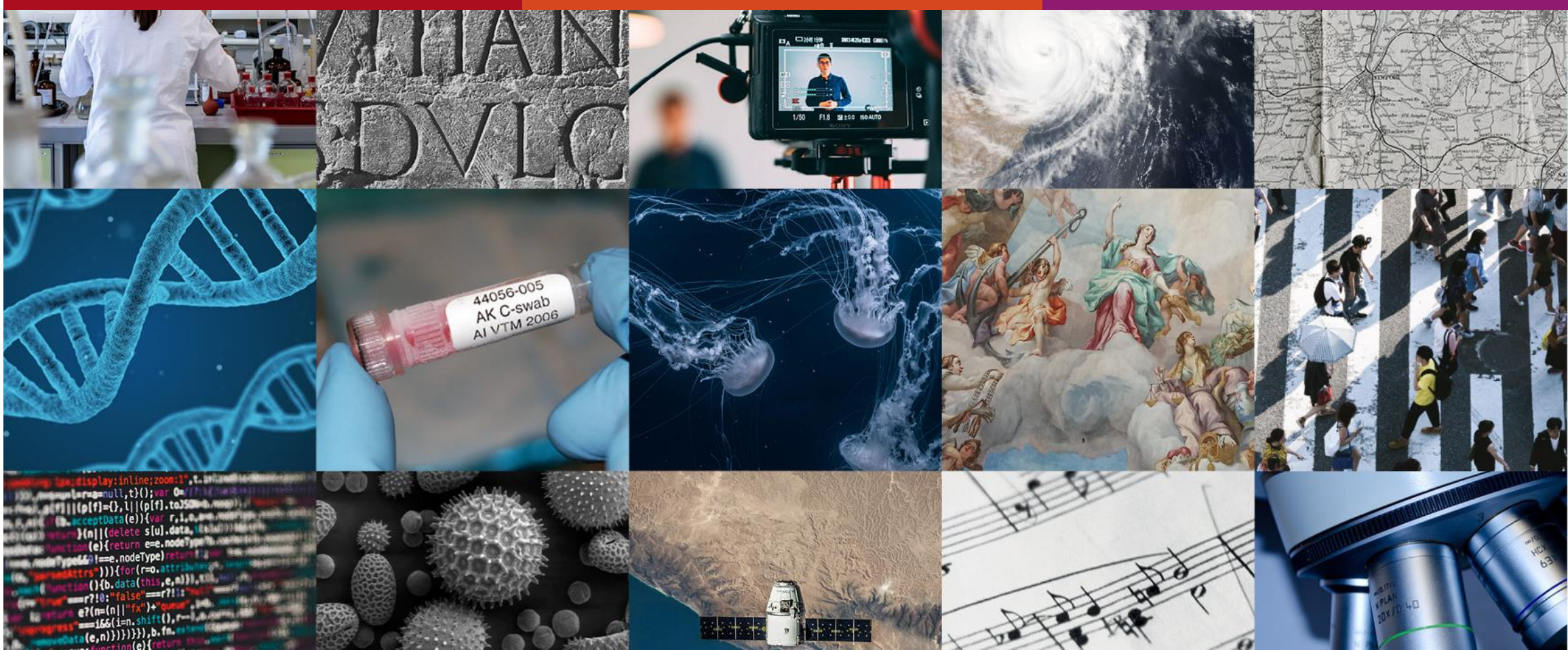




Zertifikatskurs Forschungsdatenmanagement

FDM in den Sozialwissenschaften



This work is licensed under [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



ZBIW.

Technology
Arts Sciences
TH Köln

Wer sind wir: Dr. Michaela Rizzolli

Titel Vorname Nachname

Dr. Michaela Rizzolli

Tätig an folgender
Einrichtung (seit)

FDZ Qualiservice (Universität Bremen),
seit 2022

Im FDM tätig seit

2019

FDM-Aufgaben

- Beratung und Schulung von Forschenden zu FDM & Data Sharing
- Entwicklung von Workflows für die Archivierung und Bereitstellung von Daten aus der ethnografischen Feldforschung – in Zusammenarbeit mit dem Fachinformationsdienst Sozial- und Kulturanthropologie (FID SKA)

Meine Inhalte & Rolle im
Zertifikatskurs

Dozentin Modul 3 / FDM in den Sozialwissenschaften

Kontakt

rizzolli@uni-bremen.de



Bildquelle: Miriam Klingl

Wer sind wir: Dr. Sebastian Netscher

Titel Vorname Nachname

Dr. Sebastian Netscher

Tätig an folgender
Einrichtung (seit)

GESIS – Leibniz-Institut für Sozialwissenschaften,
seit 2006

Im FDM tätig seit

2009

FDM-Aufgaben

- Beratung und Schulung zum FDM
- Servicemanager von da|ra – Registrierungsagentur für sozial- und wirtschaftswissenschaftliche Daten bei DataCite
- Forschung: Standardisierung des FDM, Aufwendungen zum FDM

Meine Inhalte & Rolle im
Zertifikatskurs

Dozent Modul 3 / FDM in den Sozialwissenschaften

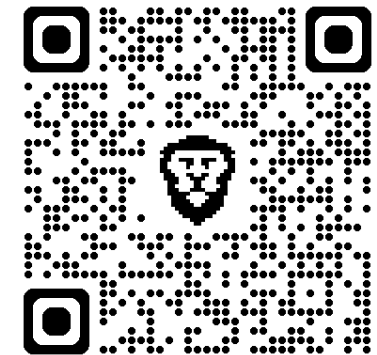
Kontakt

sebastian.netscher@gesis.org



NFDI – Nationale Forschungsdaten Infrastruktur

- Nationale Forschungsdaten Infrastruktur
 - gemeinnütziger Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. (Hannover, März 2021)
 - mehr als 270 Mitgliedsinstitutionen, wie Hochschulen, außeruniversitäre Forschungseinrichtungen, Fachgesellschaften
- Ziele
 - Schaffung von Rahmenbedingungen für eine rechtskonforme, interoperable und nachhaltige Dateninfrastruktur
 - Verbesserung des Forschungsdatenmanagement und der (Nach-)Nutzungsmöglichkeiten von Daten



<https://www.nfdi.de/>

NFDI – Konsortien und Sektionen

Konsortien

Geistes- und Sozialwissenschaften



[BERD@NFDI](#)

Betriebswirtschafts-,
Volkswirtschaftslehre & verwandte
Daten



[KonsortSWD](#)

Sozial-, Bildungs-, Verhaltens- &
Wirtschaftswissenschaften



[NFDI4Culture](#)

materielle und immaterielle
Kulturgüter



[NFDI4Memory](#)

historisch arbeitende
Geisteswissenschaften



[NFDI4Objects](#)

materielle Hinterlassenschaften
der Menschheitsgeschichte



[Text+](#)

Sprach- & textbasierte
Forschungsdateninfrastruktur

Naturwissenschaften



[DAPHNE4NFDI](#)

Daten aus Photonen- und
Neutronen Experimenten



[FAIRmat](#)

Physik der kondensierten Materie
& chemische Physik fester Stoffe



[NFDI4Cat](#)

Wissenschaften mit Bezug zur
Katalyse



[MaRDI](#)

mathematische
Forschungsdateninitiative



[NFDI4Chem](#)

Fachkonsortium Chemie



[NFDI4Earth](#)

Erdsystemforschung



[PUNCH4NFDI](#)

Teilchen, Universum, Kerne &
Hadronen

Lebenswissenschaften



[DataPLANT](#)

Pflanzen-
Grundlagenforschung



[FAIRagro](#)

Agrosystemforschung



[NFDI4Immuno](#)

Immunologie



[GHGA](#)

Deutsches Humangenom-
Phenomarchiv



[NFDI4Biodiversity](#)

Biodiversität, Ökologie und
Umweltdaten



[NFDI4BioIMAGE](#)

Mikroskopie und Bildanalyse



[NFDI4Health](#)

personenbezogene
Gesundheitsdaten



[NFDI4Microbiota](#)

Mikrobiota-Forschung

Ingenieurwissenschaften



[NFDI4DataScience](#)

Datenwissenschaften und
Künstliche Intelligenz



[NFDI4Energy](#)

interdisziplinäre
Energiesystemforschung



[NFDI4Ing](#)

Ingenieurwissenschaften



[NFDI-MatWerk](#)

Materialwissenschaft &
Werkstofftechnik



[NFDI4CS](#)

Computer Science

Sektionen

[Base4NFDI](#)

Basisdienste für NFDI (Konsortienverbund)

[Common Infrastructures](#)

(section-infra)

[Ethical, Legal and Social Aspects](#)

(section-ELSA)

[\(Meta\)daten, Terminologien, Provenienz](#)

(section-metadata)

[Training & Education](#)

