



# DATENDOKUMENTATION

Katarzyna Biernacka



# AGENDA

- Wiederholung
- Einführung in die Datendokumentation
- Metadaten
- Kontrolliertes Vokabular und Normdaten
- Dokumentationsformen
- Anwendung

WIEDERHOLUNG



Go to [www.menti.com](http://www.menti.com) and use the code 37 07 40

 Mentimeter

## Wozu wird eine Datendokumentation benötigt?

Nachvollziehbarkeit verbessern  
bzw. überhaupt erst  
ermöglichen

Wiederfindbarkeit von FD  
Strukturierung von FD

Reproduzierbarkeit der  
Ergebnisse

Zur Nachnutzung der Daten

Nachvollziehen des eigenen  
Arbeitens/des Arbeitens  
anderer

Bessere Nachvollziehbarkeit

Leichtere Übergabe bei

um die FD verstaendlich zu  
machen

Zur Reproduzierbarkeit <(^.^<)

Als Schnittstelle zu Tools



Slide is not active

Activate

Pause scroll

 16

# GRUPPENARBEIT

Welche Informationen fehlen Ihnen um diese Daten zu verstehen?

Gruppe 1 & Gruppe 3

<http://bit.ly/2T7hRqI>

Gruppe 2 & Gruppe 4

Handout

# 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 Datenbereinigungsverfahren wurden durchgeführt?
- Wie wurde die Qualität der Daten sichergestellt?
- In welchen Formaten stehen die Daten zur Verfügung?
- Wie kann man auf die Daten zugreifen?

# DMP VS. DATENDOKUMENTATION

## DMP

- der Rahmen (z. B. Zeit, Ziel) und administrative Informationen zum Projekt
- Verantwortlichkeiten und Kontaktdaten
- Methoden
- Kosten
- externe Partner und Dienstleister
- genutzte Hardware und Software
- Datentypen und -formate
- Nachnutzung existierender Datensätze
- Metadaten
- Methoden zur Erstellung neuer Daten
- Datenmenge
- Datenspeicherung und -sicherung
- Datenstruktur
- Dokumentation
- Qualitätssicherung
- Datenaustausch
- Archivierung
- Zugriff und Nachnutzung

## Datendokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Detaillierte Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Versionen
- Metadaten und Metadatenstandards
- Informationen zum Zugang und Nutzungsbedingungen

# DMP VS. DATENDOKUMENTATION

## DMP

- der Rahmen (z. B. Zeit, Ziel) und administrative Informationen zum Projekt
- Verantwortlichkeiten und Kontaktdaten
- Methoden
- Kosten
- externe Partner und Dienstleister
- genutzte Hardware und Software
- Datentypen und -formate
- Nachnutzung existierender Datensätze
- [Metadaten](#)
- Methoden zur Erstellung neuer Daten
- Datenmenge
- Datenspeicherung und -sicherung
- Datenstruktur
- Dokumentation
- Qualitätssicherung
- Datenaustausch
- Archivierung
- Zugriff und Nachnutzung

## Datendokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Detaillierte Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Versionen
- [Metadaten und Metadatenstandards](#)
- Informationen zum Zugang und Nutzungsbedingungen



# DMP VS. DATENDOKUMENTATION

## DMP

- der Rahmen (z. B. Zeit, Ziel) und administrative Informationen zum Projekt
- Verantwortlichkeiten und Kontaktdaten
- Methoden
- Kosten
- externe Partner und Dienstleister
- genutzte Hardware und Software
- Datentypen und -formate
- Nachnutzung existierender Datensätze
- Metadaten
- Methoden zur Erstellung neuer Daten
- Datenmenge
- Datenspeicherung und -sicherung
- Datenstruktur
- Dokumentation
- Qualitätssicherung
- Datenaustausch
- Archivierung
- Zugriff und Nachnutzung

## Datendokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Detaillierte Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Versionen
- Metadaten und Metadatenstandards
- Informationen zum Zugang und Nutzungsbedingungen

# DMP VS. DATENDOKUMENTATION

## DMP

- der Rahmen (z. B. Zeit, Ziel) und administrative Informationen zum Projekt
- Verantwortlichkeiten und Kontaktdaten
- Methoden
- Kosten
- externe Partner und Dienstleister
- genutzte Hardware und Software
- Datentypen und -formate
- Nachnutzung existierender Datensätze
- Metadaten
- Methoden zur Erstellung neuer Daten
- Datenmenge
- Datenspeicherung und -sicherung
- Datenstruktur
- Dokumentation
- Qualitätssicherung
- Datenaustausch
- Archivierung
- **Zugriff und Nachnutzung**

## Datendokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Detaillierte Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Versionen
- Metadaten und Metadatenstandards
- **Informationen zum Zugang und Nutzungsbedingungen**

