



COFFEE LECTURE

Datendokumentation in a nutshell

Sofern nicht anders angegeben, stehen die Inhalte dieser Präsentation unter einer unter CC0 1.0 Universell (CC0 1.0) Public Domain Dedication (<https://creativecommons.org/publicdomain/zero/1.0/deed.de>).

EINSTIEG

Gehe auf www.menti.com und benutze den Code 69 71 52 82

5 Thesen zur Datendokumentation

1. Mit der Zeit werden Informationen vergessen.
2. Dokumentation ist wesentlicher Bestandteil der guten wissenschaftlichen Praxis.
3. Datendokumentation ist lästig und sinnlos.
4. Dokumentation erleichtert die Zusammenarbeit.
5. Ohne Dokumentation sind Forschungsdaten nichts wert.

UNBESCHRIEBENES BLATT

Research Data ScaryTale No 31



E. Böker / [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

Die Daten existierten, konnten aber trotz größter Bemühungen nicht nachgenutzt werden.

GUTE WISSENSCHAFTLICHE PRAXIS

Leitlinie 12: Dokumentation

„Wissenschaftlerinnen und Wissenschaftler dokumentieren alle für das Zustandekommen eines Forschungsergebnisses relevanten Informationen so nachvollziehbar, wie dies im betroffenen Fachgebiet erforderlich und angemessen ist, um das Ergebnis **überprüfen** und **bewerten** zu können. [...]“

Quelle: Deutsche Forschungsgemeinschaft. (2019). Leitlinien zur guten wissenschaftlichen Praxis. Kodex. <http://doi.org/10.5281/zenodo.3923602>, S. 17.
Lizenziert unter Creative Commons Attribution-ShareAlike International 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/legalcode>).

FAIR DATA PRINZIPIEN

Auffindbar
(Findable)



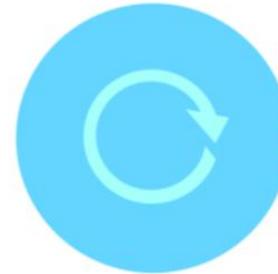
Zugänglich
(Accessible)



Interoperabel
(Interoperable)



Wiederverwendbar
(Reusable)



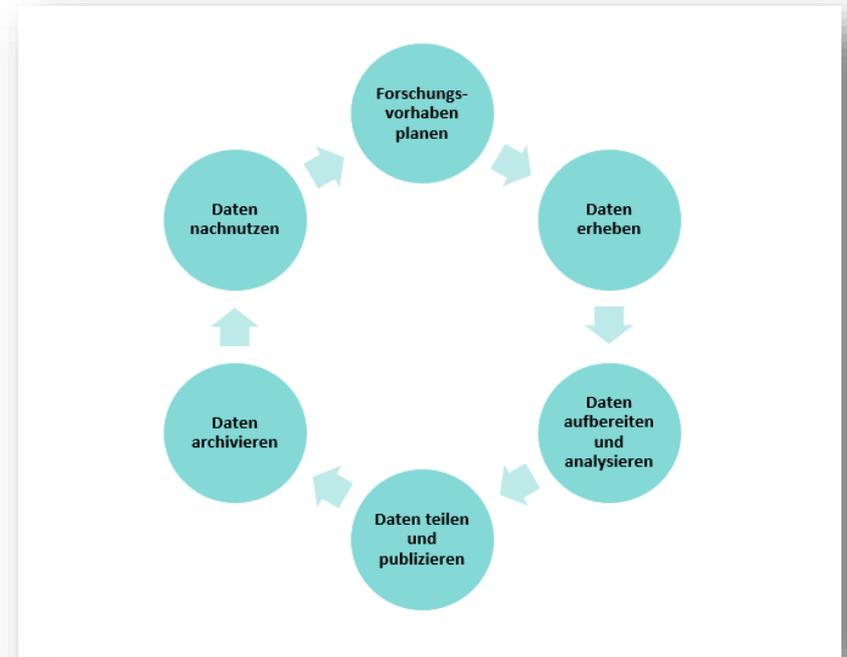
Graphik: [Paulina Halina Sieminska](#) / Bearbeitung: Dr. Ilona Lang / [CC BY-SA 4.0](#)

WANN UND WAS?