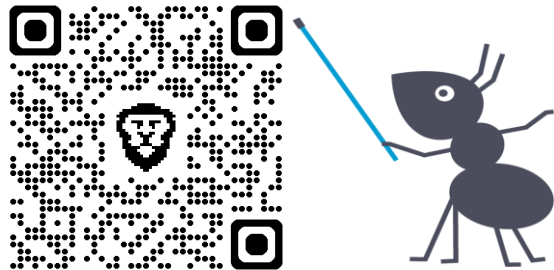
(section-edutrain)

[Industry Engagement](#)

(section-industry, seit 2023)

- Konsortien
Vernetzung und gemeinschaftliche
Bearbeitung von fachinternen Themen
(wissenschaftliche Fachbereiche)
- Sektionen
Vernetzung und gemeinschaftliche
Bearbeitung von Themen, die für alle oder
mehrere Konsortien relevant sind

KonsortSWD



<https://www.konsortswd.de/>

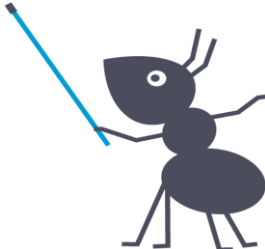
KonsortSWD

- Konsortium für die Sozial-, Verhaltens-, Bildungs-, und Wirtschaftswissenschaften
- 1 von 26 thematischen Konsortien der NFDI

Ziele von KonsortSWD

- **Forschungsdaten**infrastruktur zur Beforschung der Gesellschaft stärken, erweitern und vertiefen
- Technisch und inhaltlich bei Nutzung und Nachnutzung (**sensibler**) Daten unterstützen
- **Forschungsdatenmanagement**
- Sicherstellung der Datenverfügbarkeit

KonsortSWD Services

CODI	Forum4 MICA	Harmonized Variables	Linking Textual Data	Open Data Format	Qualidata Net	Data Findability	FDI Ausschuss
RDCnet	RDM COMPAS	RDM Grants	RDM Support (Qual.)	Stamp (Standardis. DMP)	Open API	PID-Services	RDC Support
Rat für Sozial- und Wirtschafts- daten (RatSWD)	Helpdesk	Access to Firm Data	Daten zu Krisen	Guides Forschungs- synthesen	RDC MoL/ MO RE Data		

QualidataNet – Netzwerk für qualitative Daten

[HOME](#)[ÜBER UNS](#)[DATEN SERVICES ▾](#)[TOOLS & TRAINING ▾](#)[INFOTHEK](#)[NEWS & EVENTS](#)[KONTAKT](#)[EN](#)

<https://www.qualidatanet.com/de/>

Qualitative Forschungsdaten.

Suchen, sammeln, nutzen, archivieren.
Gemeinsam im Netz.

Durchsuchen Sie die Datenbestände der Netzwerkpartner von QualidataNet. Der Bestand an Daten für die qualitative Forschung wird sukzessive erweitert.



DATEN TEILEN



DATEN FINDEN



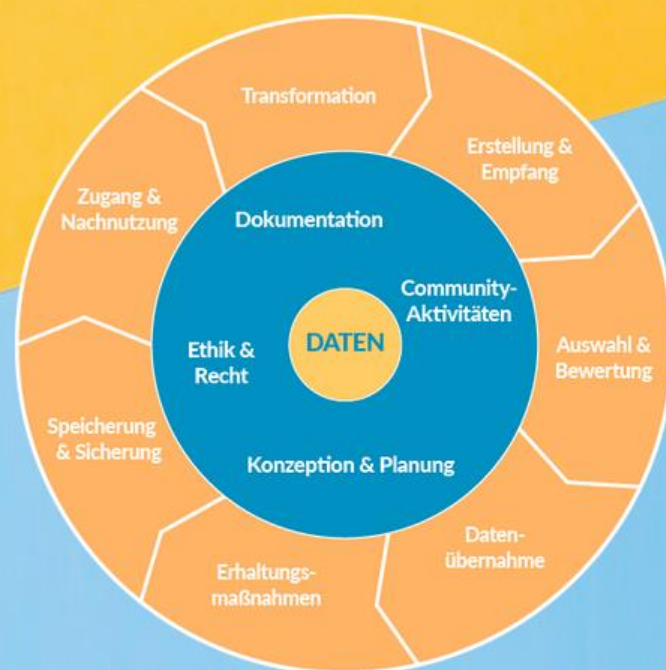
DATEN NUTZEN



DATEN MANAGEN

FDM Knowledge Base

In der Knowledge Base finden Sie Informationen zum generischen Forschungsdatenmanagement für Datenkuratierende, sowie disziplinspezifische Angebote für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften.



Die Angebote von [RDM Compass](https://rdm-compas.org/) sind auf die Tätigkeit der [Datenkuration](#) in Forschungsdatenzentren ausgerichtet. Daher nutzen wir ein einfaches Modell des [Datenkurationslebenszyklus](#) als Grundlage.

Um direkt zu den Informationsmaterialien der einzelnen Prozessschritte der Datenkuration zu gelangen, klicken Sie unten auf den entsprechenden Abschnitt. Oder wechseln Sie einfach direkt in die Knowledge Base, wo Sie alle Beiträge gesammelt finden.

- > [Erstellung & Empfang](#)
- > [Auswahl & Bewertung](#)
- > [Datenübernahme](#)
- > [Erhaltungsmaßnahmen](#)
- > [Speicherung & Sicherung](#)
- > [Zugang & Nachnutzung](#)
- > [Transformation](#)

[Zu der FDM Knowledge Base](#)

<https://rdm-compas.org/>



Fragen?

Agenda



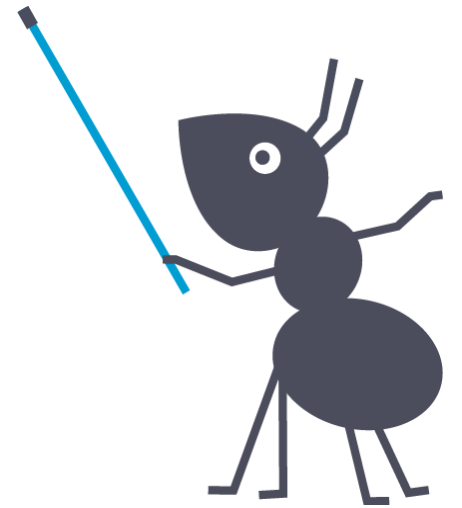
Datenschutz und Anonymisierung



Datendokumentation



Datenarchivierung/-bereitstellung



Agenda



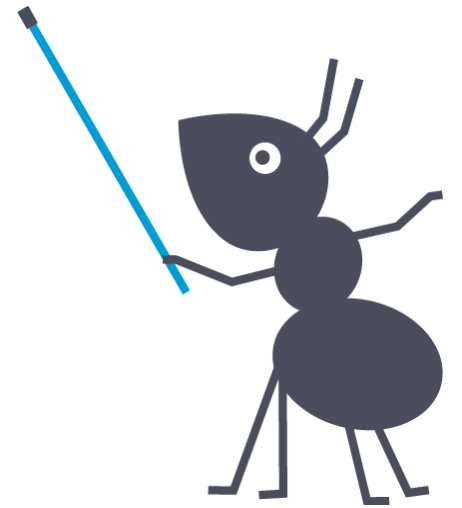
Datenschutz und Anonymisierung von quantitativen Daten



Datendokumentation



Datenarchivierung/-bereitstellung



Forschungsdaten in den Sozialwissenschaften

jede Art von systematisch strukturierten Informationen, z. B. Kennzahlen, Texte, Bilder usw., die zu Forschungszwecken gesammelt, z. B. durch Interviews, Umfragen, Beobachtungen, Web Scraping usw., und verarbeitet werden

Qualitativer Ansatz (Daten)

- induktiver Ansatz zum Verständnis und zur Erklärung sozialer Phänomene
- komplexe, heterogene Analyse zur Generierung von Hypothesen (Theorie)
- (relativ) unstrukturierte Daten, kleine Fallzahlen

Quantitativer Ansatz (Daten)

- deduktiver Ansatz zur Prüfung von Hypothesen (Theorie)
- Operationalisierung, Messung und statistische Analyse (Generalisierung)
- (relativ) strukturierte Daten, große Fallzahlen

