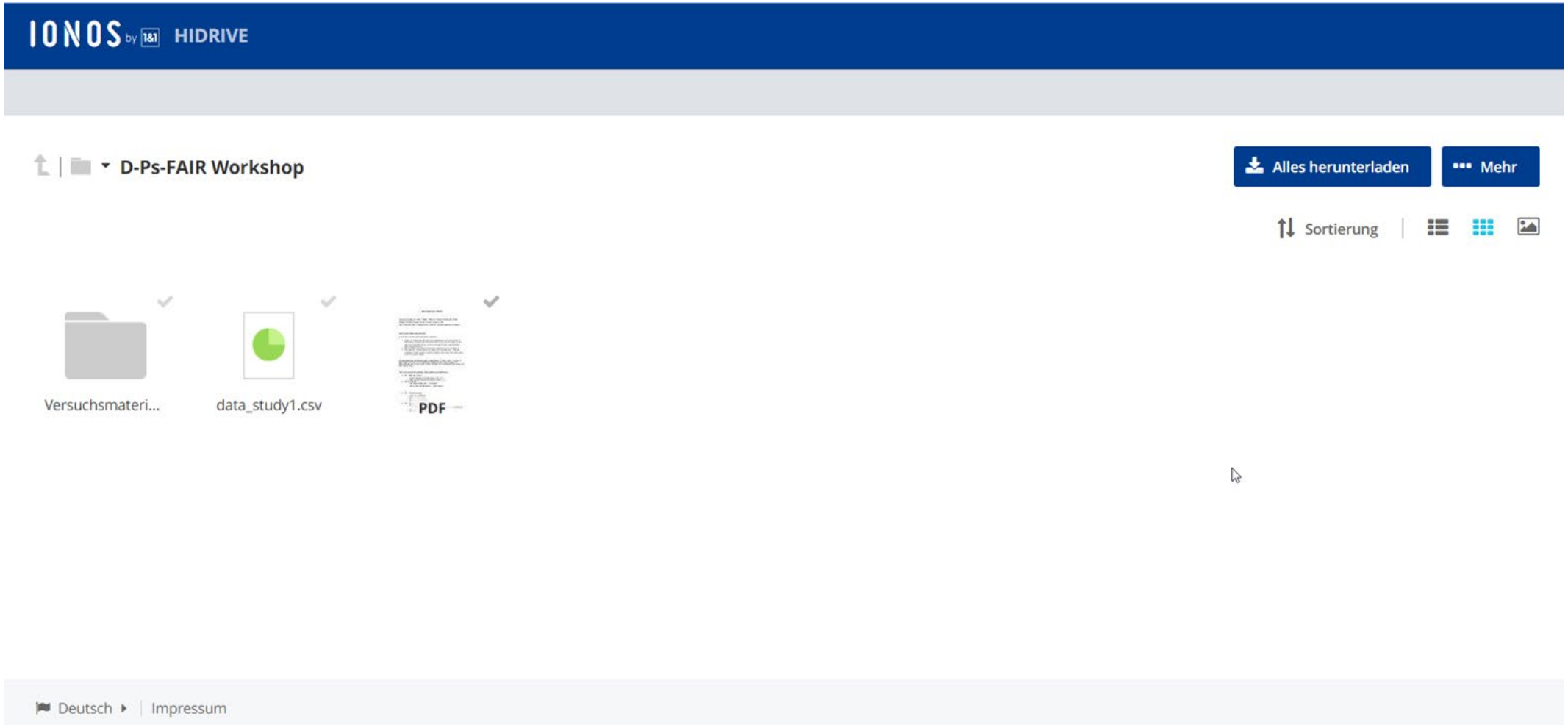
- Have you registered to *DataWiz* yet?
- Are the materials at hand?



Registration & account login



HIDRIVE Folder



Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Block A - Creating a study description with D-Psy-FAIR

Why is a study description so important?

- contains relevant metadata and all additional information necessary for the interpretation of the actual data
- provides important metadata enabling reusability and also findability of the published research data
- relevant metadata: e.g. bibliographic metadata, descriptive metadata

Block A - Creating a study description with D-Psy-FAIR

A detailed study description according to the D-Psy-FAIR standard includes information on the following aspects of your study:

- Information about the authors / creators of the dataset
- Publications associated with the dataset
- Objectives & hypotheses
- Sample & design:
  - Inclusion & exclusion criteria
  - Population
  - Sampling method & sampling size
  - Setting
  - Methodological approach
  - Number of measures
  - etc.

Registration in DataWiz

DataWiz beta

Research data management made easy!

Dashboard Data documentation User Profile Logout

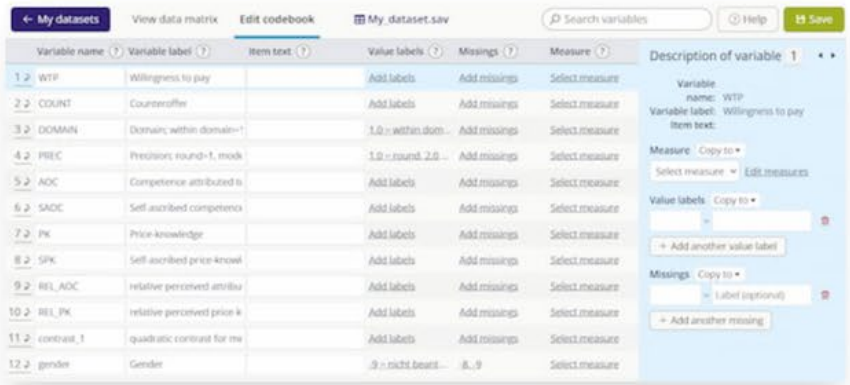
DataWiz — the data management tool for researching psychologists

Here's how DataWiz helps you during the whole research process:

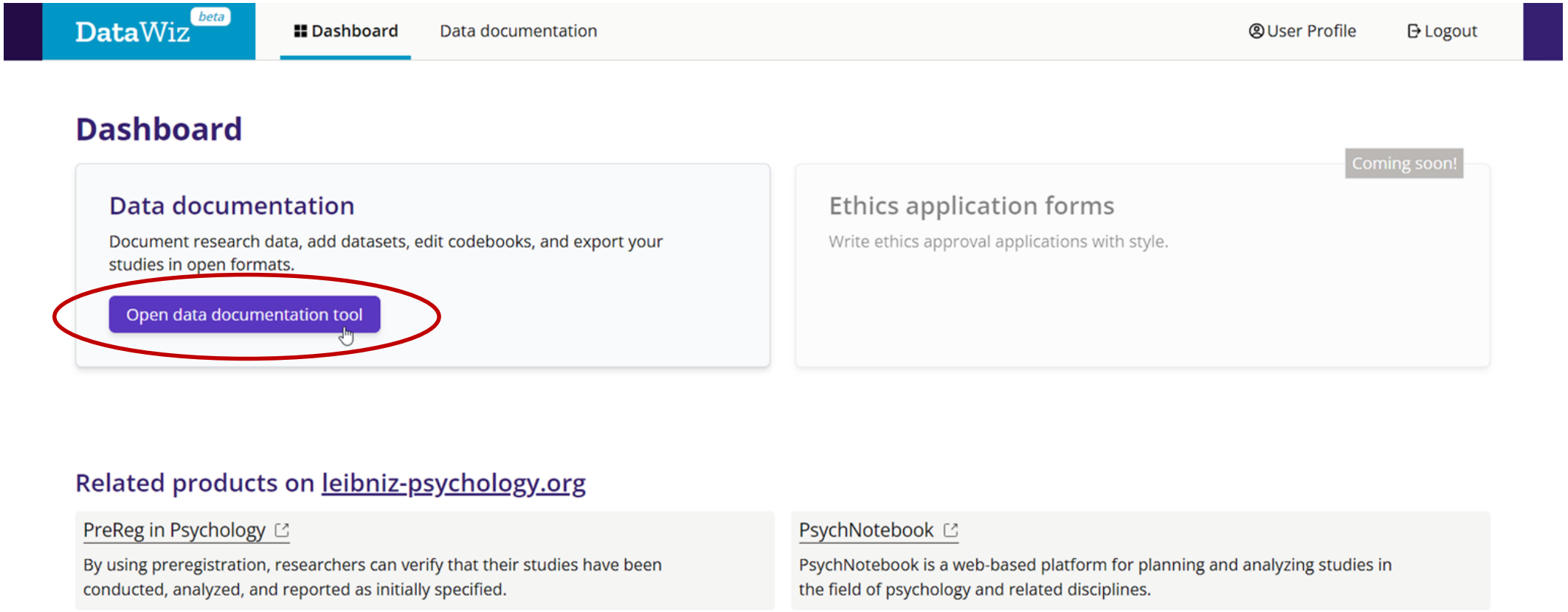
Data documentation

Years of experience with the research data center PsychData have shown that data management is most effective when it is instated early, preferably concomitant with the research process.

