Handout "Data quality"

What is it about?

Data quality can consist of the following four dimensions and concerns the following aspects in each case:

- **Intrinsic** data quality (= concerns the data itself)
  - ➔ Credibility, accuracy, objectivity
- **Contextual** data quality (= purpose-dependent)
  - ➔ Added value, relevance, topicality, completeness, appropriate scope
- **representational** data quality (= representation-related)
  - ➔ Unambiguous interpretability, comprehensibility, uniform presentation, clarity
- **Access quality** (= system-related):
  - ➔ Workablilty, accessibility,

Why is this important?

The highest possible data quality must be ensured so that the design of scientific models or the der-ivation of scientific findings is sound and meaningful. In the case of software and scripts, it must also be ensured that they can be run on any computer.

However, it is not only about the data itself. Rather, the quality of the descriptive data (see Chapter 4 "Metadata and metadata standards") and the infrastructures (see Chapter 5 "FAIR principles and CARE principles") through which the data are made available, also play an important role.

How do I implement this?

The following measures to improve data quality can be included in your research process, for example:

- Implementation of a quality assurance concept (especially for larger projects)
- First-Time-Right principle: test routines during creation
- Uniformity and standardisation in data creation (incl. documentation of standards)
- Metadata creation
- Duplicate check
- Plausibility checks
- Peer review by colleagues (especially in the case of software and scripts, these should be run on other computers to check whether the dependencies are correct)