Quantitative Forschungsdaten in den Sozialwissenschaften

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	
1	1.040e+14	4	26	104012	99	Variable	1	1973	2	1642	1040
2	1.040e+14	1	19	104008	99		2	1943	1	99	99
3	1.040e+14	4	98	104003	99		1	1990	1	99	99
4	1.040e+14	2	98	104010	99		1	1983	2	1756	1040
5	1.040e+14	1	18	104007	99		2	1927			
6	1.040e+14	2	19	104007	99		1	1983			
7	1.040e+14	4	15	104005	99		2	1970			
8	1.040e+14	4	19	104006	99		1	1942			
9	1.040e+14	4	15	104005	99		1	1965			
10	Untersuchungsobjekt			104004	99	2	1955				
11	1.040e+14	7	15	104003	99	2	7777				
12	1.040e+14	4	18	104005	99	2	1938				
13	1.040e+14	7	17	104005	99	1	1945				
14	1.040e+14	4	18	104005	99	2	1949				
15	1.040e+14	2	15	104003	99	2	7777				
16	1.040e+14	4	32	104007	99	1	1974				
17	1.040e+14	4	98	104006	99	1	7777	1	99	99	
18	1.040e+14	4	25	104010	99	2	1968	1	99	99	
19	1.040e+14	4	40	104007	99	1	1967	1	99	99	
20	1.040e+14	1	23	104004	99	1	1932	1	99	99	
21	1.040e+14	2	18	104003	99	2	1965	1	99	99	
22	1.040e+14	1	27	104011	99	1	1956	1	99	99	
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	

INTERVIEWER-ANWEISUNG
KODIEREN

Q63 Geschlecht der Zielperson
1 – Männlich
2 – Weiblich
8 – Weiß nicht (spontan)

Q64 In welchem Jahr sind Sie geboren?
.....1986
8888 – Weiß nicht (spontan)
9999 – Keine Antwort

Variable

INTERVIEWER-ANWEISUNG Q63: GESCHLECHT NUR DURCH BEOBACHTUNG KODIEREN

Q63 Geschlecht der Zielperson

1 – Männlich

2 – Weiblich

8 – Weiß nicht (spontan)

(v225)

Q64 In welchem Jahr sind Sie geboren?

.....1986

8888 – Weiß nicht (spontan)

9999 – Keine Antwort (spontan)

(v226)

Merkmalsausprägung

Anonymisierung: Direkte und indirekte Identifikatoren

personenbezogene Daten:

„alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (...) beziehen; (...) die **direkt oder indirekt**, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann“ (Art. 4.1 DSGVO)

zwei Arten von Identifikatoren

direkt

Name,
Adresse,
Bild etc.



indirekt

Universität



Institut



Position



Selbstlernteil: Anonymisierung quantitativer Daten

Nr.	Art der Information	Text	Antwort	Weiter mit
001	Frage	Geburtsdatum		002
		__/__/____		
002	Frage	Geschlecht		003
		1: Männlich		
		2: Weiblich		
		3: Divers		
		9997: Verweigert		
003	Frage	Familienstand		004
		1: Verheiratet		
		2: Eingetragene Lebenspartnerschaft		
		3: Geschieden/eingetragene Lebenspartnerschaft aufgehoben		
		4: Verwitwet/eingetragene(r) Partnerin verstorben		
		5: Ledig (NIE verheiratet oder in einer eingetragenen Lebenspartnerschaft)		
		9998: Weiß nicht		
		9999: Keine Angabe		
004	Frage	Leben in Ihrem Haushalt Kinder unter 16 Jahren, und wenn ja, wie viele?		005
		1: Ja, und zwar __ Kind(er) unter 16 Jahren		
		2: Nein		
		9999: Keine Angabe		
005	Frage	In welchem Land sind Sie selbst geboren?		ENDE
		1: Deutschland		
		2: Anderes Land, nämlich:		
		9998: Weiß nicht		
		9999: Keine Angabe		

Kritische Fragen im Kontext der Anonymisierung?

Worin liegt das Re-Identifizierungsrisiko?

Wie könnten die Daten adäquat anonymisiert werden?

Selbstlernteil: Kritische Fragen

Frage	Risiko	Anonymisierung
001. Geburtsdatum	erhöhtes Risiko der Re-Identifikation	Beseitigen (Löschen) von Information, z. B. <i>Geburtsjahr</i> statt <i>Geburtsdatum</i>
002. Geschlecht (Gender)	Risiko der Identifikation sexueller Orientierung („ <i>divers</i> “)	Aggregation von (einzelnen) Antwortkategorien, z. B. <i>divers</i> und <i>keine Angabe</i>
003. Familienstatus	Risiko der Identifikation sexueller Orientierung („ <i>Partnerschaft</i> “)	Aggregation von (einzelnen) Antwortkategorien, z. B. <i>verheiratet</i> und <i>Partnerschaft</i>
004. Anzahl an Personen unter 16 Jahren	erhöhtes Risiko der Re-Identifikation	Aggregation der ‚oberen‘ Antwortkategorien, z. B. <i>6 Kinder und mehr</i>
005. Geburtsland	erhöhtes Risiko der Re-Identifikation	Aggregation ‚seltener‘ Geburtsländer, z. B. <i>Maghrebstaaten</i> anstatt <i>Marokko, Tunesien</i> etc.

Selbstlernteil: Es kommt auf die Grundgesamtheit an!

Ausländer, Deutschland, Stichtag, Geschlecht/Altersjahre/
Familienstand, Ländergruppen/Staatsangehörigkeit

GENESIS-Table: 12521-0002

foreign nationalities: martial status

date of reference: 31.12.2021

	single	married	widowed (married)	divorced (married)	civil partnership	divorced (partnership)	widowed (partnership)
Egypt	19930	15305	220	1480	20	5	-
Equatorial Guinea	80	30	5	10	-	-	-
Ethiopia	12890	5285	225	775	40	-	-
Afghanistan	186710	85160	4790	3035	185	10	5
Albania	38145	36525	425	1945	50	5	-
Algeria	8385	7925	250	1795	15	5	-
Andorra	10	5	-	5	-	-	-
Angola	3350	1275	85	330	5	-	-
Antigua and Barbuda	10	15	-	-	-	-	-
...							

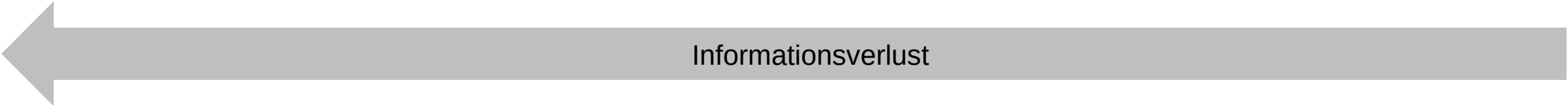
Anonymisierungsstrategien

direkte Identifikatoren
(wie Namen, Adresse etc.)
immer getrennt von
Umfragedaten aufbewahren

sensible Daten nicht
erheben bzw. löschen,
wenn nicht benötigt

Daten aggregieren
z. B. Altersgruppen

Zusammenfassen von
Kategorien
(oben bzw. unten)
z. B. Personen unter 16
im Haushalt



Informationsverlust

alle Änderungen dokumentieren!



Fragen?

Agenda



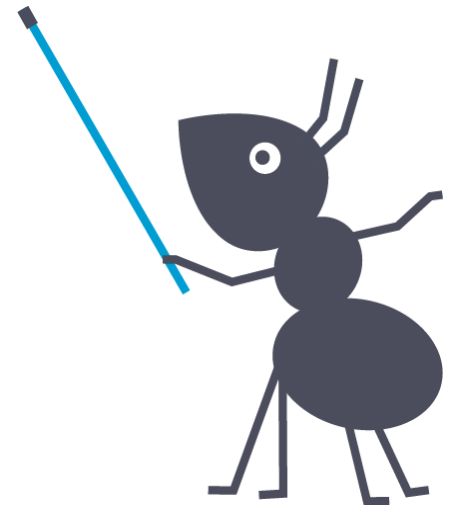
Datenschutz und Anonymisierung von qualitativen Daten



Datendokumentation



Datenarchivierung/-bereitstellung

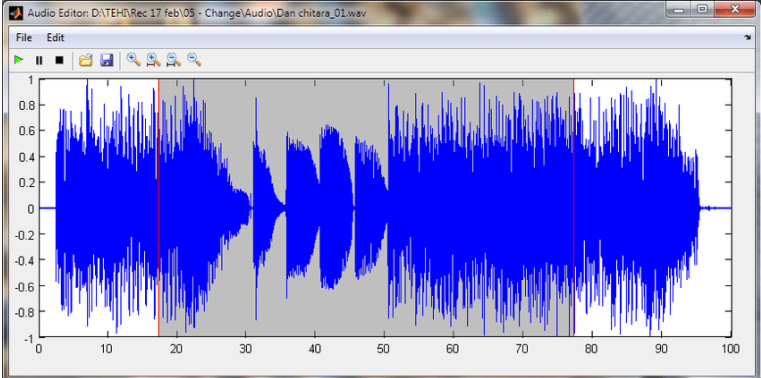


Beispiele für qualitative Forschungsdaten (I)

Wiedergabe (k)

Interviewed by Paul Thompson, 2003

0:01 / 0:51



INTERVIEW TRANSCRIPT

I: If we could go back to the beginning. You were born in 1922?

