

Daten dokumentieren

README & Datenbeschreibungsverzeichnis (Data Dictionary)

FDM-Campus Hannover | 09.07.2026



Sofern nicht anders angegeben.
<https://creativecommons.org/licenses/by/4.0/>

Dr. Max J. Hassenstein

 <https://orcid.org/0000-0002-5977-388X>



FDM-CAMPUS HANNOVER

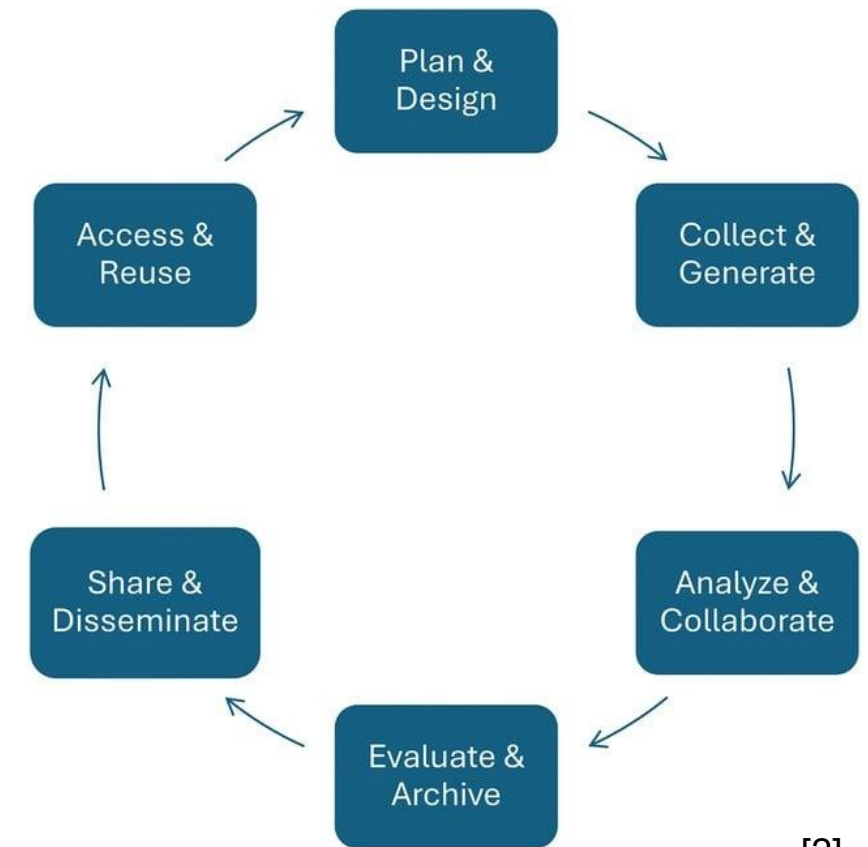


Agenda

- **Das Problem der „Daten-Hieroglyphen“:** Warum Daten ohne Dokumentation (oft) wertlos werden
- **Die README-Datei:** Aufbau und Inhalt eines effektiven Begleitdokuments
- **Das Datenbeschreibungsverzeichnis (Data Dictionary):** tabellarische Datensätze strukturiert beschreiben
- **Live-Demo:** Data Dictionary in Excel

Warum Forschungsdatenmanagement (FDM)?

- Forschungsdaten intrinsisch Wertvoll
- Datenbereitstellung von Förderern und vielen Journals verlangt
- FDM: Daten über den gesamten Lebenszyklus dokumentieren, speichern, teilen und nachnutzbar gestalten
- **FAIR-Prinzipien** [1]:
 - Findable: auffindbar über eindeutige PID, aussagekräftige Metadaten
 - Accessible: klar geregelter, dauerhafte Zugriffsmöglichkeit
 - Interoperable: offene Formate, standardisierte Vokabulare
 - Reusable: Dokumentation und Kontext, Lizenz



[2]

→Heutiger Fokus im FAIR-Kontext: README & Data Dictionary als Fundament für „R“

1. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
2. Hassenstein MJ, Jung K (2025) Ten simple rules for effective research data management. *PLOS Computational Biology* 21(12): e1013779. <https://doi.org/10.1371/journal.pcbi.1013779>

Das Problem der „Daten-Hieroglyphen“

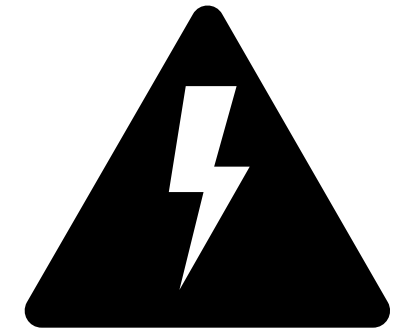
- Szenario: Sie erhalten eine Anfrage zu den Daten eines Projekts, aber die Projektleitung ist außer Haus
- Im ungeordneten Projektordner finden Sie „daten_final2.xlsx“
 - Ist das die richtige Datentabelle?
 - Welche Gruppen (grp) gibt es, und was ist t1?
 - Variablenausprägungen, Wertebereiche (uvm.) unklar

id	grp	sex	t1	med	dat
P001	1	2	12.4	0	03/14/24
P002	2	-99	18.9	9	2024-03-02
P003	1	1		1	4/07/24
P004	2	1	25.3	0	03/28/24

