GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

VerbundFDB
forschungsdaten-bildung.de

Organisation und Speicherung von Forschungsdaten

Dr. Marcus Eisentraut

Timo Hutflesz

GESIS - Leibniz-Institut für Sozialwissenschaften

Online-Seminar

DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation

gesis Leibniz-Institut für Sozialwissenschaften

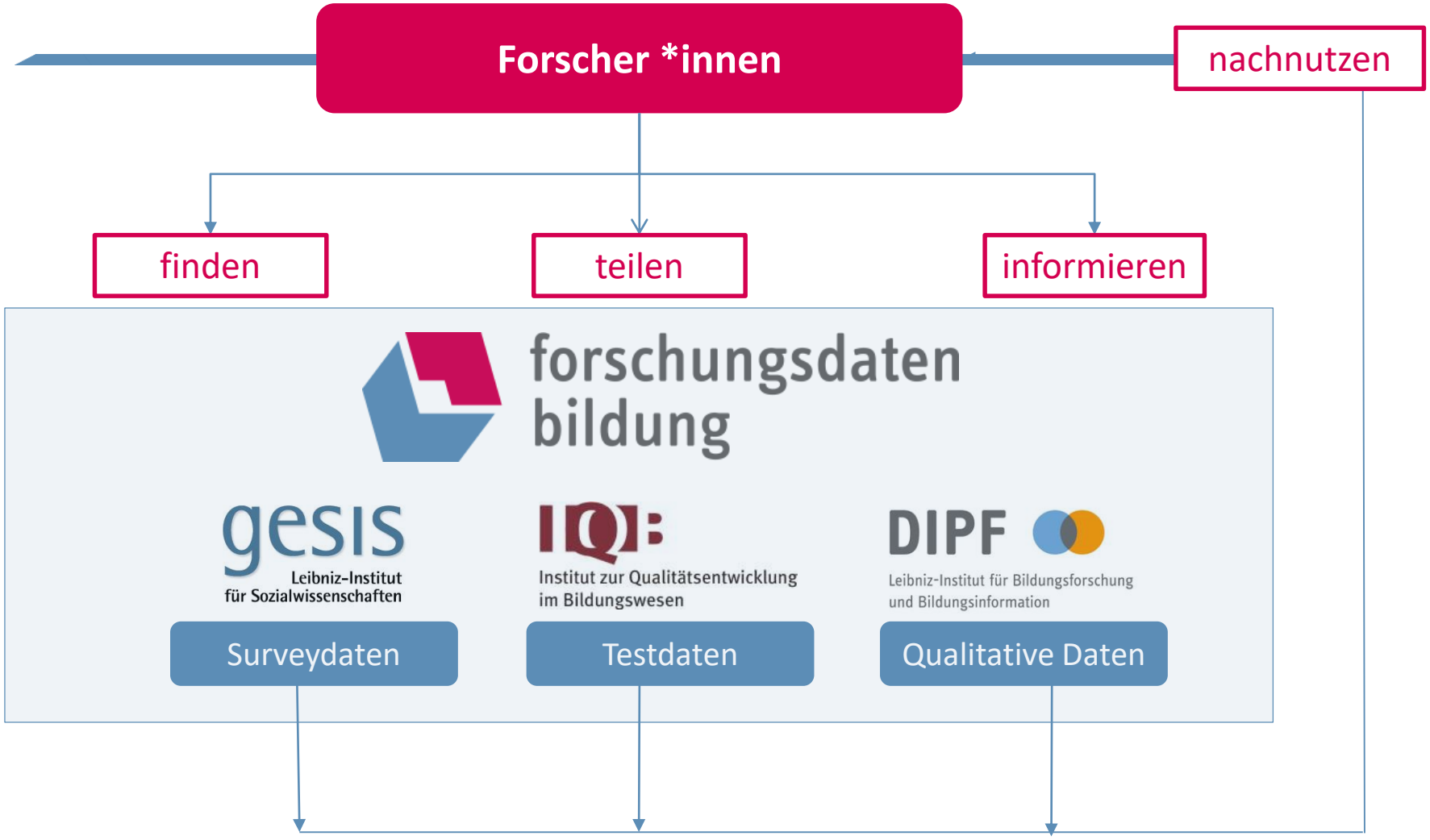
IQB Institut zur Qualitätsentwicklung im Bildungswesen