R: That's right.

I: And where was that?

R: In London, in England.

I: Yes. And could you tell me about your family background? I don't know how far back you remember. Your grandparents, for instance?

R: Yes. Well, I knew my maternal grandmother very well, as is quite commonly the case. She lived well into her nineties. My maternal grandfather died when my mother was about three months pregnant, so I never knew him, and she was devastated by this, because she was greatly attached to him.

I: What was his profession, then?

R: I think that he was a salesperson. It's a little bit difficult to tell from what I've read, but he had some kind of business activity, and I had heard, at one time, it was to do with boots and shoes - but the description gets vague. What I am clear about is that the family was Methodist, and my grandmother had converted to Methodism because of my grandfather. I don't know what her religion had been before. But what I do know is that she was -

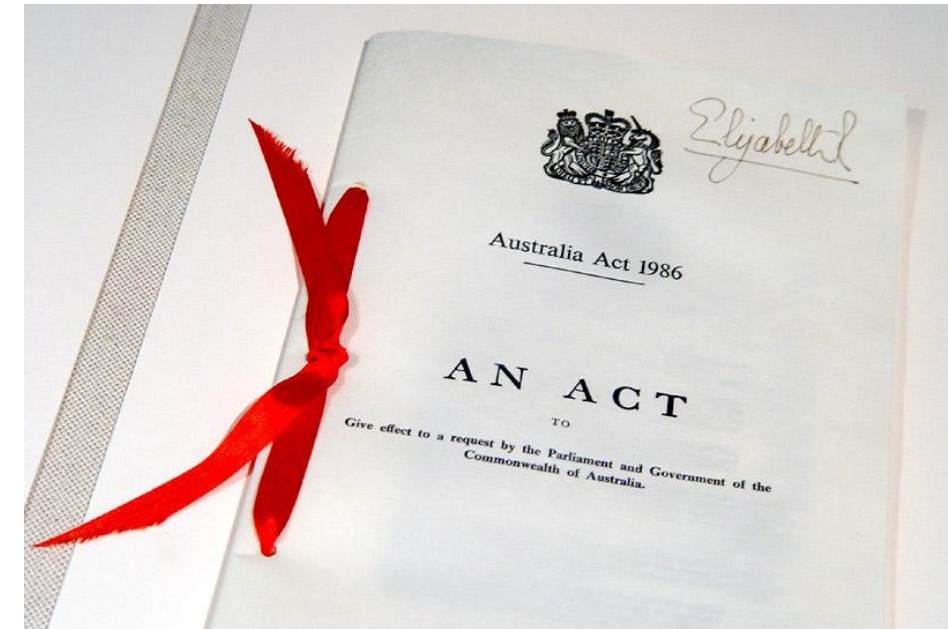
Thompson, P. (2019). *Pioneers of Social Research, 1996-2012*. [data collection]. 4th Edition. UK Data Service. SN: 6226, [DOI: 10.5255/UKDA-SN-6226-6](https://doi.org/10.5255/UKDA-SN-6226-6)

Beispiele für qualitative Forschungsdaten (II)

Beobachtungsprotokolle

6/21/91
 Bob and I have decided that for the field project at Second Woods we will employ some new methods for the mapping of artifacts and features in units. We will use the system of triangulation instead of rectangular coordinates. This requires that we have some new forms for recording the distance measurements. It also requires us to have drawing compasses on site. We decided that it is too risky not to actually plot the points in the field. We need to be able to catch mistakes right away.

Feldnotizen: The Digital Archaeological Record (tDAR), doi:10.6067/XCV86Q1VQ6



Dokumentarten (z.B. Policy-Dokumente, Broschüren, Website-Informationen)

Merkmale von qualitativer Sozialforschung & Daten

- Forschungsziele: oft Exploration, (dichte) Beschreibung, Theorieentwicklung
- Verstehen von Sinn – Analyse von Prozessen, Mechanismen, Konfigurationen etc.
- Deshalb: Offenheit des Forschungsprozesses (Fallauswahl, Methodenspezifikation)
- Resultat: viele, heterogene, sehr informationsreiche, relativ unstrukturierte Daten
- Besondere datenschutzrechtliche (und forschungsethische!) Herausforderungen
- Schwierigkeit I: Vertraulichkeit und Schutz von Forschungsteilnehmenden
- Schwierigkeit II: Informierte Einwilligung u.U. nicht möglich bzw. nicht schriftlich
- Schwierigkeit III: Anonymisierung von informationsreichen unstrukturierten Daten

Anonymisierungsübung im Selbstlernteil

Die Aufgabe lautete:

Markieren Sie im unter „Kursmaterialien“ verlinkten Auszug des Transkripts eines qualitativen (genauer: narrativ-biographischen) Interviews jedes Wort, jede Phrase oder Textbestandteil, bei denen Sie eine Anonymisierung für erforderlich halten, um eine Identifizierung der jeweiligen Person zu verhindern.

Überlegen Sie sich, wie bei der Anonymisierung vorgegangen werden könnte, um das Risiko einer Identifizierung von Personen zu verringern und zugleich Nachnutzungen der Daten zu ermöglichen.

Fragen / Diskussion zur Anonymisierungsübung

- Wie viele Stellen haben Sie gekennzeichnet?
- Bei welchen Stellen hatten Sie die größten Schwierigkeiten (und warum)?
- Wie könnte bei der Anonymisierung/Pseudonymisierung der markierten Stellen vorgegangen werden, um die personenbezogenen Daten der Personen zu schützen und zugleich (wissenschaftliche) Nachnutzungen dieser Forschungsdaten zu ermöglichen bzw. befördern?

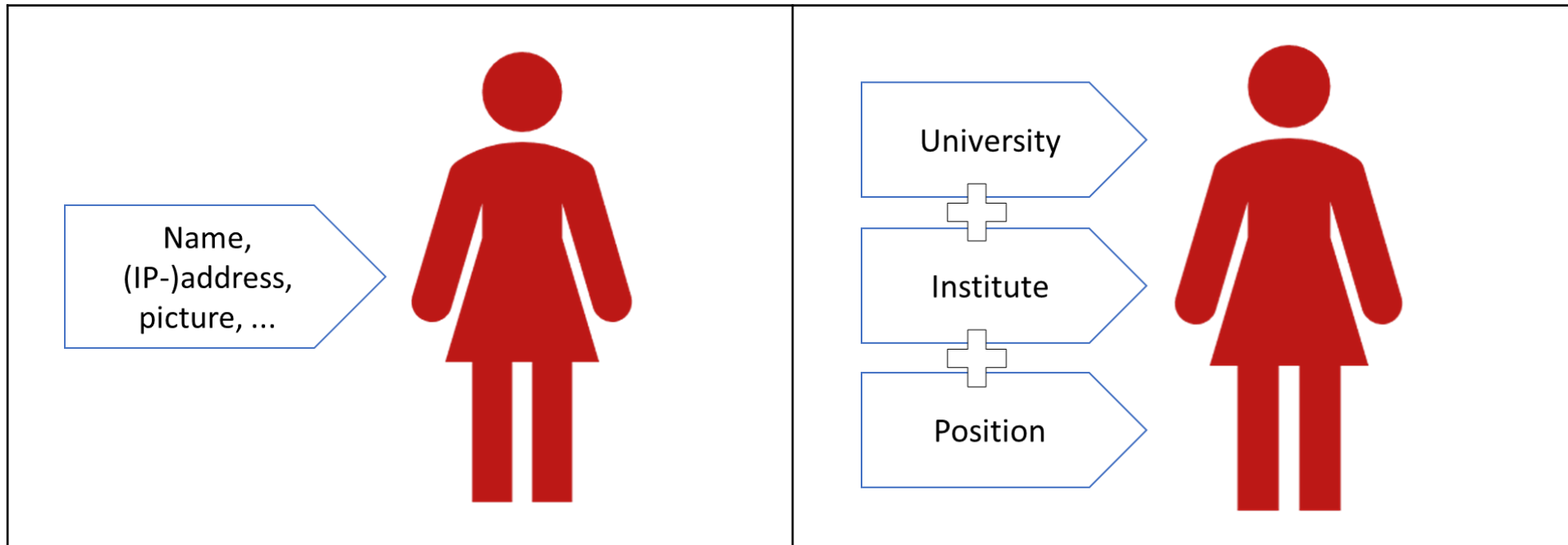
Anonymisierung/Pseudonymisierung – Grundlagen

- **Anonymisierung:** Verändern personenbezogener Daten derart, dass diese Daten nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können (keine eindeutige Definition – DSGVO)
- **Pseudonymisierung:** Ersetzen von direkten und indirekten Identifikatoren durch Pseudonyme, um die Feststellung der Identität des Betroffenen auszuschließen oder wesentlich zu erschweren → Unterschied zur Anonymisierung: reversibel (Art. 4 Nr. 5 DSGVO)

