

Handout "Datenorganisation"

Worum geht es?

Um Datenorganisation, einem Begriff, der im Forschungsdatenmanagement alle Strategien zur Strukturierung von Daten umfasst und demnach Teil eines strukturierten Arbeitens ist. Dies umfasst vor allem das Erstellen von sinnvollen und erweiterbaren Verzeichnisstrukturen, eine aussagekräftige Benennung auf Dateiebene und das Erarbeiten und Einhalten eines Versionierungskonzepts.



Warum ist das wichtig?

Je mehr Daten über einen bestimmten Projektzeitraum anfallen, desto wichtiger werden einheitliche Konzepte zur Datenorganisation und deren Dokumentation, um ein strukturiertes Arbeiten sicher zu stellen.



Eine gute Datenorganisation führt zu einer besseren Nachvollziehbarkeit und erlaubt auch noch nach Jahren sicher sagen zu können, was, wie und weshalb getan wurde. Hinzu kommt, dass in kollaborativen Projekten mit entsprechenden Konzepten die gemeinsame Arbeit an den Daten erleichtert wird, da jeder einer gleichen Benennungs- und Ablagestrategie folgt. Dadurch können Daten auf den Datenträgern schneller gesucht und gefunden werden.

Die Verwendung von Versionierungskonzepten erlaubt es außerdem, dass der aktuelle Stand einer Untersuchung oder eines bestimmten Experiments direkt eingesehen werden kann und alte Messungen als Vergleichswerte weiterhin vorhanden sind.

Wie setze ich das um?

Bei der **Dateibenennung** sollten Sie sich entweder am Camel Case (= ExperimentNummerEins) oder dem Pothole Case bzw. Snake Case (= experiment_nummer_eins) orientieren. Datumsformate sollten in den Formaten YYYY-MM-TT oder YYYYMMTT geschrieben werden, um eine klare Sortierung von alt nach neu zu gewährleisten. Außerdem müssen Sie festlegen, welche Informationen in den Dateinamen gehören und wie diese abgekürzt werden. Zuletzt sollten Sie für diese Konventionen eine ReadMe-Datei im entsprechenden Ordner anlegen.



Zur **Versionierung** sollten Sie sich am Major.Minor.Revision-Versionierungskonzept orientieren: experiment_v1_0_0. Der erste Wert gibt eine starke Veränderung an, der zweite eine kleinere und der dritte Wert beispielsweise eine Korrektur eines Rechtschreibfehlers. Entstehen an einem Tag aber nicht mehrere gleiche Messungen, reicht es oft auch aus, das Datum als Versionsindikator zu nutzen.

Bei **Verzeichnisstrukturen** sollten Sie darauf achten, dass das Konzept weder zu feingliedrig ist, noch, dass es überhaupt nicht erweiterbar ist. Auf der rechten Seite sehen Sie ein Beispiel für eine Verzeichnisstruktur, die als Grundlage genutzt werden kann.

