Handout “Data storage and archiving”

What is it about?

It is about data storage and data archiving measures. The former refers to measures that relate more to the short-term storage of data, i.e. while it is still being actively worked with, usually on a daily basis, and quick access is necessary. This includes storage on the software side, but also, for example, considerations about hardware security.
The latter refers to measures for long-term storage. In comparison to data storage, data archiving deals with the storage and availability of research data after the research work has been completed and in this respect, places demands on data formats and detailed documentation of metadata. A distinction is made between medium-term availability and so-called long-term archiving.

Why is this important?

Data storage measures, for example, help to prevent the risk of data loss. It is therefore important to know (automatic) backup strategies and apply them to one's own project data in order to also attend to the duty of documentation and verification in terms of good scientific practice. Many experiments or data can only be conducted or collected once, which is why a definitive loss of this data can then jeopardise the positive outcome of an entire project.
Data archiving is necessary to ensure the long-term availability of research data in accordance with FAIR principles, in order to be able to access, interpret, understand and re-use it. This also includes a critical examination of file formats and knowledge of the requirements of long-term archiving.

How do I implement this?

Data storage measures:

- Possible storage locations and media:
  - ➔ Institutional storage location: university cloud service, network drive
  - ➔ External storage location: Dropbox, Google Drive
  - ➔ Mobile storage medium: USB stick, external hard drive
  - ➔ Local workplace computer or own PC
- Clarified data access: for projects at least two persons with password knowledge
- Regular backups according to 3-2-1 rule: 3 copies on 2 storage media and 1 of them decentralised

Data archiving measures:

- Learn about and use suitable repositories
- Ask the operator of the selected archive or repository for suitable data formats for long-term archiving at an early stage
- Document the entire survey process (software, hardware, migration history)