Zentrale Herausforderung: nicht nur *formale* Anonymisierung/Pseudonymisierung (d.h. Verändern/Ersetzen direkter Identifikatoren wie Name, Adresse etc.), sondern *faktische* Anonymisierung/Pseudonymisierung (d.h. auch indirekte Identifikatoren)

Anonymisierung/Pseudonymisierung – Grundlagen (II)

Direkte Identifikatoren vs. indirekte Identifikatoren



Das Konzept der „flexiblen Anonymisierung“

- Ziel: Balance schaffen zwischen Datenschutz und Ermöglichung von Nachnutzung
- Löschen, Schwärzen, „Verfälschen“ erschwert Forschung, führt zu falschen Resultaten
- Deshalb: Konzept der „flexiblen Anonymisierung“ (Pseudonymisierung) von Textdaten
 1. **Ersetzen (sensibler) personenbezogener Daten durch sozialwissenschaftlich relevante (abstrahierte) Informationen** (ggf. unterschiedliche Abstraktionsebenen, z.B. durch hierarchische Listen, Standardklassifikationen etc.)
 2. **Dokumentation der Ersetzungen** (= Pseudonymisierung), so dass Informationen auch wieder „geöffnet“ bzw. genauer spezifiziert werden können (und geschlossen)
 3. **Abhängig vom Forschungsinteresse der Nachnutzenden können bestimmte Informationen weniger pseudonymisiert/abstrahiert werden („geöffnet“)** und im Gegenzug andere Informationen stärker pseudonymisiert/abstrahiert werden, um Nachnutzungen zu ermöglichen und zugleich eine Identifikation zu vermeiden

Bsp.: Ersetzungskategorien bei flexibler Anonymisierung

Ggf. zu anonymisieren:

- Charakteristika der Person
- Ortsangaben
- Zeitangaben
- Ausbildung
- Berufliche u.ä. Tätigkeiten
- Organisationen/Institutionen
- Individuelle Besonderheiten
- ...

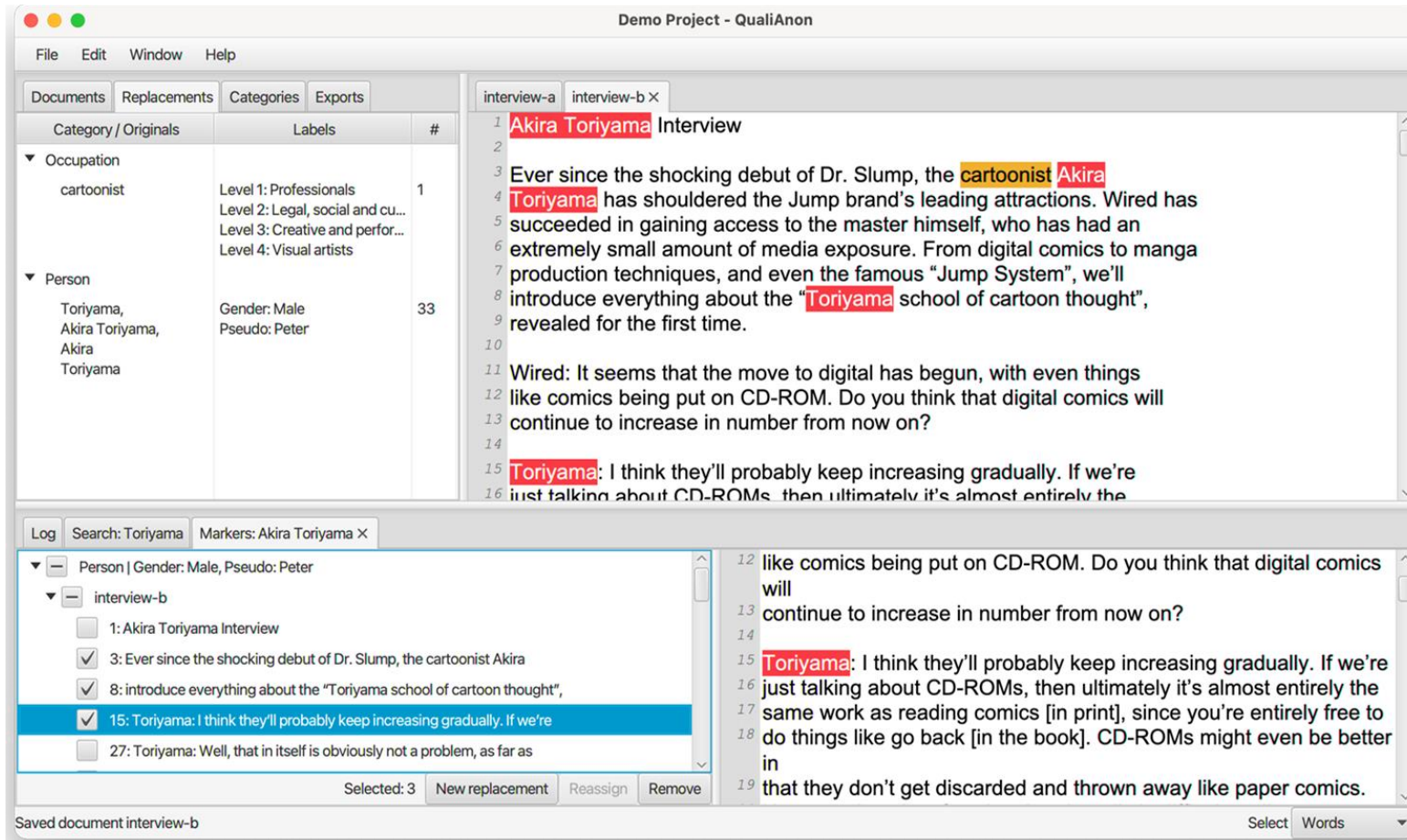
Original	Ersetzungsvorschläge
Ich bin Altenpflegerin.	Ich bin [[Nennung des Berufs]]. Ich bin [[Pflegerin/ Pflegekraft]] Ich bin [[Nennung des Berufs, Dienstleistungsberufe und Verkäufer, ISCO-08: Code 05:]] Ich bin [[Nennung des Berufs, Betreuungsberufe im Gesundheitswesen, ISCO08 Code 532:]]. Ich bin [[Nennung des Berufs, Betreuungsberufe im Gesundheitswesen, Haus- und Familienpfleger, ISCO08 Code 5322:]]

Tool für die (flexible) Anonymisierung von Text: QualiAnon

- Markieren und Ersetzen von sensiblen Informationen
- Halb-automatisch; Kontrolle durch Nutzende(n)
- Dokumentation der Ersetzungen
- Nutzung von Listen/Klassifikationen für Ersetzung
- Beim Export: Auswahl des Abstraktionsgrads
- Separates Speichern der jeweiligen Dateien
- Anlegen, Teilen, Nachnutzen von Ersetzungslisten
- Open Source



Tool für die Anonymisierung von Text: QualiAnon (II)



Für Informationen und Zugang zum GitHub-Repository:

qualianon@uni-bremen.de

<https://github.com/pangea-data-publisher/qualianon>



Fragen?

Agenda



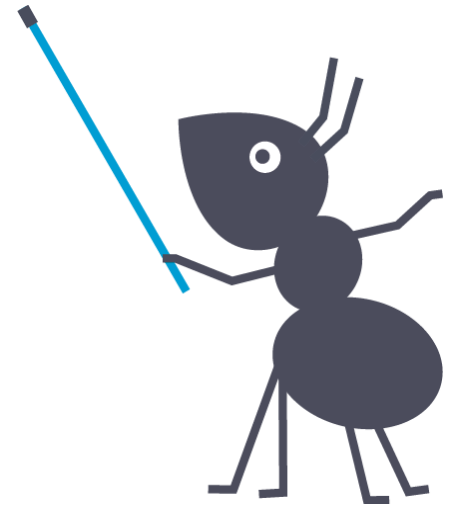
Datenschutz und Anonymisierung



Datendokumentation



Datenarchivierung/-bereitstellung



Agenda



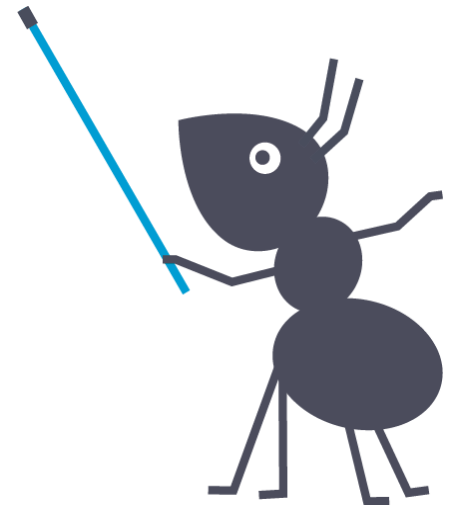
Datenschutz und Anonymisierung



Datendokumentation von quantitativen Daten



Datenarchivierung/-bereitstellung



Worum geht es in diesen Daten?