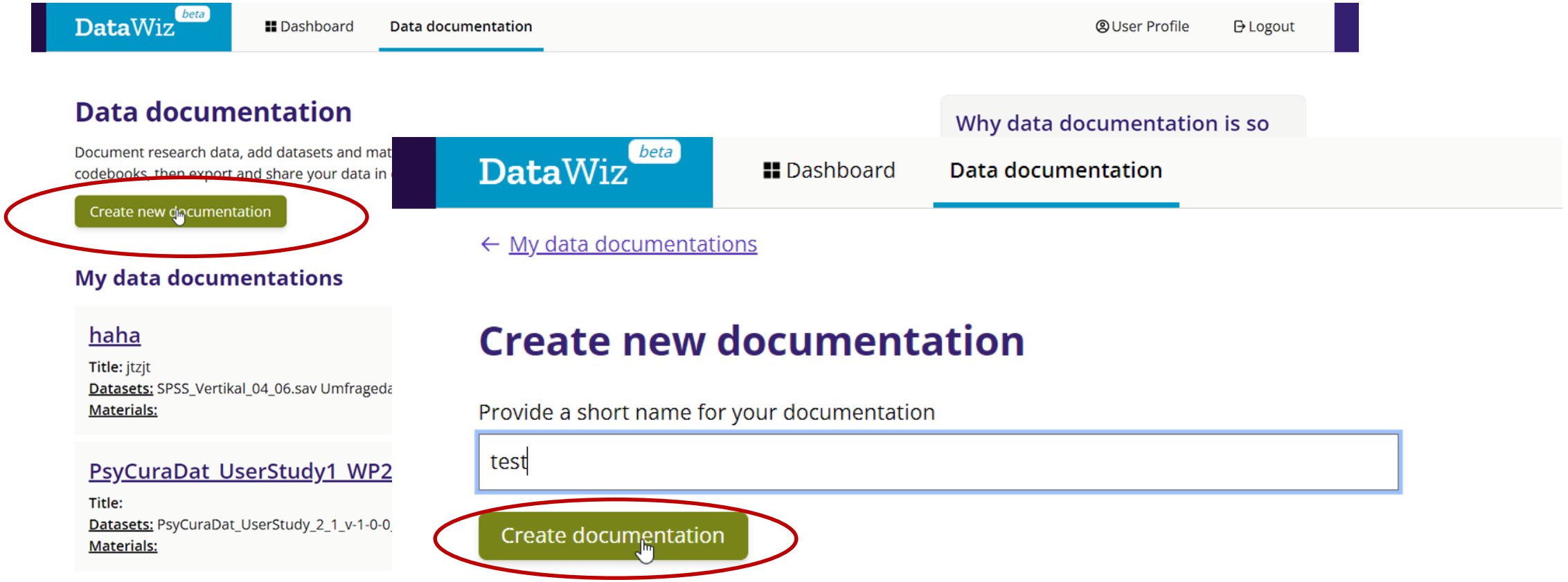
- Document your studies
- Add and edit codebooks
- Upload materials
- Export in open data formats



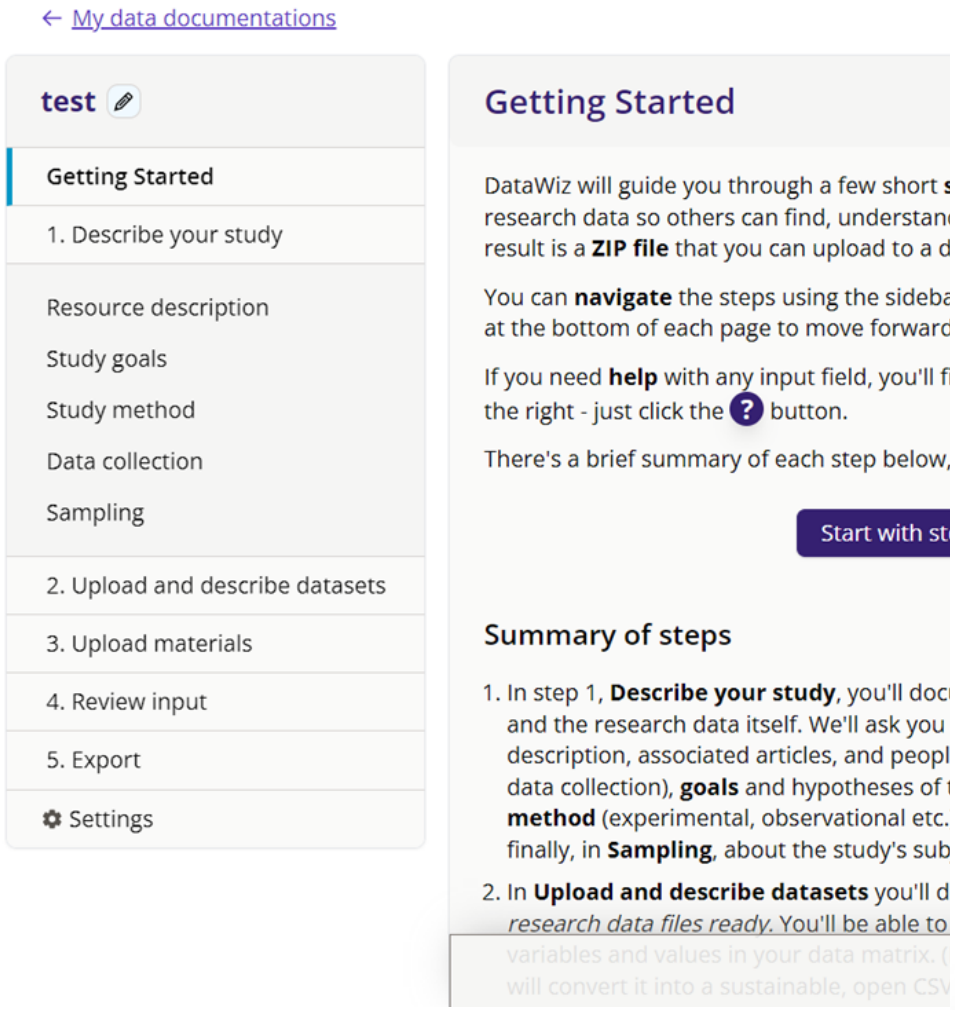
Using DataWiz2 to create a study description

