Dateibenennung und Ordnerstruktur an der TiHo – eine Empfehlung

Forschungsdatenmanagement

Dr. Max Hassenstein
forschungsdaten@tiho-hannover.de

1. Einleitung

Eine einheitliche, logische Ordnerstruktur und eindeutig benannte Dateien bzw. Dokumente können den Arbeitsalltag erheblich erleichtern. Hürden, die durch ein Dateichaos entstehen, werden von Beginn an reduziert. Im Idealfall wird im Vorfeld des Projektstarts oder zum Zeitpunkt der Datensammlung bzw. Anlage eines Projektordners eine Namenskonvention herangezogen und festgelegt.

Eine organisierte Ordnerstruktur vermeidet die Dateisuche nach dem Versuch und Irrtum-Prinzip und reduziert somit die Suchdauer. Gleiches Prinzip gilt für die Dateibenennung: Die gesuchte Datei kann direkt ausfindig gemacht werden und zudem über die globale Dateisuche gefunden werden.

Insbesondere für das gemeinsame Arbeiten mit dem Kollegium oder externen Personen kommt eine Ordnerstruktur und Dateibenennung zum Tragen. Auch sich selbst ermöglicht man das bessere Zurechtfinden in der Zukunft.

2. Ordnerstruktur

Eine gut organisierte Ordnerstruktur dient zweierlei: dem schnellen Wiederfinden von benötigten Dateien und Dokumenten und auch dessen logischer Trennung, sprich der Reduktion von digitalem Chaos. Diese Unordnung erschwert nicht nur Aspekte wie Datenzugriffsregeln (die Projektadministration benötigt ggf. keinen Zugriff auf sensible Originaldaten), sondern kann Unklarheiten im direkten Umgang mit Daten verursachen – wie die Frage nach der *richtigen* Tabelle mit den wertvollen Originaldaten.

Wir empfehlen eine Ordnerstruktur mit drei Ebenen (Abb. 1) als Ausgangsstruktur:

- Die Ebene 1 ist Ihr Projektordner, welcher sämtliche Daten und Dokumente des Projekts enthält. Daten bzw. Dokumente sollten möglichst nicht an anderer Stelle zusätzlich abgelegt werden, abgesehen von der Datensicherung. Diese gesammelte Ablage fördert eine eindeutige Organisation. Falls die Ablage in einem anderen Verzeichnis oder extern notwendig ist, dann sollte dies z. B. in der readme-Datei (s.3) dokumentiert werden. Selbst erzeugte oder von extern erhaltene Dateien sollten eindeutig benannt (s. 3) und entsprechend der Ordnerstruktur abgelegt werden.

- Ebene 2 ist das Arbeitsverzeichnis, von welchem in die entsprechenden Unterordner navigiert werden sollte, z. B. in die Administration oder den Auswertungs- und -analyseordner. Abgesehen von der readme-Datei sollten möglichst keine weiteren Dateien und Dokumente, außer ggf. zwingend notwendige, abgelegt werden, sondern direkt in die entsprechenden Unterordner einsortiert werden.
- Auf der Ebene 3 ist für den Auswertungs und –analyseordner eine weitere Ordnerstruktur vorgesehen, in welche jeweils Analysecode, abgeleitete Daten und schließlich auch Ergebnisdaten abgelegt werden können. Auf dieser Ebene können zusätzliche Ordner erforderlich werden, z. B. bei einer großen Anzahl unterschiedlicher Originaldaten oder kleinteiligen Projektberichten.