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	26		104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	2		104008	99	2	1943	1	99	99	99	1	1	6	4	5
3	1.040e+14	2		104003	99	1	1990			99	99	1	6	3	99	11
4	1.040e+14	2	98	104010	99	1	1983			99	99	1	6	3	99	11
5	1.040e+14	1	18	104007	99	2	1927			99		1	1	6	4	1
6	1.040e+14	2	19	104007	99	1	1983	1	99	99	99	1	6	2	5	3
7	1.040e+14	4	15	104005	99	2	1970	1	99	99	99	1	1	2	4	6
8	1.040e+14	4	19	104006	99	1	1942	1	99	99	99	2	1	6	5	88
9	1.040e+14	4	15	104005	99	1	1965	2	1040	1276	1040	7	4			5
10	1.040e+14	7			99	2	1955	1	99	99	99	1				5
11	1.040e+14				99	2	7777	1	99	99	99	1				77
12	1.040e+14	4			99	2	1938	1	99	99	99	1			5	5
13	1.040e+14	7		104005	99	1	1945	1	99	99	99	1	1	6	5	4
14	1.040e+14	4	18	104005	99	2	1940	2	1040	1380	1040	1	4	6	5	5
15	1.040e+14	2	15	104003	99				99	99	99	1	1	2	4	5
16	1.040e+14	4	32	104007	99				99	99	99	1	6	2	5	5
17	1.040e+14	4	98	104006	99			1	99	99	99	1	6	3	99	11
18	1.040e+14			104010	99	2	1968	1	99	99	99	1	1	2	5	1
19	1.040e+14			104007	99	1	1967	1	99	99	99	1	1	2	6	2
20	1.040e+14		23	104004	99	1	1932	1	99	99	99	1	1	6	4	3
21	1.040e+14	2	18	104003	99	2	1965	1	99	99	99	1	6	2	5	5
22	1.040e+14	1	27	104011	99	1	1956	1	99	99	99	2	1	1	5	2
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	99	1	3	6	3	3
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	99	1	4	2	5	3
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	1040	1	1	6	3	3

Wer wurde befragt?

???

Wann wurde befragt?

Wie wurde befragt?

Was bedeuten die Codes?

Wer hat befragt?

Elemente der Datendokumentation

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	4	26	104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	1	19	104008	99	2	1943	1	99	99	99	1	1	6	4	5
3	1.040e+14	4	98	104003	99	1	1990	1	99	99	99	1	6	3	99	11
4	1.040e+14	2	98	104010	99	1	1983	2	1756	1040	1040	1	6	3	99	11
5	1.040e+14	1	18	104007	99	2	1927	1	99	99	99	1	1	6	4	1
6	1.040e+14	2	19	104007	99	1	1983	1	99	99	99	1	6	2	5	3
7	1.040e+14	4	15	104005	99	2	1970	1	99	99	99	1	1	2	4	6

nachvollziehen & beurteilen



Methodenbericht
(zu den Daten)

verstehen & interpretieren



Codebuch & Fragebogen
(Beziehung zwischen Fragen
und Variablen)

auffinden & zitieren



Metadaten
(zur Registrierung)

... und alles, was das Verständnis der Daten fördert
(z. B. Angaben zu Interviewenden, Showcards, Stimuli etc.)

Methodenbericht und Datenqualität

Frage	Berichtsteil	Fehlerquelle (Datenqualität)
Für welchen Zweck wurden die Daten erhoben?	Zielsetzung und Design	
Wie wurden die Befragten ausgewählt?	Zielpopulation und Stichprobenziehung	Erfassungsfehler Stichprobenfehler Unit-Non-Response-Fehler
Wie wurden die Daten erhoben?	Datenerhebung	Messfehler
Welche Informationen wurden erhoben?	Erhebungsinstrument	Messfehler Item-Non-Response-Fehler
Wer hat die Daten wann und wo erhoben?	Feldarbeit	Messfehler des Interviewers
Wie wurden die Daten bearbeitet, kodiert und gewichtet?	Datenaufbereitung	Verarbeitungsfehler
Wurde der Datenschutz berücksichtigt?	Datenschutz und ethische Fragen	rechtliche Aspekte ignoriert



Fragen?

Agenda



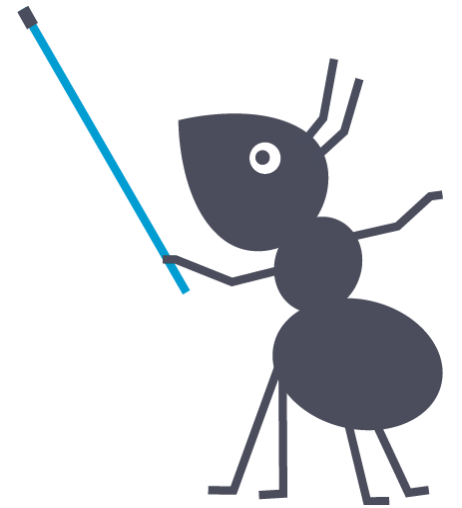
Datenschutz und Anonymisierung



Datendokumentation von qualitativen Daten



Datenarchivierung/-bereitstellung



Dokumentation qualitativer Daten: Beispiel

Wiedergabe (k)

Interviewed by Paul Thompson, 2003

0:01 / 0:51

Pioneers of Social Research: Interview Guide

Pre-interview
Look for an outline biography of cv of the person you are recording (Wikipedia, Who's Who, or they may send you a cv)

Try to give a hard look at what seem to be the researcher's key books

For every interview it is essential to establish in the early part

- a) the date and place of birth of the interviewee, their family background and occupations;
- b) what led them towards research and university work;
- c) what were they key influences in leading them towards their pioneering research themes.
- d)

Opening questions

Date of birth, family and own occupations.

INTERVIEW TRANSCRIPT

I: If we could go back to the beginning. You were born in 1922?

R: That's right.

I: And where was that?

R: In London, in England.

I: Yes. And could you tell me about your family background? I don't know how far back you remember. Your grandparents, for instance?

R: Yes. Well, I knew my maternal grandmother very well, as is quite commonly the case. She lived well into her nineties. My maternal grandfather died when my mother was about three months pregnant, so I never knew him, and she was devastated by this, because she was greatly attached to him.

I: What was his profession, then?

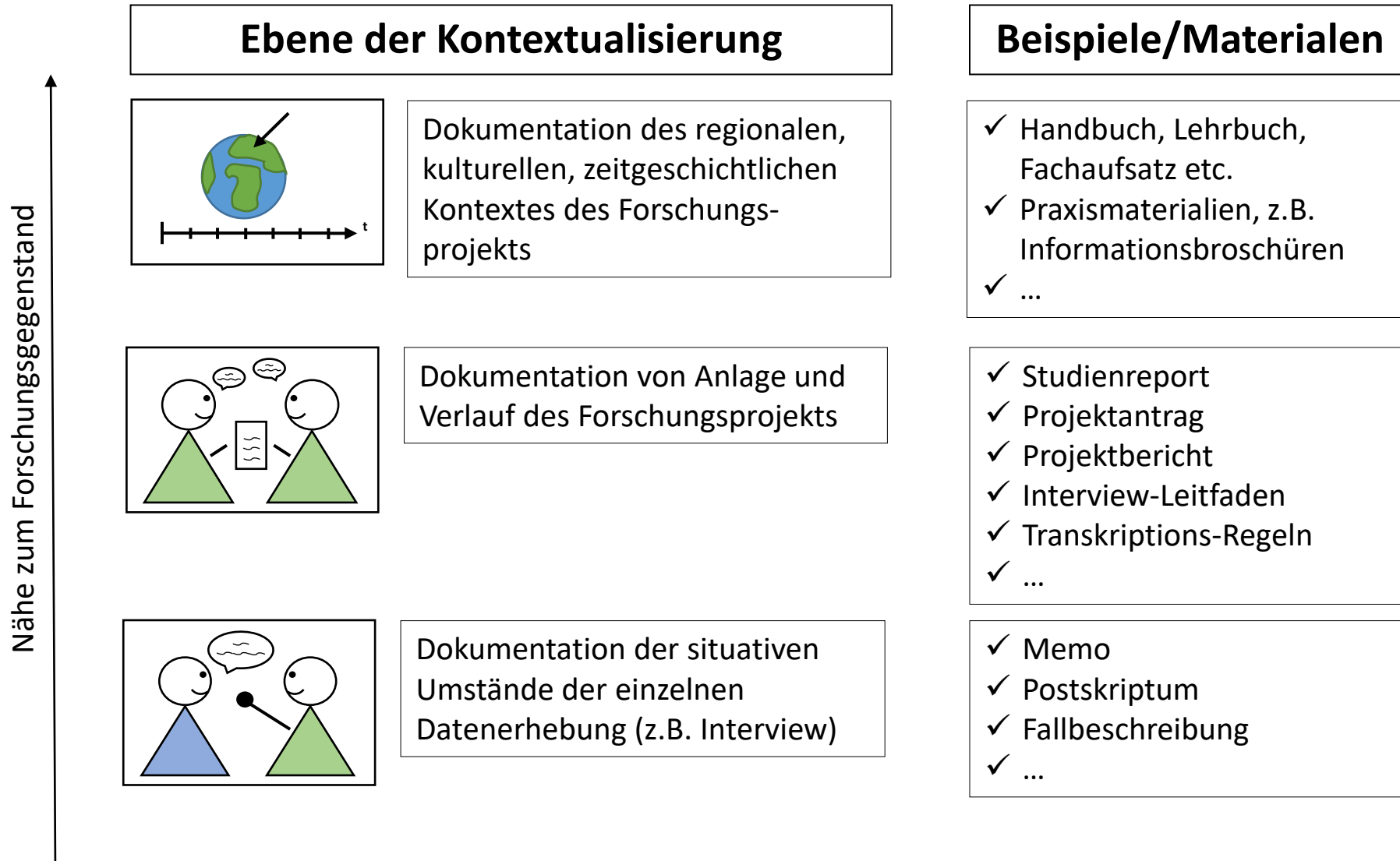
R: I think that he was a salesperson. It's a little bit difficult to tell from what I've read, but he had some kind of business activity, and I had heard, at one time, it was to do with boots and shoes - but the description gets vague. What I am clear about is that the family was Methodist, and my grandmother had converted to Methodism because of my grandfather. I don't know what her religion had been before. But what I do know is that she was -