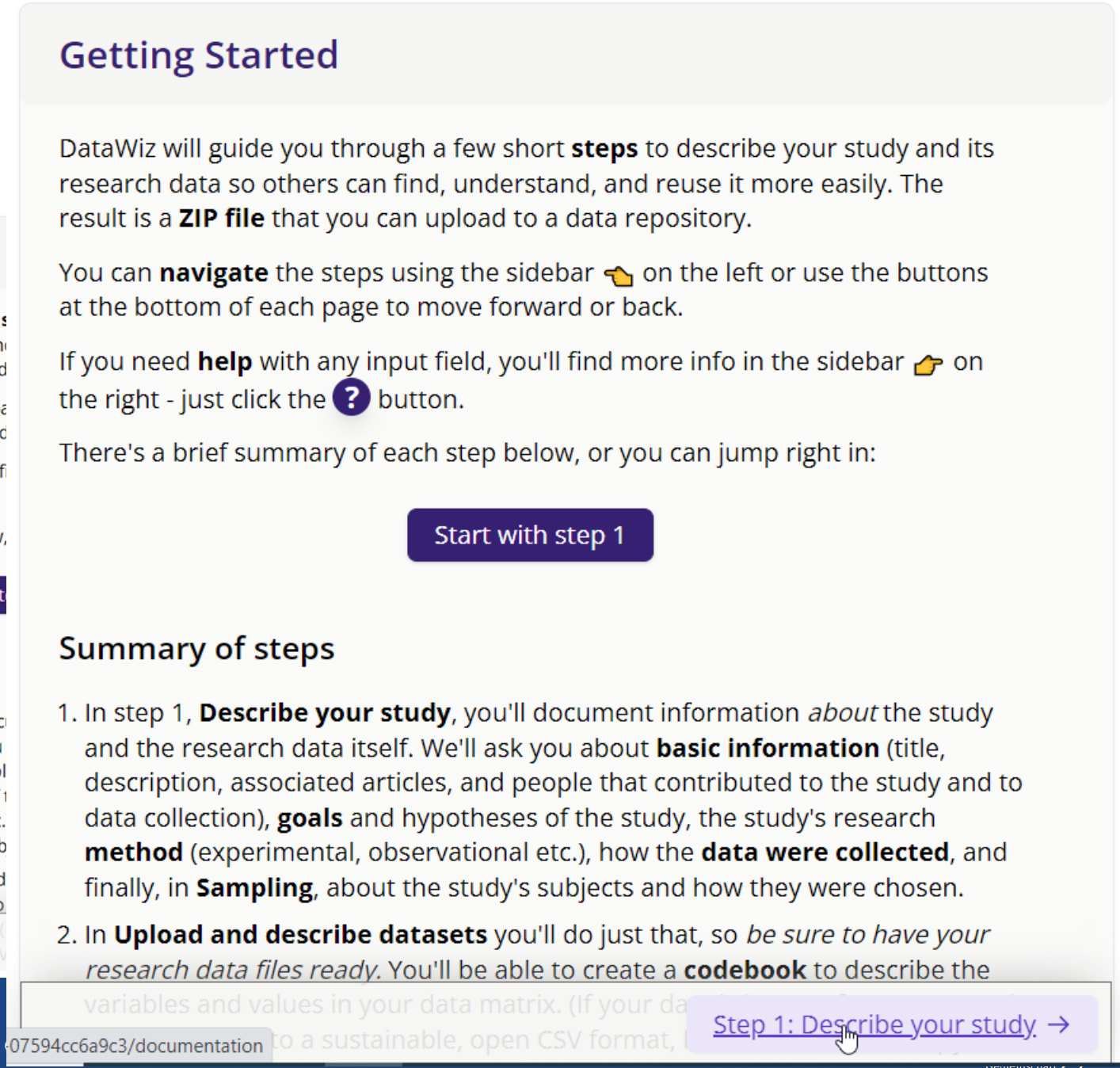


First steps in DataWiz

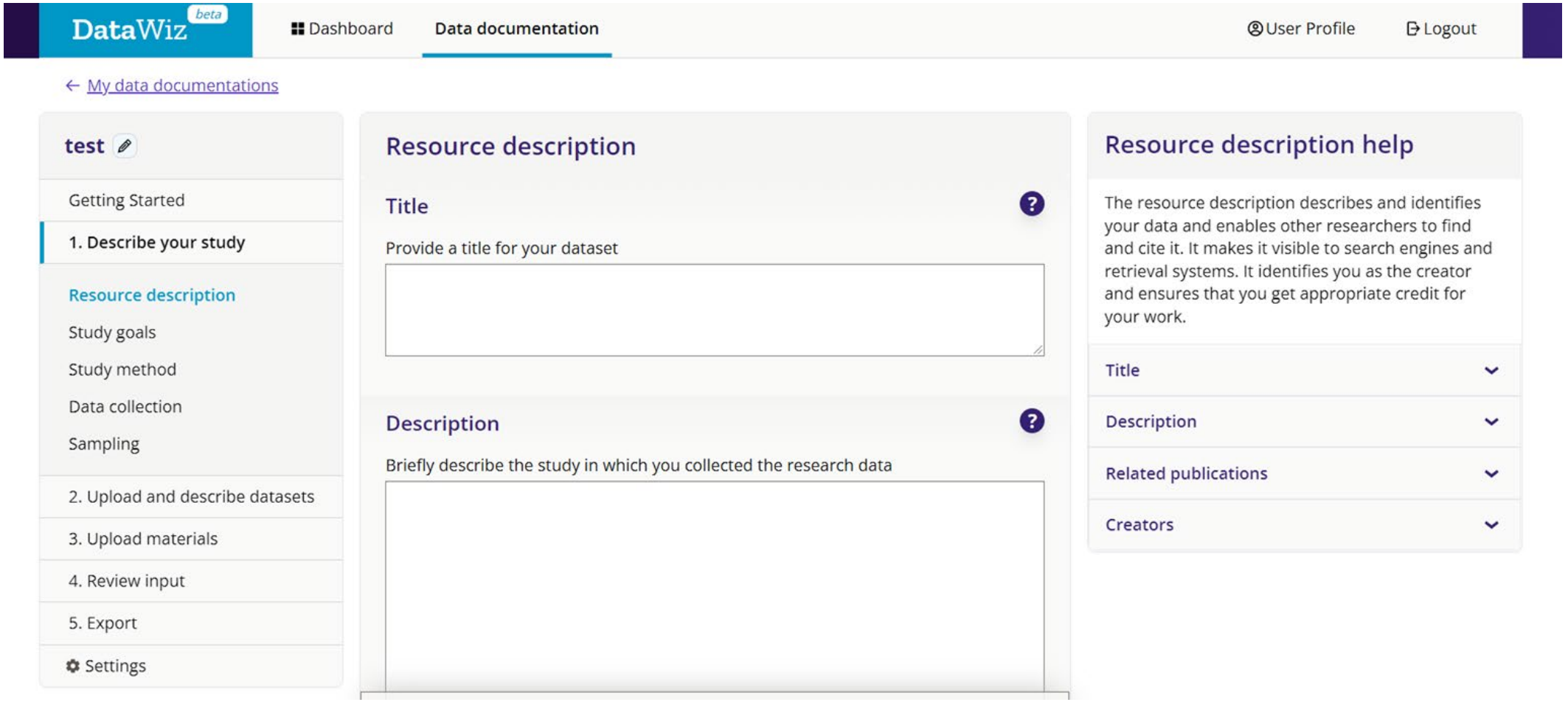


First steps in DataWiz

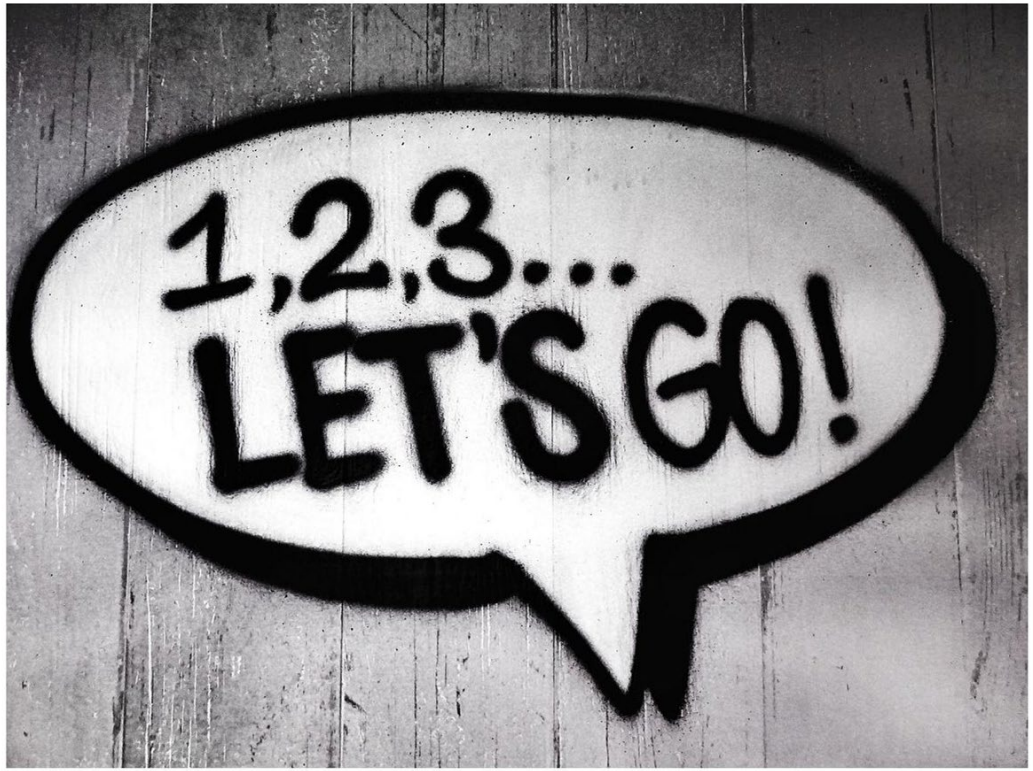




Creating a study description



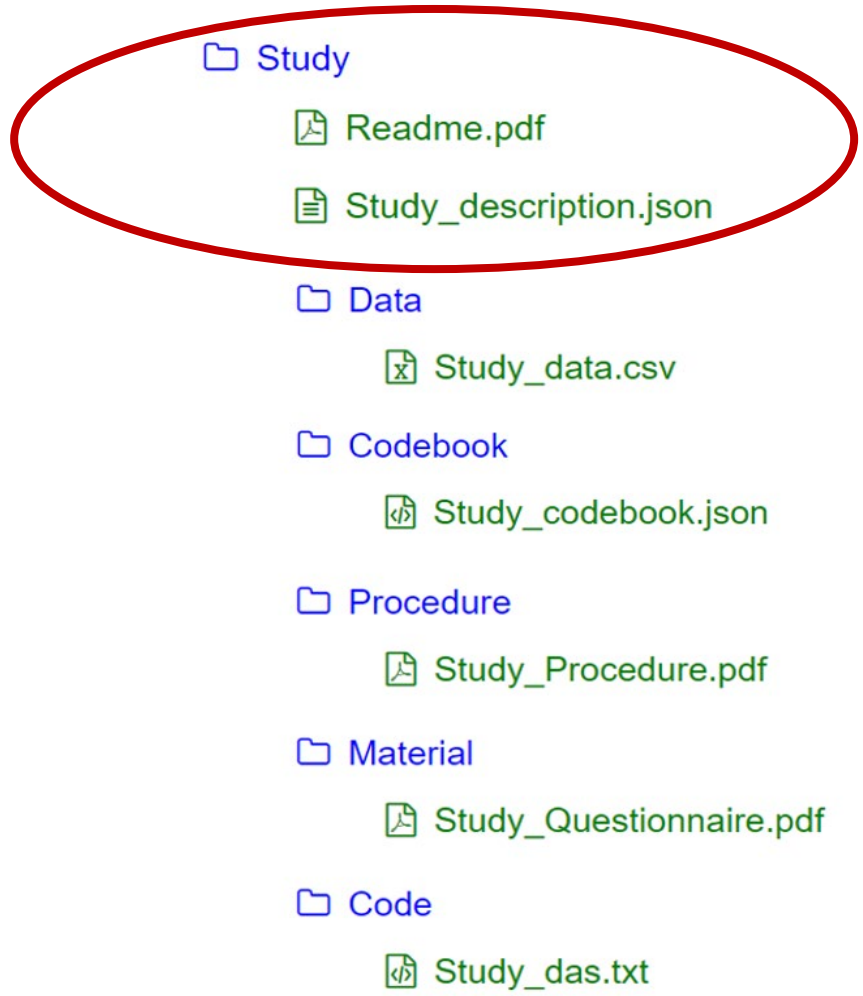
Theory into practice



And now it's your turn!

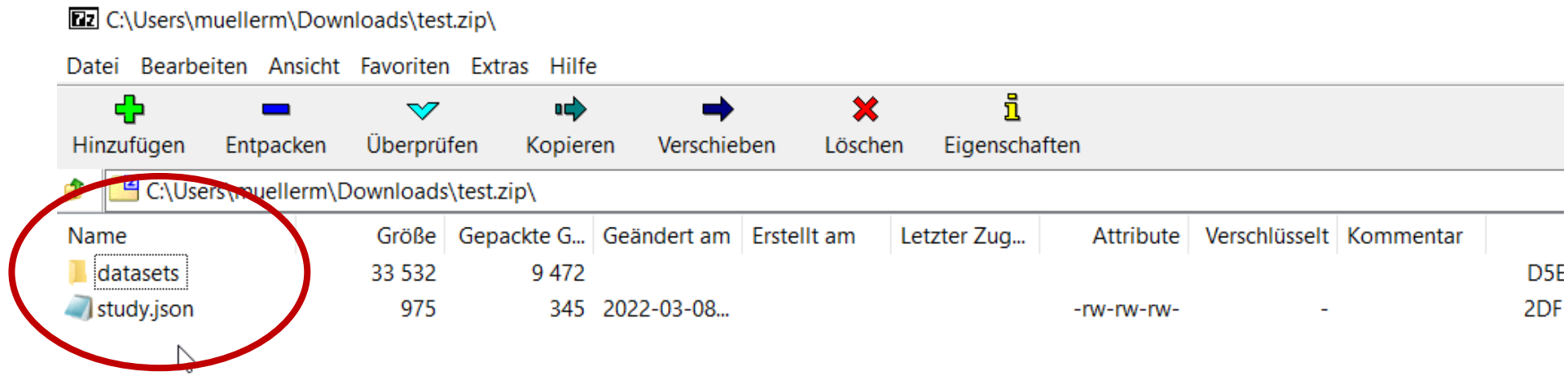
Managing your data documentation with D-Psy-FAIR

D-Psy-FAIR structure



Managing your data documentation with D-Psy-FAIR

DataWiz output / documentation export



Evaluation & discussion

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break → See you in 10 minutes!
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Block B - Creating a codebook with D-Psy-FAIR

What is a codebook and why is it so important?