*Auszug aus
daten_final2.xlsx*

Das Problem der „Daten-Hieroglyphen“

- Kernproblem: Kontextwissen vergänglich
 - Oft nur im Kopf einer Person
 - Nimmt mit der Zeit ab (Personalwechsel, Projektende)
- Folgen fehlender Dokumentation:
 - Fehlinterpretation, mangelnde Nachvollziehbarkeit
 - Mehraufwand
 - Datenverlust



- Dokumentation steht im Einklang mit guter wissenschaftlicher Praxis!
- Nachnutzungspotenzial erhöht
- So dokumentieren, dass auch unbeteiligte Dritte die Daten nachvollziehen können

Zwei Ebenen der Dokumentation

README = Landkarte
Data Dictionary = Legende

README (allgemeines Begleitdokument)

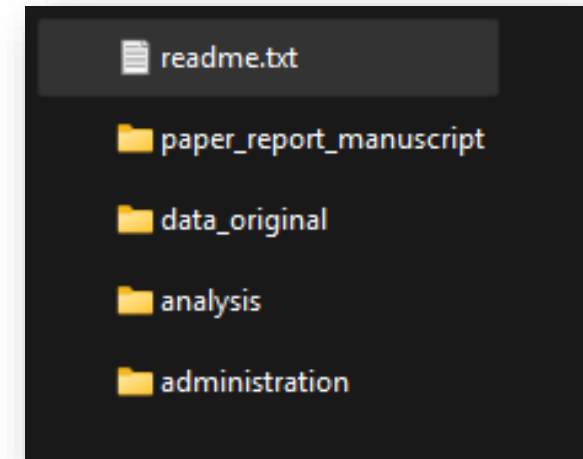
- Kontext auf Projekt- und Datensatzebene
- Info zum Projekt allgemein, Methoden, Dateiübersicht, Nutzungsbedingungen
- Freitext
- Üblicherweise eine Readme pro Projektverzeichnis

Data Dictionary (Datenbeschreibungsverzeichnis)

- Struktur auf Variablenebene
- Kontext auf Datenebene: Variablennamen, Labels, Datentypen, Einheiten, Wertebereiche, Kodierung, NA-Werte
- Tabellarisch
- Üblicherweise ein Data Dictionary pro Datentabelle

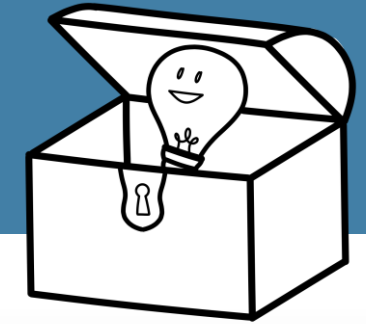
Die README-Datei: Zweck und Form

- Erste Anlaufstelle auf oberste Ordner Ebene
- Reines Textformat mit eindeutiger Benennung (z. B. README.txt oder LIESMICH.md)
 - Ohne Spezialsoftware lesbar
 - Plattformunabhängig
 - Langfristig unterstützt