Thompson, P. (2019). *Pioneers of Social Research, 1996-2018*. [data collection]. 4th Edition. UK Data Service. SN: 6226, [DOI: 10.5255/UKDA-SN-6226-6](https://doi.org/10.5255/UKDA-SN-6226-6)

Dokumentation qualitativer Daten: „Kontextualisierung“

- Dokumentation von qualitativer Forschung: mehr als „nur“ Metadaten
- Begriff für die Übermittlung relevanten Kontextwissens von Datenproduzenten (z.B. Primärforschung) zu Datennutzenden: KONTEXTUALISIERUNG
- Qualitative Forschung zielt auf sinnhaftes Verstehen von Handlungen bzw. Äußerungen in ihrem Bedeutungszusammenhang („Kontext“)
- Kontexte können unterschiedlich gut zu dokumentieren sein (z.B. informelles und intuitives Wissen, körperliches Erleben)
- Es gibt nicht einen Kontext, sondern vielfältige (z.B. institutionelle, inhaltliche, methodische und situative) Kontexte

Ebenen der Kontextualisierung (mit Beispielmaterialien)

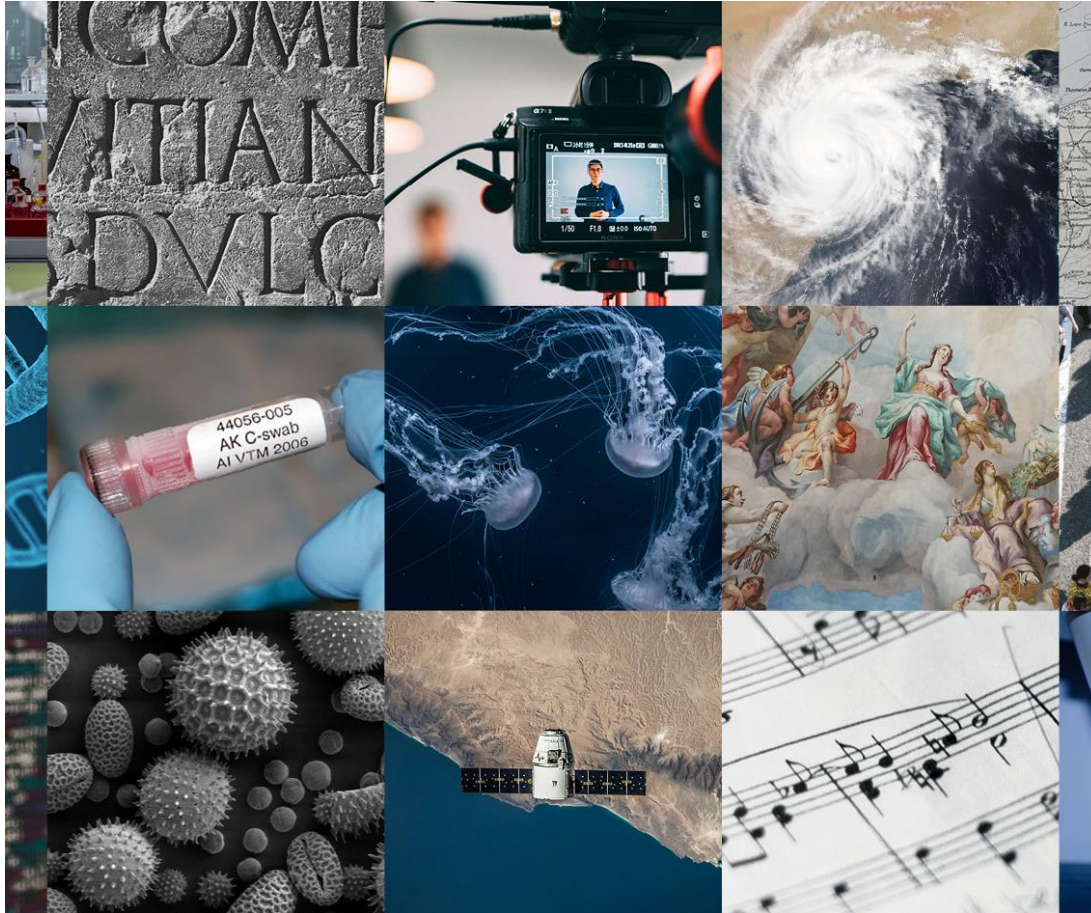


Empfehlungen zur Kontextualisierung qualitativer Daten

- Es gilt, sich in die Situation von nicht am Forschungsprojekt beteiligten Wissenschaftler:innen hineinzusetzen und zu überlegen, welche Informationen diese zum Verständnis der Forschungsdaten benötigen
- Es gilt, eine Balance zu finden zwischen möglichst geringem Aufwand für Primärforschende und möglichst hohem Nutzen für Sekundärnutzende
- Idealerweise fortlaufend (d.h. als Bestandteil des Forschungsprozesses) eine Sammlung von (möglichen) Kontextmaterialien anlegen
- Jeden Schritt des Forschungsprozesses dokumentieren während er durchgeführt wird (weniger Zeitaufwand und höhere Qualität)
- Ideal (nicht nur für das „Data Sharing“): Verfassen eines „Studienreports“

Der „Studienreport“ als zentrales Kontextmaterial

- Für Nachnutzende ist es besonders hilfreich, wenn wesentliche Informationen in einem Dokument vorliegen („Studienreport“, „Daten-Methoden-Bericht“ o.ä.)
- Beim Erstellen können Datengebende in der Regel auf viele bestehende Materialien zurückgreifen (Anträge, Berichte, Publikationen etc.)
- Für Datengebende: Publikation / für Nutzer:innen: „Data fit“ einschätzen
- Vorschlag für die Strukturierung des Studienreports (an Bedarf anzupassen):
 1. Informationen zur Studie und zum Forschungskontext
 2. Datengenerierung: Vorbereitung und Durchführung
 3. Aufbereitung, Auswertung, Nachnutzungspotentiale der Daten
 4. Weitere Kontextmaterialien zur Studie bzw. zu den Daten



Fragen?

Agenda



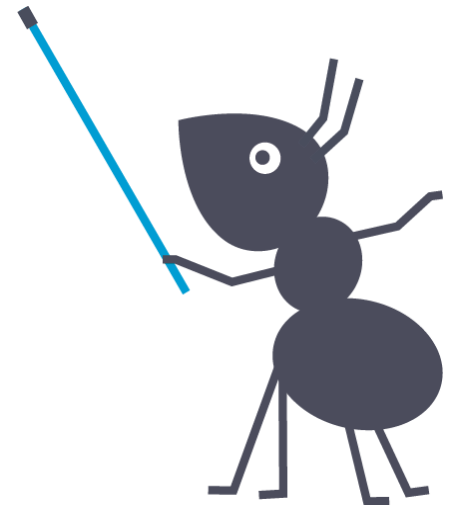
Datenschutz und Anonymisierung



Datendokumentation



Datenarchivierung/-bereitstellung



Ein geeignetes FDZ finden



<https://www.re3data.org/>

KonsortSWD 
Konsortium für die
Sozial-, Verhaltens-, Bildungs- und
Wirtschaftswissenschaften

<https://www.konsortswd.de/datenzentren/alle-datenzentren/>



Auswahlkriterien:

Disziplin-/Datenspezifisches FDZ

Repository der eigenen Institution

Generisches Repository (z.B. Zenodo)

→ Kosten, Anforderungen, Zertifizierung

→ Langzeitarchivierung? Schutz/Zugriff?