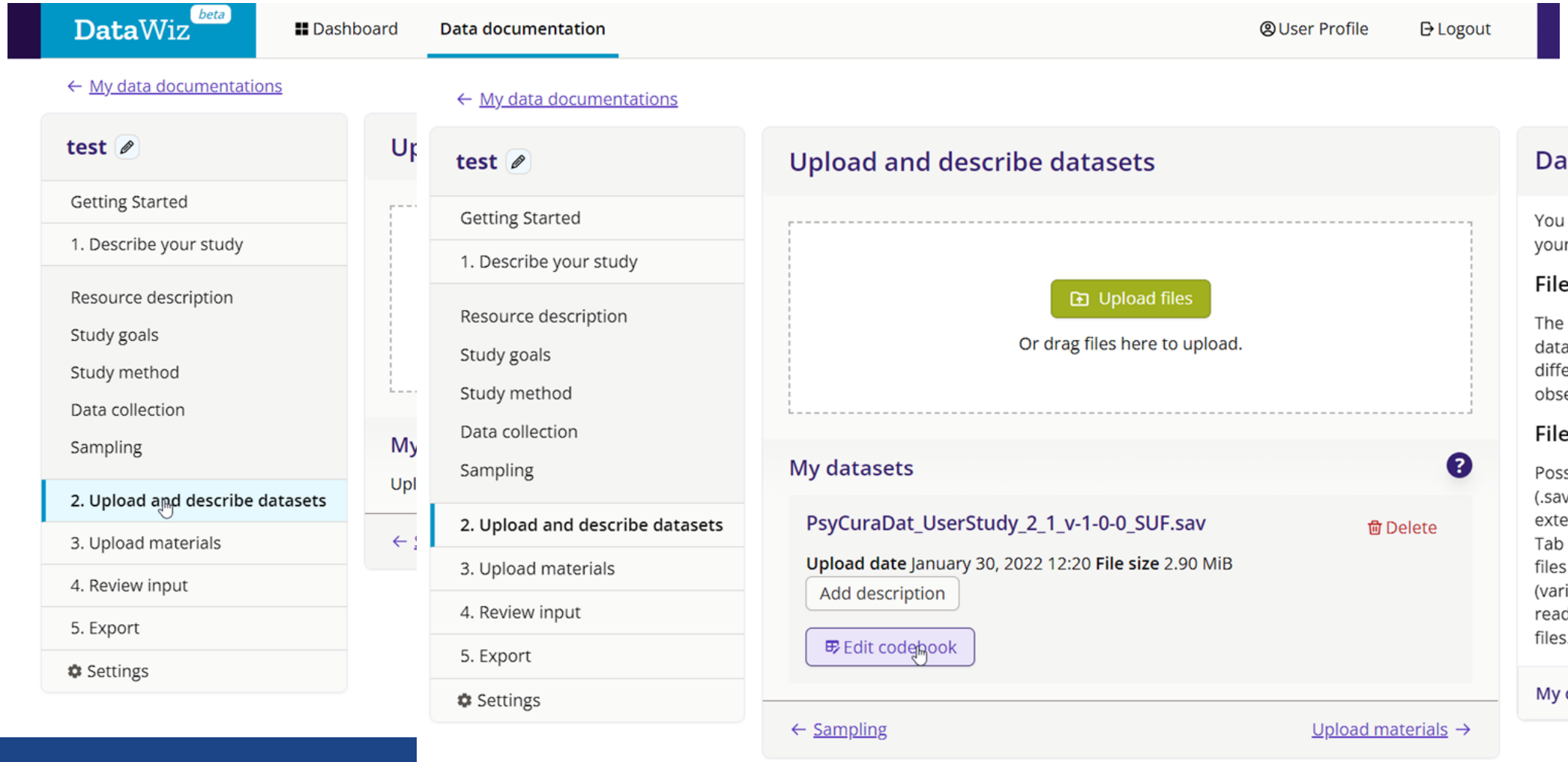
- formally the codebook can be best described as a tabular list describing the contents, structure, and layout of a data collection by providing detailed variable descriptions
- provides important metadata enabling reusability and also findability of the published research data
- ideally it enables the user to fully understand and replicate the realized procedure of the study: therefore, the order of the variables in the codebook should reflect the actual procedure

Block B - Creating a codebook with D-Psy-FAIR

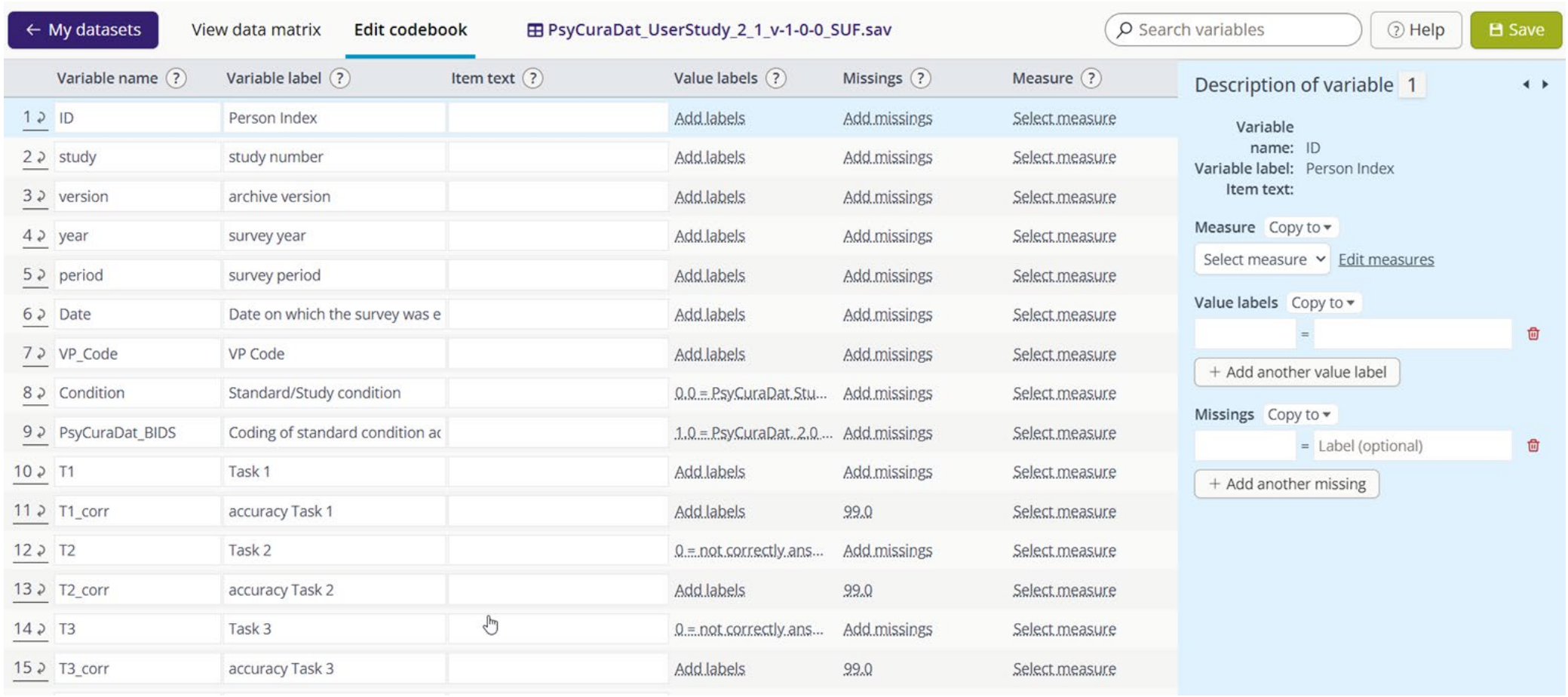
Therefore, and according to the D-Psy-FAIR standard a codebook should contain the following information about each variable within the dataset:

- Name = a short meaningful name of the variable, relating to the variable´s content; only small letters should be used; instead of space, a underline character should be applied (e.g. us_pos)
- Label = a more comprehensive description of the variable (e.g., positive unconditioned stimuli)
- Item text = the exact question in surveys; the instruction text
- Values = describes the value range of a given variable; in case of paradata (e.g., instructions, material) the relevant file name wherein the concrete information should be indicated
- Missing = specifies missing values (e.g., 99 for unanswered questions)
- Measure = the measure used for the dependent variable as defined in the study documentation