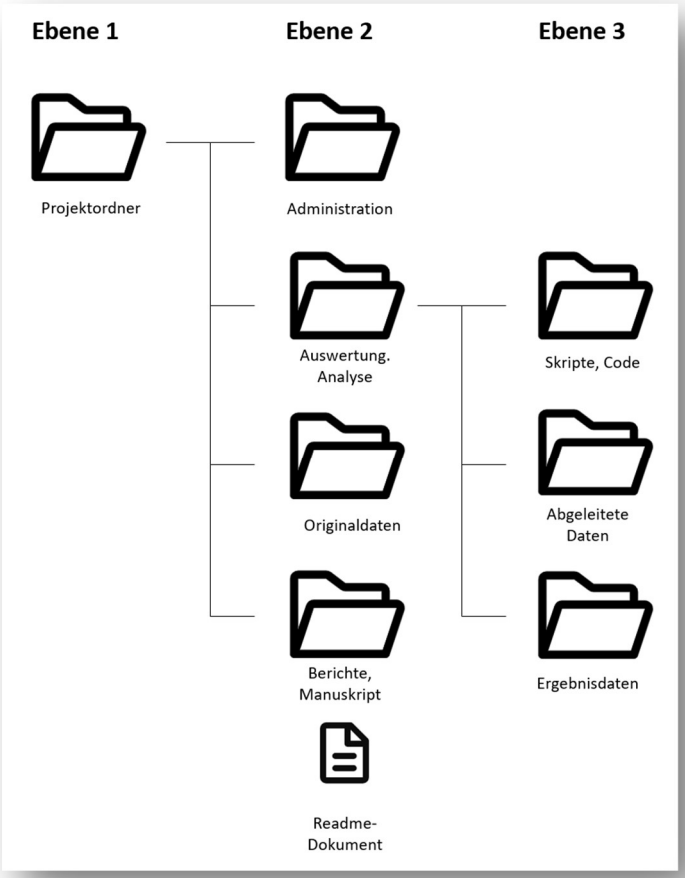

Abbildung 1: *Beispielhafte Ordnerstruktur für ein Forschungsprojekt*

Die in Abb. 1 vorgeschlagene Ordnerstruktur wird für viele Projekte gut funktionieren, in anderen Projekten werden zusätzliche Unterordner oder gar alternative Ordner erforderlich sein. Die Freiheit, diese Struktur anzupassen, liegt bei den Forschenden. Es ist allerdings zu beachten, dass eine starke Abweichung von einer grundlegenden Ordnerstruktur die Vorteile einer solchen zunichtemachen kann, wie z. B. das schnelle Zurechtfinden bedingt durch die Familiarität

mit der generischen Ordnerstruktur. Eine Ebene 4 ist nicht vorgesehen, kann aber je nach Projekt notwendig sein. Es ist abzuwägen, ob die Schaffung zusätzlicher Ebenen tatsächlich der Organisation zuträglich ist.

Grundsätzlich empfiehlt sich, zusätzliche Ordner erst dann anzulegen, wenn eine gewisse Anzahl an Dokumenten zusammengekommen ist, oder sich diese stark von anderen Abgrenzen. Als Beispiel sind hier Antragsdokumente zu nennen, die oftmals aus einem Antrag in PDF bestehen, gesonderten Anlagen und ggf. noch einer Word-Datei. Zeichnet sich im Projektverlauf ab, dass bestimmte Ordner nicht benötigt werden, können diese gelöscht werden.

2.1 Readme

Ein readme (dt. „lies mich“) soll als schnelle Orientierung dienen und Informationen zum Projekt enthalten, wie bspw. den Verantwortlichkeiten und den Inhalten, d. h. es stellt eine Kurzübersicht des Projektordners dar. Eine solche Datei sollte bestenfalls in jedem alleinstehenden Projektordner vorhanden sein. Dafür empfiehlt sich eine reine Textdatei (z. B. im TXT- oder RTF-Format), die eine Kompatibilität in unterschiedlichen Systemen und im Zeitverlauf gewährleisten. Ein readme kann zudem beschreiben, ob und welche ggf. technischen Voraussetzungen zur Interaktion notwendig sind (z. B. Zusatzsoftware zum Auslesen bestimmter Daten).

Sofern keine weitere Dokumentation zur Methodik der Forschungsdatengenerierung enthalten ist, z. B. in Form eines Manuskripts, sollte im Readme beschreiben werden, auf welche Weise die enthaltenen Daten erzeugt und gesammelt wurden. Im Falle einer Datennachnutzung sollte konkret die Quelle der Daten beschrieben werden, möglichst auch mit einer gesonderten Beschreibung dessen Erhebungs- oder Erzeugungsmethodik.

3. Dateibenennung

Eine eindeutige Dateibenennung unterstützt das schnelle Auffinden des benötigten Dokuments. Eine strukturelle Benennung (beispielhaft in Tab. 1) ermöglicht es, wichtige Informationen über das Dokument direkt in den Dateinamen zu integrieren. Somit wird direkt z. B. der Autorenname oder die Versionsnummer deutlich. Als Beispiel: Der in Excel durch Anna Schneider am 30.06.2021 angelegte Projektplan des Projekts „Wildtierbeobachtungsmeldesystem (WMS)“ in der Version 1 mit integrierten Kommentaren des Kollegen Peters wird „Projektplan_WMS_Schneider_v1_Peters.xlsx“ benannt (s. Tab. 1).