CC BY

Vorstellung

Dr. Marcus Eisentraut

Timo Hutflesz


Forscher *innen
nachnutzen
finden
teilen
informieren
forschungsdaten bildung
gesis
Leibniz-Institut für Sozialwissenschaften
IQB
Institut zur Qualitätsentwicklung im Bildungswesen
DIPF
Leibniz-Institut für Bildungsforschung und Bildungsinformation
Surveydaten
Testdaten
Qualitative Daten

Online-Seminare zu unterschiedlichen Themen:
https://www.youtube.com/channel/UChIMIqtkZDEXKQsJ9SVD--w

VerbundFDB
forschungsdaten-bildung.de

Forschungsdaten Bildung

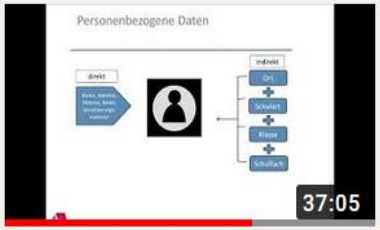
52 Abonnenten

KANAL ANPASSEN | VIDEOS VERWALTEN

ÜBERSICHT | VIDEOS | PLAYLISTS | KANÄLE | DISKUSSION | KANALINFO

Uploads ALLE WIEDERGEBEN

SORTIEREN NACH



Webinar: Datenschutz in der Praxis - die informierte...

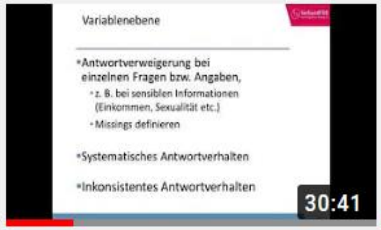
1186 Aufrufe • vor 2 Jahren

Webinar: Datendokumentation und...

698 Aufrufe • vor 2 Jahren

Webinar: Daten teilen - Wo fange ich an?

482 Aufrufe • vor 2 Jahren



Webinar: Datenaufbereitung und Anonymisierung

453 Aufrufe • vor 1 Jahr

Webinar: Lizenzierung und Urheberrecht

341 Aufrufe • vor 1 Jahr

Sekundärdaten finden und nutzen – das Angebot des...

243 Aufrufe • vor 7 Monaten

Aufbereitung und Anonymisierung qualitativer...

176 Aufrufe • vor 4 Monaten

Forschungsdatenmanagement während des Forschungsdatenlebenszyklus

Organisation und Speicherung von Forschungsdaten

Warum?

- Arbeitserleichterung
- Notwendig zum Arbeiten in Forschungskooperationen
- Gute wissenschaftliche Praxis
- Ggf. rechtliche Auflagen

☺ frühzeitige Planung (Projektplanungsphase)

cc-by http://hochstenbach.wordpress.com

Hochstenbach, P. (2015): FOSTER Cartoon: Backups. FOSTER-Project (EU-Funding 612425).
Verfügbar unter: https://www.fosteropenscience.eu/foster-taxonomy/research-data-management. (CC-by).

Inhalte der Sitzung

Bilder: www.pixabay.org (CC-0).

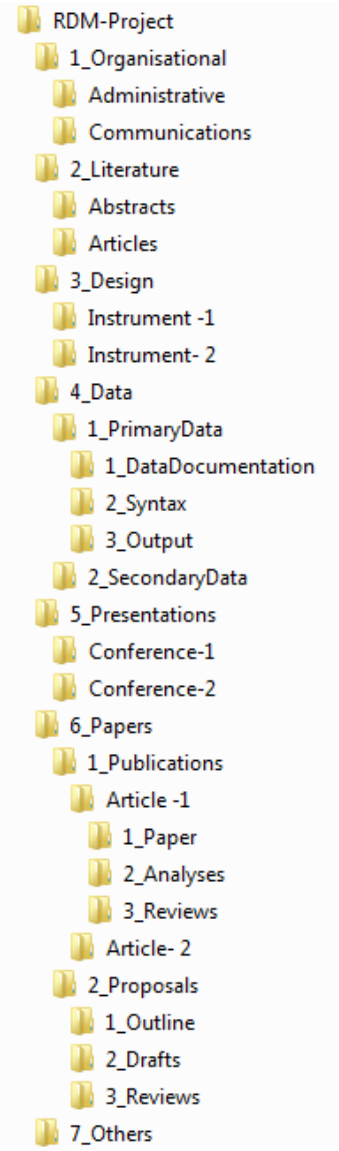
Organisation von Dateien

- Zumeist besitzen Forschende bereits entsprechende Verfahren
- Hierarchische Ordnerstruktur, z. B. anhand von Themenkomplexen, Datentypen, Prozessen etc.
- Ordnerbenennung sollte selbsterklärend sein

