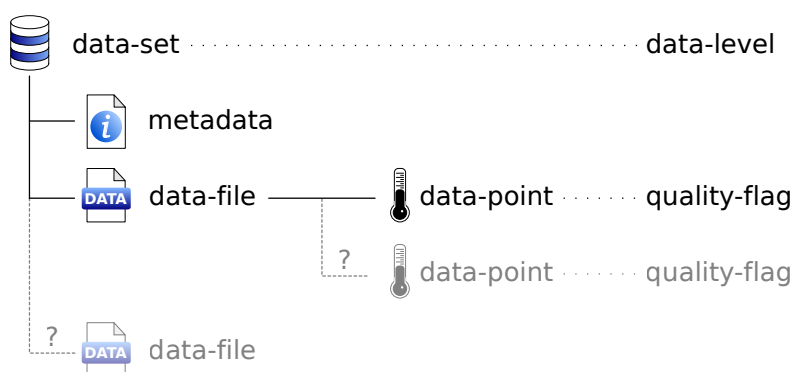


Data Scheme

Background & Motivation

Scientific data are often highly complex: heterogeneous data-sources, numerous processing steps, and challenging amounts of data are factors demanding for systematic data-management, especially when sharing data. To maximise the benefit for all users of the data, guidelines, principles, and rules are needed as outlined in a *Data Scheme*.

The *data-set* is the central entity: it consists of at least one *data-file* and corresponding *metadata*. A data-set can also comprise multiple data-files. To encode the overall state of a data-set it is assigned a *data-level* (see document on *Data-Levels* for details).



Each data-file contains at least one *data-point*. A data-point is a measurement (e.g. temperature-value) along with information for its identification or interpretation (e.g. sample-number, time & date, position). Additionally, each data-point has a *quality-flag*, which encodes for its validity (see document on *Data Quality-Flags* for details).

The metadata inform about how to interpret the actual data (e.g. parameter measured, unit, offset) and explains the *data-format*, i.e. the internal organisation of data-files (e.g. meaning of columns in a spreadsheet). Meta-data also comprises the method used to collect, check, and correct the data, such that corresponding method-sections in publications can be drafted. Further, it should inform about persons responsible for the data and might include historical notes about changes to the data.

Specifics on data-handling like

- data-format
- extend of metadata
- file-names / data-set identifiers
- folder hierarchy / database structure

are best chosen according to specific needs (see document on *Project Data-Guidelines* for details).