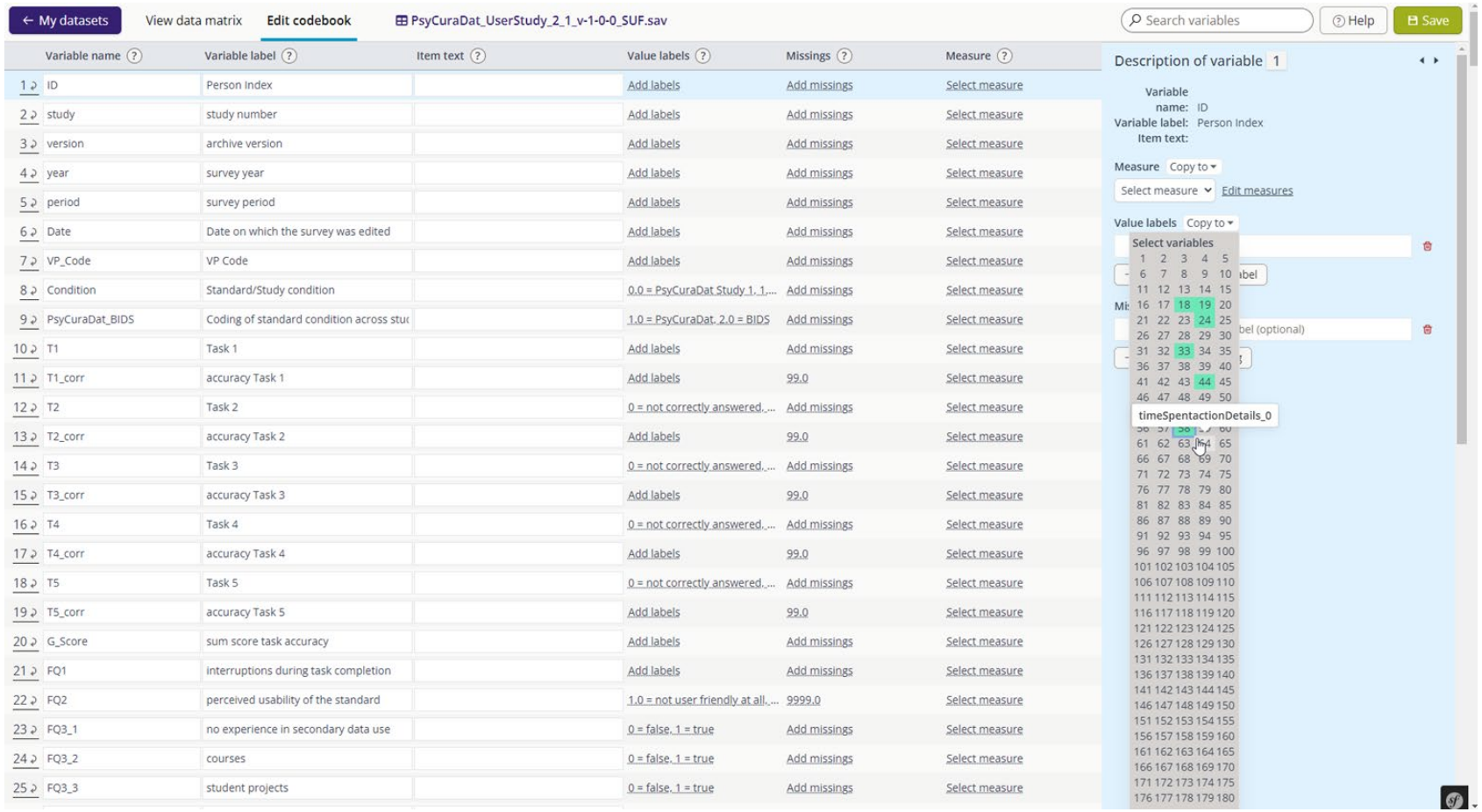
Creating a codebook in DataWiz



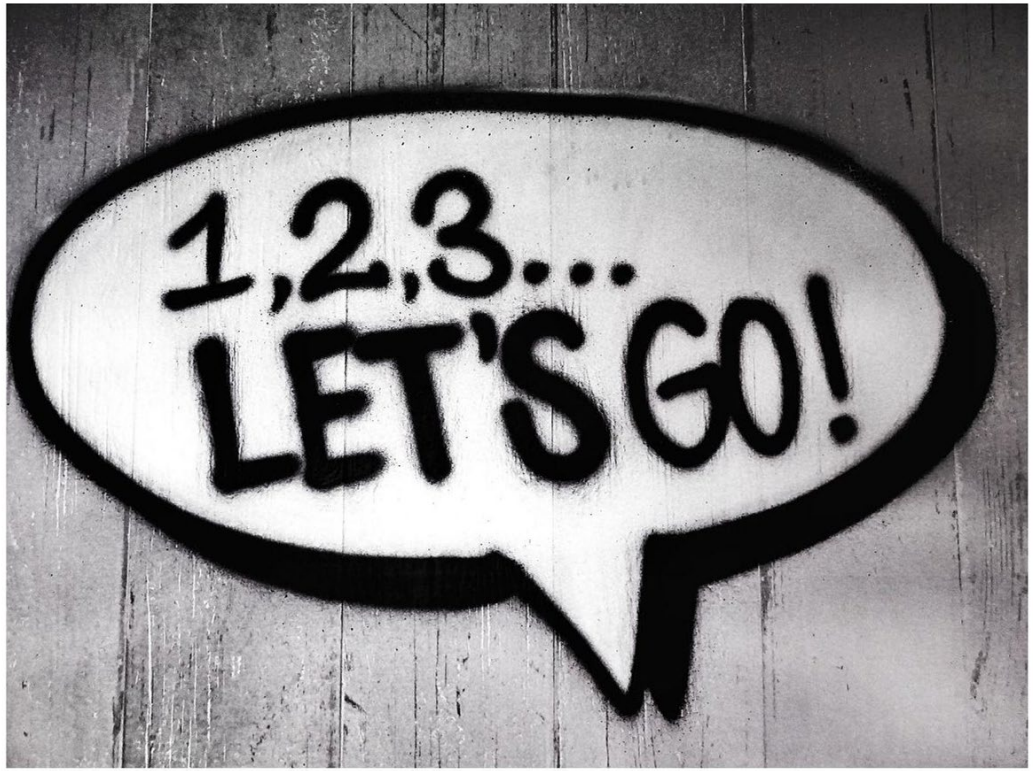
Creating a codebook in DataWiz



Creating a codebook in DataWiz



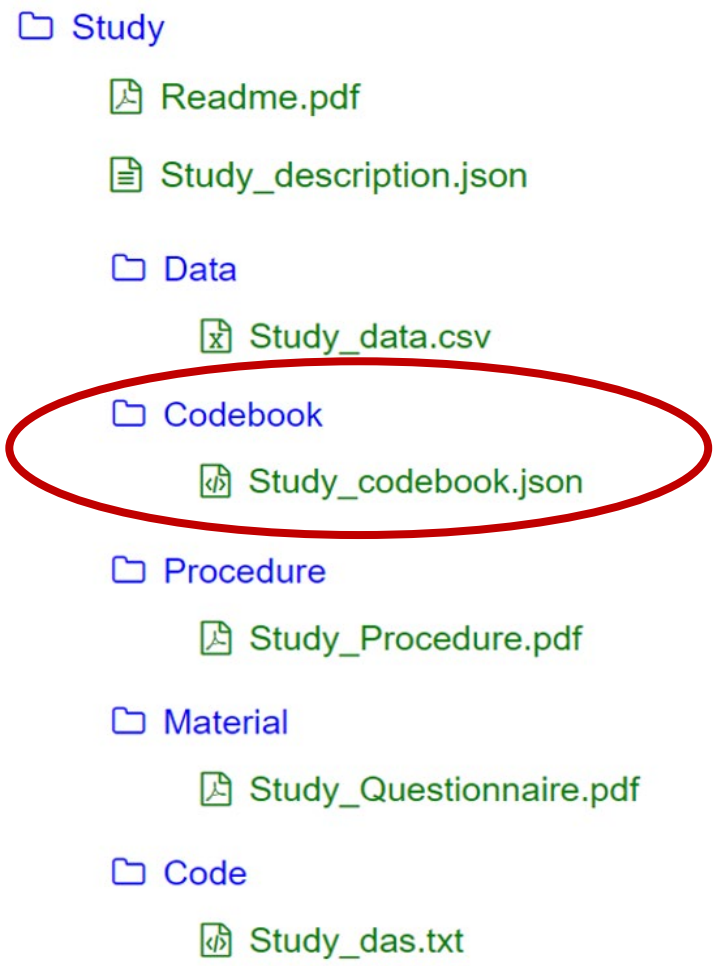
Theory into practice



And now it's your turn!