☺ Richtlinien frühzeitig entwickeln und konsistent umsetzen

☺ nicht zu viele Ordnerebenen

☺ evtl. readme-Datei mit einem Dateiverzeichnis des Projektordners



Dateinamen und Versionierung

- Information in Dateinamen
  - Projekttitel, Ersteller*in, Datum des Erstellens, Datentyp, Version der Datei etc.
  - aber sollte nicht zu lang sein und keine Sonder- bzw. Leerzeichen
- Versionierung zur Identifikation der aktuellen Arbeitsdatei
  - als Teil des Dateinamens, in der Kopfzeile der Datei, in Versionierungsdokument etc.
  - klare Versionierungsregeln

Datenversionierung nach DDI

(1.0.0)

major.minor.revision

☺ major (beginnt 1): große Veränderungen im Datensatz, z. B. Hinzuspielen neuer Variablen oder Beobachteter

☺ minor (beginnt mit 0): kleinere Veränderungen, z. B. Rekodieren einer Variable

☺ revision (beginnt mit 0): sehr kleine Veränderungen, z. B. Korrektur von Rechtschreibfehlern in den Variablenlabels

www.ddialliance. org.

Dateiformate

Datentyp	Bevorzugt	Akzeptiert
quantitative Daten	• SPSS Portable (*.por; *.sav) • STATA (*.dta) • R (*.R) • SAS Transport (*.sas) • Tabulator-, Komma- oder Spalten-getrennte Textdatei (“csv”)	• OpenDocument-Tabellendokument (*.ods) , MS Excel (*.xls, *.xlsx), MS Access (*.mdb, *.accdb) • CSV-Formate ohne zusätzliche Datendefinitionsdateien (Setup, Syntax, Command file) • Column Binary-Format
Textdateien	• PDF/A (*.pdf) • Plain Text-Formate (ASCII)	• OpenDocument Text (*.odt) • PDF (*.pdf) • MS Word (*.doc, *.docx) • RichTextFormat (*.rtf) • HTML (*.htm, *.html)
Bilder	• TIFF Version 6 unkomprimiert (*.tif)	• JPEG (*.jpg, *jpeg), PNG (*.png), GIF (*.gif), BMP (*.bmp) • PDF/A, PDF (*.pdf)
Audio	• MPEG-1 Audio Layer 3 (*.mp3)	• Wave Audio Format WAV (*.wav)
Video	• MPEG-4 (*.mp4), MPEG-2 (*.mpg)	• Audio Video Interleave AVI (*.avi) • Windows Media Video WMV (*.wmv)

Abhängig u. a. von

☺ Bedarfen der Forschenden

☺ Forschungsdisziplin

☺ (Kooperations-)Partnern*innen

☺ (Daten-)Archiven / Repositorien

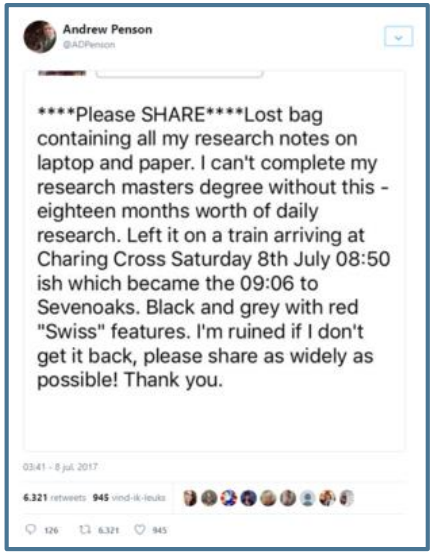
https://www.forschungsdaten-bildung.de/formate.

Inhalte der Sitzung

Bilder: www.pixabay.org (CC-0).

Dateien gegen Verlust sichern

Sind Ihre Dateien ausreichend gegen Verlust gesichert?



The tweet (Penson, 2017) in the image dates from the 7th of July 2017. Although the tweet is real, the scenario about the contents of the backpack is fictional and based on this scenario in a blog post by Peter Murray-Rust (2011).