Den gesamten Prozess dokumentieren

Relevante Informationen/Metadaten identifizieren

Kontextinformationen gemeinsam mit den Daten speichern



Graphik: forschungsdaten.info (Abruf 15.02.2022)

INHALTE EINER DOKUMENTATION

- Für welchen Zweck wurden die Daten erstellt?
- Was beinhaltet der Datensatz?
- Wie wurden die Daten erhoben?
- Wer hat die Daten erhoben?
- Wann wurden die Daten erhoben?
- Wie wurden die Daten verarbeitet?
- Welche Datenbereinigungsprozesse wurden durchgeführt?
- Wie wurde die Qualität der Daten sichergestellt?
- In welchen Formaten stehen die Daten zur Verfügung?
- Wie kann man auf die Daten zugreifen?

DOKUMENTATIONSFORMEN

- README Files
- Data Dictionaries
- Codebooks
- Als Teil des DMP
- Electronic Lab Notebooks
- Artikel in einem Data Journal

→ Große Bandbreite

README FILE

- Menschenlesbar
- Durchsuchbar
- Metadaten
unstrukturiert, aber
besser als keine
Metadaten
- Am besten im .txt.
oder .md Format

```

Readme Example

Creator:
Maxi Mustermann, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin

Reason(s) for data analysis:
The data was collected within project XYZ which investigates, why students don't document their data.

Creation date of file(s): 11-2019

Used method(s): Survey among 500 students

Used Software (incl. version and add-ons) and tools:
- R x64 version 3.2.3

Data:
- Project-XYZ-survey.csv
  - Anonymized dataset of the survey
  - Outliers were deleted

Code:
- Project-XYZ-survey.r
  - R code for data analysis and creation of fig. 1-3

Additional files:
- Project-XYZ-survey-questionnaire.pdf
- Project-XYZ-survey-variable-description.pdf

License:
All data, files and code are licensed under a Creative Commons At-tribution 4.0 International (CC BY 4.0)
license (https://creativecommons.org/licenses/by/4.0/).

Notes:
- Non-anonymized data is available from the creator for secondary analysis.
  
```

DATA DICTIONARY & CODEBOOK

File name	Data Type	Method	Creator	Date	Description	Rights	Long-term availability

Show rows with cells including:

Variable	Variable name	Mesaurement unit	Allowed values	Description
Participant ID number	ID	Numeric	001-999	ID number assigned to participant in sequential order
Group number	GROUP	Numeric	1-30	Group assigned to participant based on ID number
Age in years	AGE	Numeric	18.0-65.0	Age of participant in years
Date of birth	DOB	mm/dd/yyyy	1-12/1-31/1951-1998	Participant's date of birth
Gender	SEX	Numeric	1 = male 2 = female	Participant's gender
Date of survey	SURVEY	mm/dd/yyyy	01/01/2015 – 01/01/2016	When the participant completed the survey
Self-reported consumer spending	SPEND	Numeric	0-100,000,000	Self-reported average yearly expenditure
Market sentiment	SENTIMENT	Numeric	1 = negative 2 = neutral 3 = positive	Sentiment towards US domestic economy
Actual GDP growth	GDP	Numeric	-5.0-5.0	Average US yearly GDP growth

- Menschenlesbar
- Durchsuchbar
- Metadaten strukturiert(er)

Quelle: Bowman, S. How to Make a Data Dictionary. Online verfügbar: <https://help.osf.io/hc/en-us/articles/360019739054-How-to-Make-a-Data-Dictionary> (Public Domain).

DATA JOURNALS

- Forschungsdaten werden in einem Artikel dokumentiert und beschrieben
- Inhalte: Datenerhebung, Charakteristiken, Funktionen und Nachnutzungsmöglichkeiten
- Peer Review und Zitierbarkeit

- Beispiele:
 - Scientific Data
<http://www.nature.com/sdata>
 - Data in Brief
<http://www.journals.elsevier.com/data-in-brief>
 - Data
<http://www.mdpi.com/journal/data>

WIE DOKUMENTIEREN?