Managing your data documentation with D-Psy-FAIR

D-Psy-FAIR structure



Evaluation & discussion

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Block C - Documenting the data preparation and analysis steps with D-Psy-FAIR

Describe every single step you have taken to prepare your data so that it can be used for analysis.
Give a short description of each analysis that you have carried out (e.g. by adding appropriate comments in your evaluation script).

Example data analysis script (SPSS)

```
* Encoding: UTF-8.
****Use file 'Study1_Data.sav' to reproduce analyses for Study 1

**Screening for multivariate outliers

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT VPN
/METHOD=ENTER Emo_mean Mind_mean FourN_mean
/SAVE MAHAL.


COMPUTE Probability_MD=1- CDF.CHISQ(MAH_1,3).
EXECUTE.

**Set filter:
* Probability_MD has to be p <=.001

USE ALL.
COMPUTE filter_$=(VPN ~= 100).
VARIABLE LABELS filter_$ 'VPN ~= 100 (FILTER)'.
```

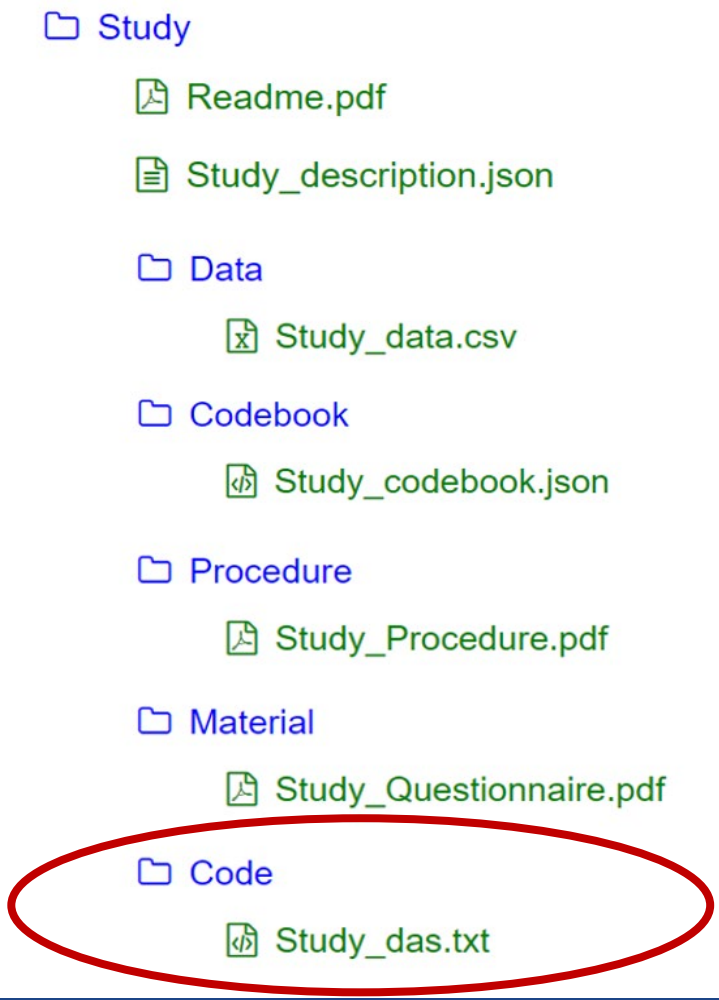
Block C - Documenting the data preparation and analysis steps with D-Psy-FAIR

Documenting data preparation and analysis

- Goal: Documenting the steps from raw(-est) data to analyzed data as best as possible
- Provide non-proprietary files
  - UTF-8 coded
  - open(ish) formats
  - the more comprehensible the better (proprietary syntax, coding habits)
- Usecase dependent
  - In-script documentation vs. external documentation

Managing your data documentation with D-Psy-FAIR

D-Psy-FAIR structure



Evaluation & discussion

Agenda

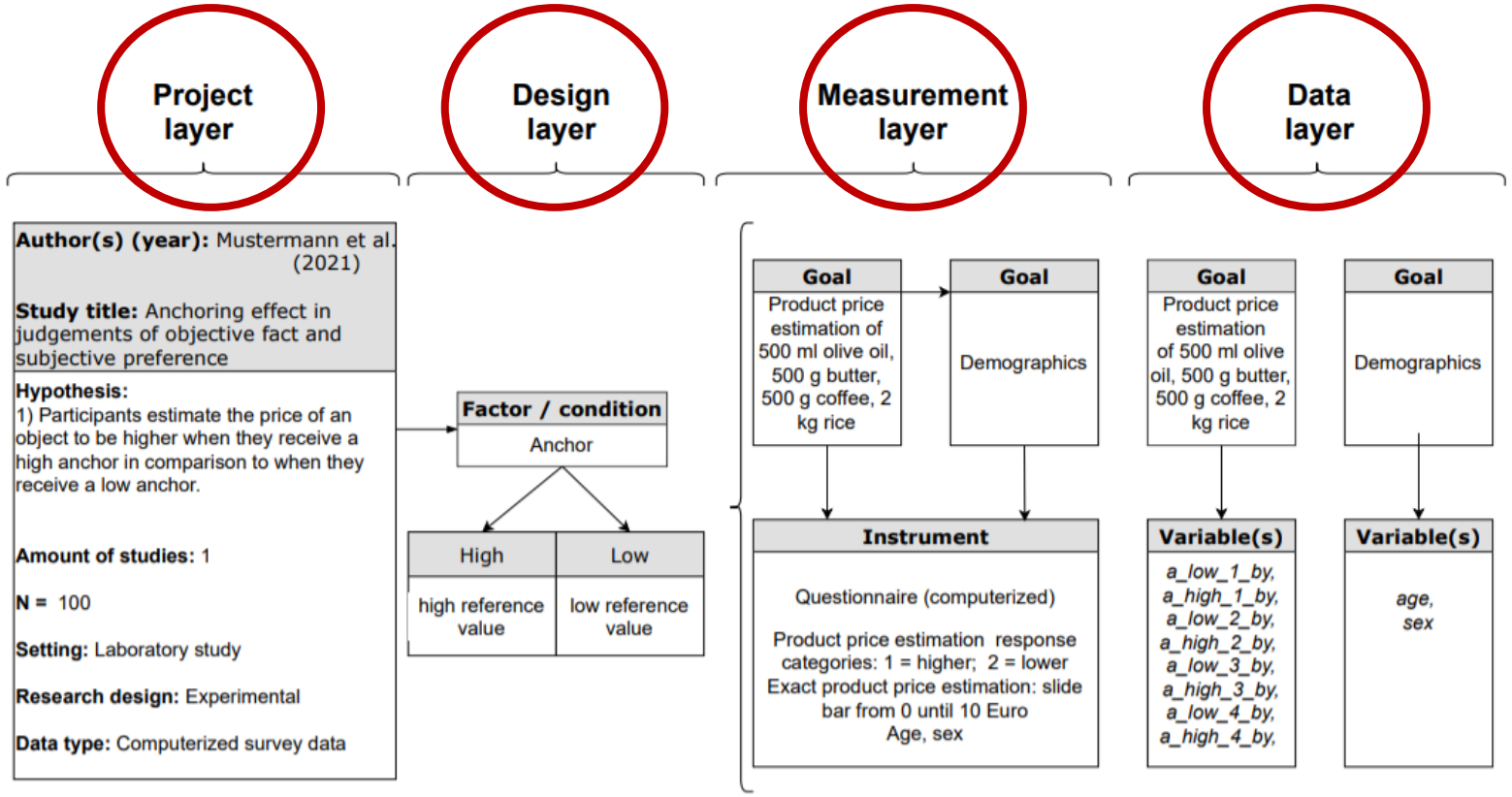
Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break → See you in 10 minutes!
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Block D - Creating a graphical overview of the research data process