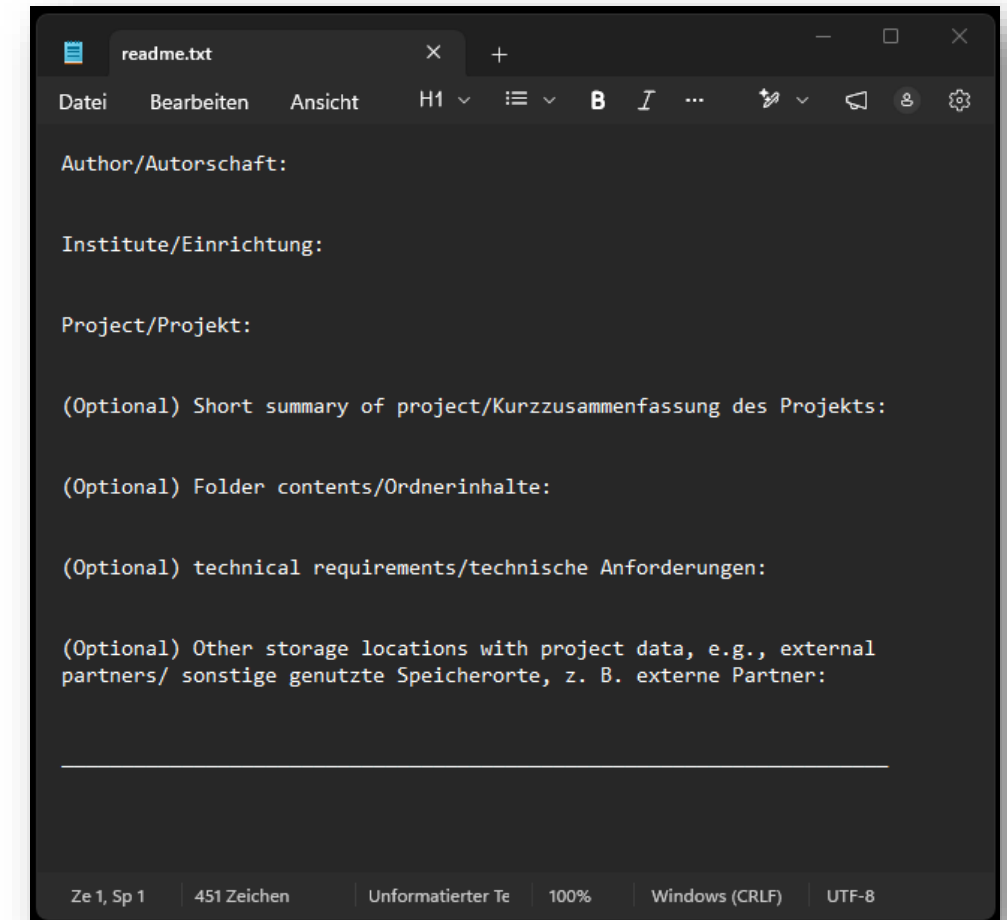


→ Lieber eine knappe, gepflegte README als eine perfekte, die nie entsteht

README: Was gehört hinein?



- **Allgemeines und Ordnerkontext**
 - Wer, wo, wozu?
- **Zusammenfassung und Methodik**
 - Projekt/Methodik und Ordnerinhalte
 - Technische Anforderungen: Geräte, Software
- **Schemata/Strukturen**
 - Kurzbeschreibung Daten, Beziehungen, Namenskonventionen



Das Data Dictionary: Was gehört hinein?

- Datensatzbezogen, pro Variable eine Zeile
 - Variablenname und Label/Bedeutung
 - Datentyp (Zeichenketten, Ganzzahlen, Gleitkommazahlen, Datum/Uhrzeit)
 - Wertebereich bzw. gültige Werte und Kodierungen (z. B. 2 = weiblich)
 - Einheit
 - Missing-Deklaration (z. B. NA = fehlend, -99 = keine Angabe)
- Format: eigene Excel-Tabelle oder separate CSV (maschinenlesbar)
- Leitfäden: MOSAIC [1] (Universitätsmedizin Greifswald), Buchanan et al. 2021 [2]

→ Kann eine dritte Person die Originaltabelle nachvollziehen?

1. Bialke M, et al. (2015) MOSAIC. A modular approach to data management in epidemiological studies. METHODS OF INFORMATION IN MEDICINE.54(4):364-371. <http://dx.doi.org/10.3414/ME14-01-0133>
2. Buchanan et al. (2021) Getting Started Creating Data Dictionaries: How to Create a Shareable Data Set. Advances in Methods and Practices in Psychological Science. 4(1). <https://doi.org/10.1177/25152459209280>

Live-Demo: Data Dictionary mit Excel erstellen

- Szenario: Sie sind die Projektleitung, die außer Haus war, und möchten nun das Data Dictionary nacharbeiten – für „daten_final2.xlsx“

Take-Home: Die 3 goldenen Regeln der Dokumentation

1. **Sofort & fortlaufend:** Während der Arbeit dokumentieren, nicht erst am Projektende, da Kontextwissen schnell verfällt.
2. **Zwei Ebenen:** Eine README für den Kontext, ein Data Dictionary für jede Datentabelle. Gerne vorlagen nutzen.
3. **Verständlichkeit:** Für Fachfremde und das „Zukunfts-ich“ schreiben: eindeutige Namen, Kodierungen, Einheiten, NA-Deklaration.

Ihre Ansprechperson vor Ort



Dr. Erik Tute

Forschungsdatenmanagement Datenintegrationszentrum
Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1
30625 Hannover
E-Mail: fdm@mh-hannover.de



Dr. Max J. Hassenstein

Referent für Forschungsdatenmanagement
AG "Genomics and Bioinformatics of Infectious Diseases"
Bünteweg 17p
30559 Hannover
E-Mail: forschungsdaten@tiho-hannover.de



Sebastian Lehmann

Stabsabteilung Forschung, Entwicklung und Transfer
Bereich Forschungsdatenmanagement
Expo Plaza 4
30539 Hannover
E-Mail: fdm@hs-hannover.de



Dr. Samuel Mund

Hochschulbibliothek
Hochschule für Musik, Theater und Medien Hannover
Leisewitzstraße 39
30175 Hannover
Loebensteinstraße 2
E-Mail: fdm@hmtm-hannover.de

Vielen Dank – Fragen?

Gefördert durch:

Der Aufbau von Basisstrukturen im Forschungsdatenmanagement an den beteiligten Hochschulen wird im Rahmen der Maßnahmensäule 2 der Landesinitiative FDM Niedersachsen durch das Förderprogramm zukunft.niedersachsen und Hochschule.digital Niedersachsen gefördert.

zukunft.
niedersachsen

H  **N** **HOCHSCHULE.DIGITAL**
Niedersachsen



FDM-CAMPUS HANNOVER

