Describing data

In order to understand your data easily, it is important to **document your data well**. The type of information included in the documentation and its format will depend on the type of data that you work with. In some cases, a simple read-me file might be sufficient, in others it might be appropriate to use tables or insert comments directly into the file (e.g., if you are describing a computer script).

If you are not sure what should be included in the **metadata** (= data about data), try asking yourself: “What would I need to know to interpret and use these data if I returned to it ten years from now?”

Documentation may include, for example:
- When and where the data were collected
- Information about the tool or version of the software used to collect or analyse the data
- Demographics of the participants, how the respondents were recruited
- List of variables / field names and their description (e.g., values they can take)
- Description of abbreviations, codebook
- Laboratory protocol
- Blank questionnaire
- Sample consent form
- License agreement
- …and a lot more

Especially if you decide to share your data, it is recommended you take into account subject specific guidelines for documenting data and use **metadata standards** and specific **ontologies** (= standardised dictionaries). This will ensure that others working in the same field are able to understand your data, and you decrease the chance that your data is misinterpreted. Moreover, if you use standardised description, it will be easier to combine data from different sources.

To find suitable metadata standards, you may use the FAIRsharing platform, the DCC list of metadata standards , or the Czech formal open standards .