Handout “Metadata and Metadata Standards”

What is it about?

It’s about metadata, metadata standards and controlled vocabularies. Metadata are data about (research) data that describe the data in a structured way and are ideally readable for humans and computers. Metadata standards are schemas that define the ways in which data can be described by metadata via a standardised selection of descriptive categories. Controlled vocabularies are a defined collection of terms or labels that can be used within the categories specified by the metadata standard.

Why is this important?

Complete and correct metadata meet the requirements of good scientific practice. Metadata also make research data reusable for other researchers and can be found and linked in digital search systems. Moreover, well-curated research data enriched with a lot of subject-specific metadata often allow new research questions to be answered even after a project has ended.

Metadata standards are important so that researchers from the same discipline have a common repertoire of descriptive categories for their metadata, so that the data remain comparable.

Discipline-specific controlled vocabularies also ensure, looking at comparability, that the same terms are also used within these categories and, for example, that spelling errors are avoided, which could otherwise distort the results when the data are analysed by a computer.

How do I implement this?

In general, collecting metadata in a generic or subject-specific metadata schema and also linking it to a controlled vocabulary usually requires a higher level of expertise in dealing with these standards. If you’re not experienced, and do not have enough time to acquire knowledge about it, you should think about relevant categories in advance and simply save the metadata in a text file as a ReadMe. This could include the following:

- Author/researcher and other persons involved, e.g. by specifying the ORCiD ID.
- Object of research or title of the study or experiment
- Details of the experimental set-up (instruments used, settings, software & version used)
- Information on data provenance (e.g. use of already existing data sets)

If you are confident in working with subject-specific metadata schemas and controlled vocabularies, you should find out if there are any subject-specific metadata schemas and controlled vocabularies that you can use before collecting your data.