BBC News (2003) Fire Disaster at the School of Electronics and Computer Science, University of Southampton

Source: Die Welt (dpa)
http://www.welt.de/regionales/nrw/article138019133/Nach-dem-Einsturz-sind-fast-alle-Schaetze-gerettet.html

Speicherung von Forschungsdaten

- Grundlegende Fragen zur Speicherung:
  - Was soll gespeichert werden?
  - Wie soll gespeichert werden?
  - Welche Speichermedien werden genutzt?
  - Wie werden Dateien gesichert (Back-Ups)?
  - Wer braucht Zugang/Zugriff zu den Daten?
- Vielfältige Möglichkeiten, Daten zu speichern:
  - Wie viel Speicherplatz wird benötigt?
  - Gibt es Dienste und Standards an Ihrer Einrichtung?
  - Entstehen zusätzliche Kosten?
  - Wie lange sollen die Daten gespeichert werden?

cc-by http://hochstenbach.wordpress.com

Hochstenbach, P. (2015): FOSTER Cartoon: Data Management. FOSTER-Project (EU-Funding 612425). Verfügbar unter: https://www.fosteropenscience.eu/foster-taxonomy/research-data-management. (CC-by).

Speichermöglichkeiten

Portable Speicherung

☺ geringe Kosten, transportabel

☹ kurzlebig und leicht zu verlieren, evtl. Datenschutzproblem

Cloud-Dienste

☺ leichte Zugänglichkeit, Back-ups und Versionierung

☹ evtl. Datenschutz- und Urheberechtsproblem

Lokale Speicherung

☺ gut zu schützen

☹ kurzlebig und nur für die jeweilig nutzende Person zugänglich

Netzwerke

☺ zentrale Speicherung, zentraler Zugang, ggf. Back-ups

☹ hohe Kosten, ggf. Datenschutz- und Zugangsproblem (Externe)

☹ keine Pauschallösung

☺ abhängig von Projektbedarfen, institutionellen Vorgaben

☺ ggf. mehrere Speichermöglichkeiten berücksichtigen

Bilder: www.pixabay.org (CC-0).

Speichermedien

optisch

☺ geringe Kosten, transportabel

☹ kurzlebig und (relativ) instabil

portable (flash-drive)

☺ geringe Kosten, transportabel

☹ leicht zu verlieren, relativ geringe Speicherkapazität

magnetisch

☺ geringe Kosten, relativ hohe Speicherkapazität

☹ (relativ) instabil

build-in (flash-drive)

☺ relativ stabil und langlebig

☹ hohe Kosten, relativ geringe Speicherkapazität

☹ keine Pauschallösung, abhängig von Projektbedarfen

☺ mindestens zwei verschiedene Medien nutzen und regelmäßige Integritätsprüfung

☺ Medien regelmäßig ersetzen (alle 2-5 Jahre)

Bilder: www.pixabay.org (CC-0).

Inhalte der Sitzung

Bilder: www.pixabay.org (CC-0).

Geschütztes Speichern und Verarbeiten

Unter Berücksichtigung des Stands der Technik, der Implementierungskosten und der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung sowie der unterschiedlichen Eintrittswahrscheinlichkeit und Schwere des Risikos (…) [trifft der Verantwortliche] (…) geeignete technische und organisatorische Maßnahmen, um ein (…) angemessenes Schutzniveau zu gewährleisten“

(Art.32.1 DSGVO)

- Organisatorisch
  - Anonymisierung (Pseudonymisierung)
  - Trennung personenbezogener Daten von den Forschungsdaten
  - Aufbewahrung personenbezogener Daten innerhalb der EU
- Technisch
  - Zugangskontrolle
  - Passwörter
  - Verschlüsselung
  - Vernichtung nicht mehr benötigter Daten

Maßnahmen: regelmäßige Überprüfung, Bewertung und Evaluierung

Passwörter

- Ein gutes Passwort
  - hat mindestens acht bis fünfzehn Zeichen
  - ist eine zufällige Reihung von Zeichen
- Kombination von:
  - Buchstaben in
    - Großschreibung: A - Z
    - Kleinschreibung: a - z
  - Zahlen: 0 - 9
  - Sonderzeichen: ! " # $ % & ' ( ) * + , - . / : ; etc.