- Describe all the important steps in the research data process on which your data is based.
- Start with the project / study structure picturing the research design and its implementation.
- Then move on to the various research methods that you have used,
- and finish with the mapping of the data that have resulted from their application.



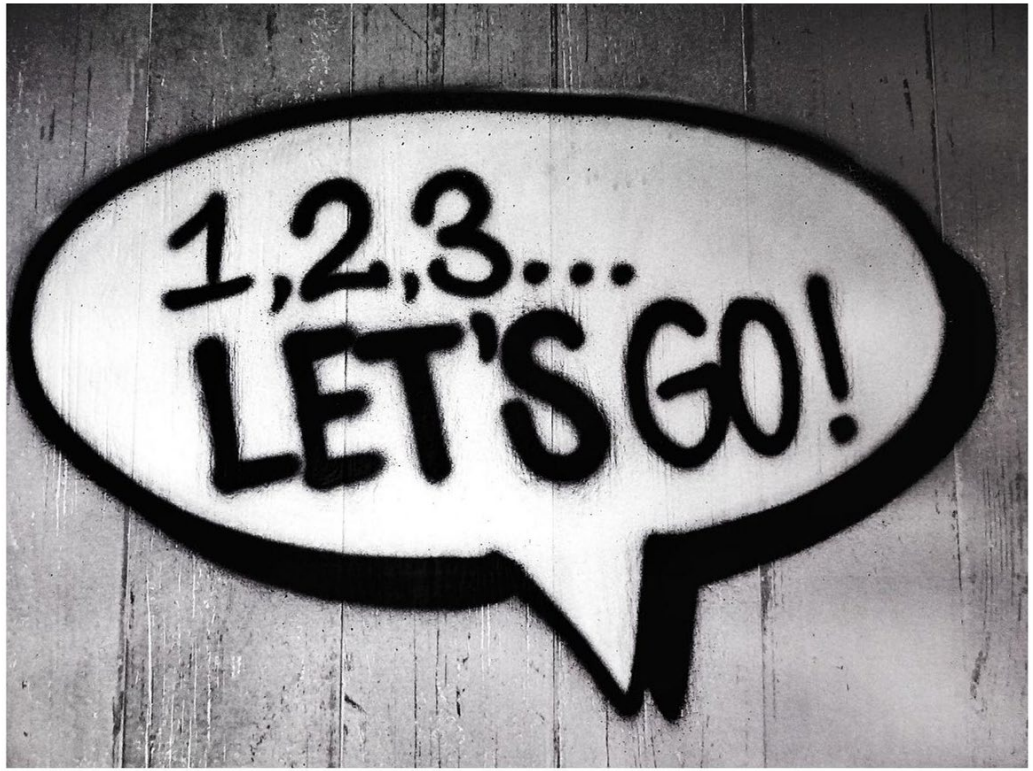
Note: The study used to exemplify this graphical overview is based on Anderson et. al (2021). Anchoring effect in judgements of objective fact and subjective preference. *Food Quality and Preference, 88*. DOI: https://doi.org/10.1016/j.foodqual.2020.104102

Block D - Creating a graphical overview of the research data process

Software recommendations

- *Open, free, Multi OS*: yEd (Desktop & Browser), draw.io (Browser), OpenOffice Draw
- *Commercial / Windows Only*: MS Office (Visio, PowerPoint)
- Filetypes
  - PDF or HTML export

Theory into practice



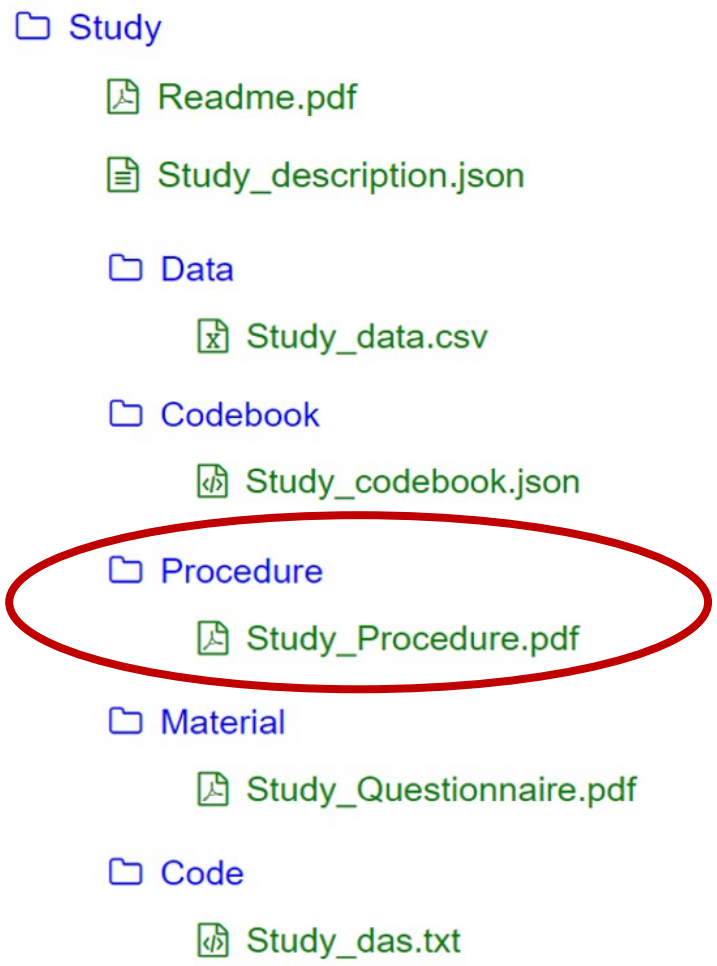
And now it's your turn!

https://app.diagrams.net/

https://www.yworks.com/yed-live/

Managing your data documentation with D-Psy-FAIR

D-Psy-FAIR structure



Evaluation & discussion

Agenda

Time (CET)	Session
13:00 - 13:15 h	Welcome & overview
13:15 - 13:35 h	**Introduction to the subject matter**
13:35 - 13:40 h	Technical break
13:40 - 14:25 h	**Block A: Writing a study description**
14:25 - 14:35 h	Coffee break
14:35 - 15:20 h	**Block B: Creating a codebook**
15:20 - 16:05 h	**Block C: Documenting the data preparation and analysis steps**
16:05 - 16:15 h	Coffee break
16:15 - 17:00 h	**Block D: Creating a graphical overview of the entire research data process**
17:00 - 17:30 h	**Discussion & concluding remarks**

Evaluation & discussion

Thank you for your participation and attention!

Contact

Do you have any questions?

<u>Please, contact:</u>
PsyCuraDat@leibniz-psychology.org

If you seek any further information, go to our project website!