Tabelle 1: *Beispielhafte Konvention zur Benennung von Dokumenten*

Dokumentbezeichnung	Projekt	Autor	Datum	Version	Review
Abschlussbericht	Borreliose in Bayern (BIB)	Emily Harris	17.06.2024	-	A. Khan
➔ Abschlussbericht_BIB_Harris_2024-06-17_Khan.docx					
Kick-off-Präsentation	Shar-Pei Gendefekt-studie (SPG)	Olivia Bennett	01.10.2006	2	-
➔ Kickoff_SPG_Bennett_2006-10-01_v02.pptx					
Projektplan	Wildtierbeobachtungs- meldesystem (WMS)	Anna Schneider	30.06.2021	1	M. Peters
➔ Projektplan_WMS_Schneider_v01_Peters.xlsx					

Dateien und Dokumente sollten ausschließlich anhand alphanumerischer Zeichen benannt werden. Die Trennung der Dateinamensbestandteile kann durch einen Unterstrich (_) oder Bindestrich (-) erfolgen. Bindestriche können bevorzugt bei der Trennung einzelner Namensbestandteile wie z. B. dem Datum „2022-11-21“ (gemäß ISO 8601) dienen. Weitere Sonderzeichen (!, @, ~,*) und Umlaute und möglichst auch Leerzeichen sollten zwecks einer verbesserten Maschinenlesbarkeit vermieden werden.

Sofern mit der vorgeschlagenen Ordnerstruktur auch sonstige interne Dateiverzeichnisse verwaltet werden sollen, kann anstelle des Projektkürzels eine alternative, passende Abkürzung (z. B. „Orga“) gewählt werden.

Die Länge der Benennung von Datei- und Dokumenten sollte so kurz wie möglich, aber so lang wie nötig sein. In einem typischen Fenster eines Betriebssystems ist die Spalte *Name* mit der Dateibenennung nur begrenzt breit und weitere Zeichen werden versteckt. Als oberer Grenzwert zwecks Darstellung werden 100 Zeichen empfohlen, aus technischen Gründen sollten aber nicht mehr als 250 Zeichen genutzt werden.

3.1 Versionierung

Zur Versionierung von Dokumenten und Daten kann eine Versionsnummer (z. B. „v01“, „V02“, oder „Vers03“) verwendet werden. Alternativ kann eine Datumsangabe genutzt werden, um den aktuellen Stand der Datei festzuhalten, oder auch eine Kombination aus Versionsnummer und Datum für noch mehr Genauigkeit (s. Tab. 1). Die Wahl des Vorgehens richtet sich an dieser Stelle nach persönlicher Präferenz oder zusätzlich der Frequenz der anfallenden neuen Versionen.

Berichte, Anträge oder Manuskripte unterlaufen typischerweise Revisionsrunden mit mehreren Beteiligten. Eine Versionierung kann in diesem Kontext folgendermaßen gestaltet werden: Die erste Version wird „Abschlussbericht_BIB_Harris_2024-06-18.docx“ benannt und enthält im Revisionsprozess das Kürzel der Mitarbeitenden, z. B. „Abschlussbericht_BIB_Harris_2024-06-18_Khan.docx“. Sofern die Revisionen eingearbeitet wurden, kann eine neue Version mit aktuellem Datum und ohne das Kürzel der Revisoren versehen werden: „Abschlussbericht_BIB_Harris_2024-06-20.docx“.

Bei der fortlaufenden zusätzlichen Abspeicherung der aktuellen Version können sich im entsprechenden Verzeichnis vielzählige veraltete Versionen ansammeln. Diese können zeitweise für eine verbesserte Sichtbarkeit in einen neu angelegten Order „Archiv“ verschoben werden. Der Vorteil der Zwischenspeicherung ist, dass sich mögliche aus Versehen unbemerkt gelöschte Textabschnitte oder Vorversionen bestimmter Formatierungen problemlos und eigenständig wiederherstellen lassen. Sofern eine finale Version eines Textdokuments erreicht wird, sollten die zwischengesicherten Vorversionen zur Übersichtlichkeit und Datensparsamkeit gelöscht werden.