# DMP VS. DATENDOKUMENTATION

## DMP

- der Rahmen (z. B. Zeit, Ziel) und administrative Informationen zum Projekt
- Verantwortlichkeiten und Kontaktdaten
- Methoden
- Kosten
- externe Partner und Dienstleister
- genutzte Hardware und Software
- Datentypen und -formate
- Nachnutzung existierender Datensätze
- Metadaten
- Methoden zur Erstellung neuer Daten
- Datenmenge
- Datenspeicherung und -sicherung
- Datenstruktur
- Dokumentation
- Qualitätssicherung
- Datenaustausch
- Archivierung
- Zugriff und Nachnutzung

## Datendokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Detaillierte Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Versionen
- Metadaten und Metadatenstandards
- Informationen zum Zugang und Nutzungsbedingungen

# METADATEN – DATEN ÜBER DATEN

- Beschreibende Metadaten:

- Titel
- Beschreibung
- Autor
- Urheberrechts-Inhaber
- Kontaktdaten
- Lizenzangaben
- Schlüsselwörter

- Technische Metadaten:

- Aufnahme datum
- Brennweite
- Blende
- Belichtungsdauer
- Geographische Koordinaten
- und viele weitere

# METADATENSTANDARDS

Eine Übersicht zu disziplinspezifischen und fachübergreifenden Metadatenstandards gibt es unter:

<http://www.dcc.ac.uk/resources/metadata-standards>

<http://rd-alliance.github.io/metadata-directory/subjects>

Beispiele:

Fachübergreifende Metadatenstandards:

Dublin Core

MARC21

Geowissenschaften:

ISO 19115

Biodiversität:

Darwin Core

Geisteswissenschaften:

Text Encoding Initiative (TEI)

Naturwissenschaften:

ICAT Schema

Cristallographic Information Framework

Sozialwissenschaften:

Data Documentation Initiative

# METADATENSTANDARDS - BEISPIELE

## Dublin Core (15 Kernelemente)

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

## Darwin Core (Erweiterung von Dublin Core)

- EventID
- ParentEventID
- SamplingProtocol
- SampleSizeValue
- SampleSizeUnit
- OrganismQuantity
- OrganismQuantityType
- TaxonID

# NORMDATEN

Dienen der eindeutigen Identifikation von Personen, Institutionen, Forschungsförderern, z.B.

- Gemeinsame Normdatei (GND)
- International Standard Name Identifier (ISNI)
- Virtual International Authority File (VIAF)
- Open Funder Registry

# KONTROLLIERTES VOKABULAR

## Klassifikation

dient der Zuordnung von Objekten in (meist hierarchisch strukturierten) Klassen. Die Klassen sind durch bestimmte Merkmale charakterisiert.

## Thesaurus

ist eine natürlich-sprachliche geordnete Sammlung von Begriffen und deren Beziehungen zueinander.

Entsprechende Thesauri und Klassifikationen können im Basel Register of Thesauri, Ontologies & Classifications gefunden werden: <http://www.bartoc.org/>





# ÜBUNG

Gehen Sie zu [bartoc.org](http://bartoc.org) und finden Sie ein kontrolliertes Vokabular, welches für Ihre Daten geeignet wäre.



# DOKUMENTATIONSFORMEN

- README Files
- Data Dictionaries
- Codebooks
- Electronic Lab Notebooks

# DOKUMENTATIONSFORMEN

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

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> [2019-01-21]



# ÜBUNG

Schreiben Sie eine Dokumentation für Ihre Daten.

*Falls keine eigenen Daten verfügbar:*

Schreiben Sie eine Anleitung für den bereitgestellten  
Gegenstand.



# ZUSAMMENFASSUNG

1. Keine Panik!
2. Beginnen Sie so früh wie möglich.
3. Schreiben Sie einfach Ihre Forschungspraktiken auf.
4. Behalten Sie im Hinterkopf: was wird benötigt um diese Daten zu verstehen?

# KATARZYNA BIERNACKA

Mail: [biernack@hu-berlin.de](mailto:biernack@hu-berlin.de)

ORCID: 0000-0002-6363-0064

Twitter: @ICTKasia

SO FERN NICHT ANDERS ANGE GEBEN, SIND DIE INHALTE AUF DIESEN FOLIEN UNTER EINER CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LIZENZ (CC-BY-4.0) LIZENZIERT.



WEITERE INFORMATIONEN:

DOLZYCKA, DOMINIKA, BIERNACKA, KATARZYNA, HELBIG, KERSTIN, & BUCHHOLZ, PETRA. (2019). TRAIN-THE-TRAINER KONZEPT ZUM THEMA FORSCHUNGSDATENMANAGEMENT (VERSION 2.0). ZENODO. [HTTP://DOI.ORG/10.5281/ZENODO.2581292](http://doi.org/10.5281/zenodo.2581292)