⇒ „Passsätze“ statt Passwörter

Die 10 schlechtesten Passwörter 2018

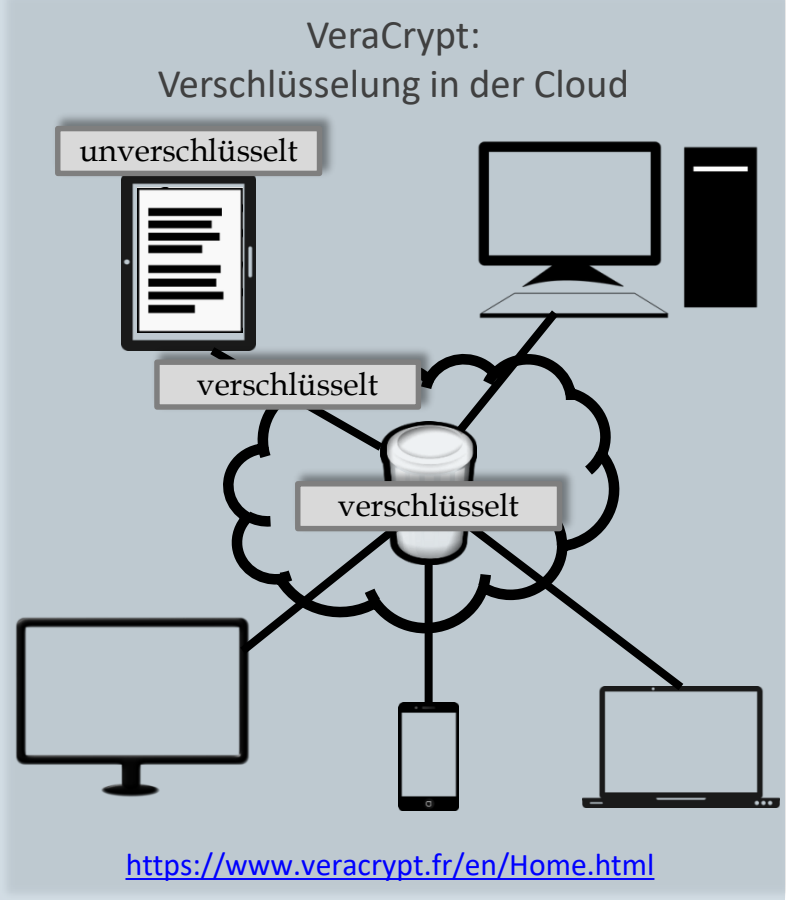
1. 123456
2. password
3. 123456789
4. 12345678
5. 12345
6. 111111
7. 1234567
8. sunshine
9. qwerty
10. iloveyou

https://www.teamsid.com/splashdatas-top-100-worst-passwords-of-2018/

Verschlüsselung

- Verschlüsselung
  - Daten werden in einen unlesbaren Code konvertiert
  - Re-Konvertierung durch einen „Schlüssel“ (Passwort)
- Benutzen Sie Verschlüsselung für
  - den Transport oder Transfer von Daten,
  - die Aufbewahrung und für das Backup von Daten,
  - Mobilgeräte (z. B. Laptops, USB-Sticks)

  z. B. *7Zip*, *Gpg4win* oder *VeraCrypt*



Bilder: www.pixabay.org (CC-0).

Datenvernichtung

- Vernichtung von Dat(ei)en
  - aufgrund rechtlicher Bestimmungen oder sonstiger Vereinbarungen
  - von temporären (Arbeits-)Dateien
  - am Projektende

Vernichtung von Dat(ei)en

- ☺ durch Überschreiben von digitalen Dateien, z. B. *dban*, *Eraser*
- ☺ durch physische Vernichtung der Festplatte
- ☺ Dokumente und Dateien im Papierformat durch Schreddern

Bild: www.pixabay.org (CC-0).

Weiterführende Ressourcen

Hinweise zur Datenorganisation & Dateibenennung des *VerbundFDB*:

- https://www.forschungsdaten-bildung.de/datenmanagement?la=de
- https://www.forschungsdaten-bildung.de/datei-benennung?la=de%20

Weiterführende Ressourcen zur Datenorganisation & Dateibenennung:

- CESSDA Training Working Group (2017 - 2018). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Unter: https://www.cessda.eu/DMGuide.
- Corti, L., van den Eynden, V., Bishop, L. & Woollard, M. (2014): Managing and Sharing Research Data. A Guide to Good Practice. London: Sage Publication Ltd.
- Recker, J. und Evelyn Brislinger (2019): *Dateiorganisation in empirischen Forschungsprojekten*. In: Jensen, U., S. Netscher und K. Weller (Hrsg.): *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten*. S. 81-96. Opladen, Berlin, Toronto : Barbara Budrich.

Hinweise zu Cloud-Diensten

- Borgmann, M., Hahn, T., Herfert, M., Kunz, T., Richter, M., Viebeg, U. und Vowé, S. (2012): *On the Security of Cloud Storage Services*. Frauenhofer Institut, SIT. Technical Report. Unter: https://www.sit.fraunhofer.de/fileadmin/dokumente/ studien_und_technical_reports/Cloud-Storage-Security_a4.pdf.

Danke für die Aufmerksamkeit!

Mail: **forschungsdaten-bildung@gesis.org**
Telefon: **0221 / 47694-473 (Dr. Marcus Eisentraut)**
0221 / 47694-457 (Timo Hutflesz)