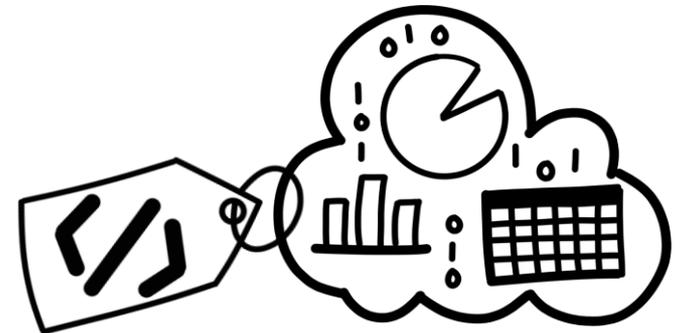
- Standardisiert & offen (Formate, Vokabularien, Sprachen)
- Mensch- und maschinenlesbar (strukturierte Metadaten)
- Eindeutig zuzuordnen (persistente Identifikatoren)



E. Böker / [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

METADATEN: SCHEMATA UND STANDARDS

- Daten über Daten
- Je strukturierter die Information, desto besser für Maschinen lesbar und verarbeitbar
 - Festgelegte Metadatenschemata
 - (Fachspezifische) Metadatenstandards
 - Übersicht z.B. unter [RDA Metadata Standards Directory Working Group](#)
 - Generisch: Dublin Core oder [Datacite](#) Schemata
- Vorgaben existent?



KONTROLLIERTES VOKABULAR



Athree23 / [Pixabay Lizenz](#)

Standardisierung der
Inhalte von
Metadatenschemata

- z.B. Wortliste, Klassifikation, Thesaurus
- fördert die Konsistenz und Genauigkeit, löst Mehrdeutigkeiten auf
- ermöglichen es, Beziehungen zwischen Begriffen und Konzepten darzustellen
- Basel Register of Thesauri, Ontologies & Classifications
<http://www.bartoc.org/>

TAKE HOME MESSAGE

- ✓ Zeit investieren & (möglichst) früh starten
- ✓ Knappe Dokumentation besser als keine Dokumentation
- ✓ So einheitlich wie möglich, so flexibel wie nötig
- ✓ Metadaten: Fachstandards beachten

Love your data.

REFERENZEN

- Biernacka, Katarzyna, Helbig, Kerstin, Senft, Matthias, & Trautwein-Bruns, Ute. (2020, September 21). Datendokumentation leicht gemacht! Ein interaktiver Online-Workshop. Zenodo. <https://doi.org/10.5281/zenodo.4037151>
- DFG „Checkliste für einen angemessenen Umgang mit Forschungsdaten in DFG-Projekten“. Zugriff am 15.02.2022. https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf
- CESSDA Training Working Group. „CESSDA Data Management Expert Guide.“ Bergen: CESSDA ERIC, 2017-2018. Zugriff am 15.02.2022. <https://www.cessda.eu/DMGuide>
- Gerlach, Roman, Rex, Jessica, Lang, Kevin, Neute, Nadine, Annett Schröter, Schwartze, Volker, Assmann, Cora, & Lehmann, Anne. (2021). Research Data ScaryTales (2.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.5636273>
- Gerlach, Roman, Schröter, Annett, & Lang, Kevin. (2021, February 24). Coffee Lecture Slides: Datendokumentation. Zenodo. <https://doi.org/10.5281/zenodo.4569528>
- forschungsdaten.info. „Datendokumentation.“ Zugriff am 14.02.2022. <https://www.forschungsdaten.info/themen/beschreiben-und-dokumentieren/datendokumentation/>.

VIELEN DANK FÜR DIE AUFMERKSAMKEIT!

... UND MORGEN
12:00-12:30 UHR COFFEE LECTURE

PRÄREGISTRIERUNG HYPOTHESEN-BASIERTER STUDIEN
VORTRAGENDE: DANIELA MERTZEN, UNIVERSITÄT POTSDAM