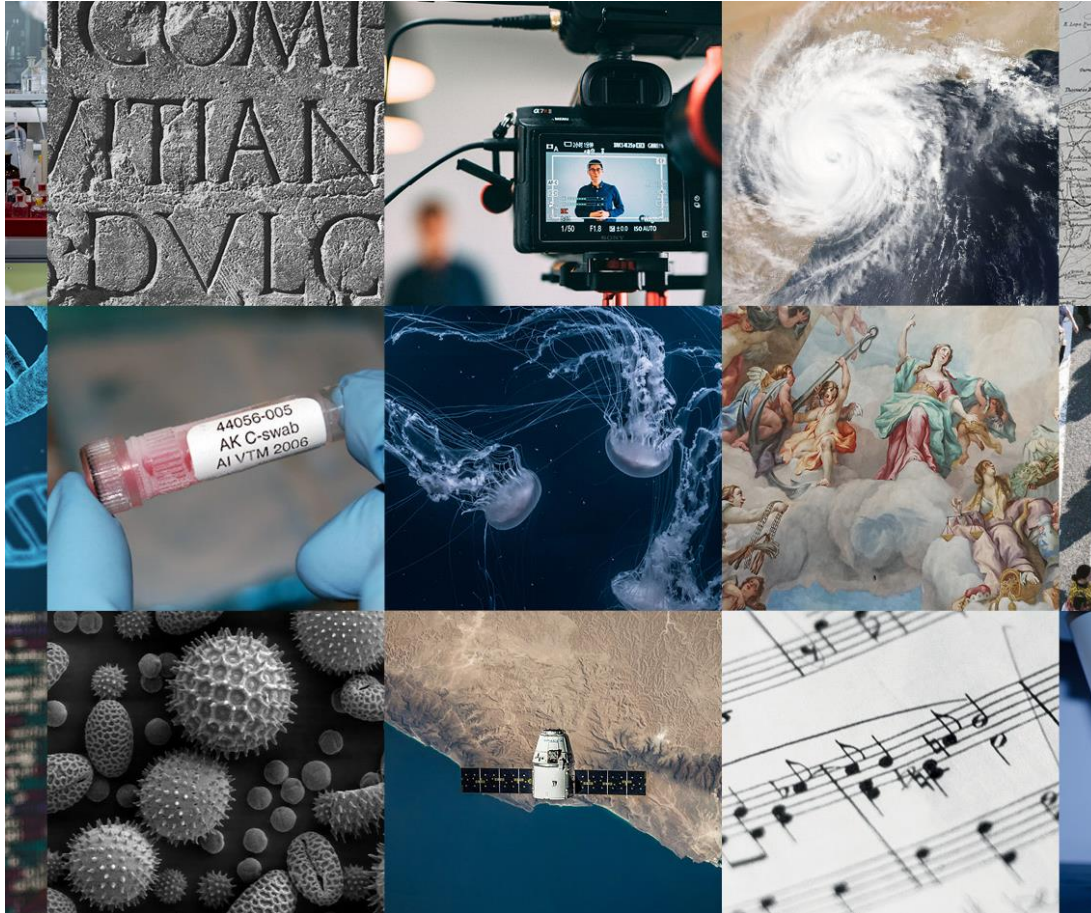
→ Schutz/Zugriff? Qualitätssicherung?

Datenarchivierung als kooperativer Prozess

- Die Archivierung von Forschungsdaten sollte als kooperativer Prozess von Forschenden/Datengebern, FDM-Personal und FDZen geschehen
- Wichtig: Forschende sollten frühzeitig auch mit möglichen FDZen Kontakt aufnehmen (vor Antragstellung!) zur Klärung von Kosten, Anforderungen etc.
- Forschende sollten sich über die Daten und Möglichkeiten/Schwierigkeiten des Data Sharings frühzeitig Gedanken machen (Einwilligungen, Zugangswege,...)
- FDZen und FDM-Personal können bei der Datenvorbereitung unterstützen
- FDZen bieten Vorlagen, Tools, Handreichungen, Schulungsangebote an
- FDZen bieten u.U. umfangreiche Kuration an (z.B. Anonymisierungsprüfung)

Bsp.: Archivierung beim FDZ Qualiservice

- Forschende nehmen vor Antragstellung Kontakt auf → Erstberatungsgespräch (über Projekt, Daten, QS-Leistungen) → QS stellt Kostenangebot aus
- Forschende bereiten Daten im Forschungsprojekt vor; QS bietet fortlaufend Beratung und Unterstützung (ergänzend: Materialien, Schulungsangebote)
- Forschende entscheiden, welche Daten und Kontextmaterialien in welcher Form archiviert und bereitgestellt werden sollen (Datenübergabevereinbarung)
- Nach Datenübergabe kuratiert QS die Daten (u.a. Anonymisierungsprüfung); offene Fragen werden mit den Forschenden besprochen (ggf. „Proofread“)
- Metadaten und Studienreport werden veröffentlicht (jeweils mit DOI); Daten sind nach Unterzeichnung einer Nutzungsvereinbarung zugänglich für wissenschaftliche Nutzungen in Forschung und/oder Lehre (SUF, CF, Onsite)



Fragen?

Genutzte Materialien

Literatur

- Ebel, T., Meyermann, A. (2015): Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdatenzentrum (FDZ) Bildung am DIPF. Frankfurt am Main (forschungsdaten bildung informiert, 3). <https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-3.pdf>.
- European Union (2016). Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung, DSGVO). <https://dsgvo-gesetz.de/>. Groves R., Fowler, F., Couper, M., Singer, E. & Tourangeau, R. (2004). Survey Methodology, New York, Wiley.
- Heuer, J.-O.; Kretzer, S.; Mozygemba, K.; Huber, E.; Hollstein, B. (2020): Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: eine Handreichung für Forschende zur Erstellung eines Studienreports. Forschungsdatenzentrum (FDZ) Qualiservice, Universität Bremen. Bremen (Qualiservice Working Papers, 1-2020). <https://doi.org/10.26092/elib/166>.
- Hollstein, B., Miller, B., Siegers, P., Wolf, Ch. (2021): KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung. In: Bausteine Forschungsdatenmanagement 2/2021: S. 10-22. <https://doi.org/10.17192/bfdm.2021.2.8330>.
- Jeding, A.; Watteler, O.; Förster, A. (2018): Improving the Quality of Survey Data Documentation: A Total Survey Error Perspective. Data, 3, 45. <https://doi.org/10.3390/data304004>.
- Jensen, Uwe; Netscher, Sebastian & Weller, Katrin (2019). Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten - Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten. Verlag Barbara Budrich. Open Access verfügbar unter <https://shop.budrich.de/produkt/forschungsdatenmanagement-sozialwissenschaftlicher-umfragedaten>.
- Medjedović, I. (2019): Qualitative Daten für die Sekundäranalyse. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung. 2. Aufl. Wiesbaden: Springer VS, S. 247–258. https://doi.org/10.1007/978-3-658-21308-4_16.
- Meyermann, A.; Porzelt, M. (2014): Hinweise zur Anonymisierung von qualitativen Daten. Forschungsdatenzentrum (FDZ) Bildung am DIPF. Frankfurt am Main (forschungsdaten bildung informiert, 1). <https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-1.pdf>.
- Mozygemba, K.; Hollstein, B. (2023) Anonymisierung und Pseudonymisierung qualitativer textbasierter Forschungsdaten – eine Handreichung. Qualiservice Working Papers 5, Bremen: Forschungsdatenzentrum Qualiservice. <https://doi.org/10.26092/elib/2525>.
- Orten, H., Mills, H., Perry, A. (2021, March 5). What can the Data Documentation Initiative (DDI) do for comparative surveys?. Zenodo. <https://doi.org/10.5281/zenodo.4584652>.
- Steinhardt, I. et al. (2020): Das Öffnen und Teilen von Daten qualitativer Forschung: eine Handreichung. Ergebnisse eines Workshops der Forschungsgruppe "Digitalisierung der Wissenschaft" am Weizenbaum-Institut in Berlin am 17. Januar 2020. Weizenbaum-Institut. Berlin (Weizenbaum Series, 6). <https://doi.org/10.34669/wi.ws/6>.
- ZBW, GESIS & RatSWD (2014). Auffinden, Zitieren, Dokumentieren: Forschungsdaten in den Sozial- und Wirtschaftswissenschaften, <https://doi.org/10.4232/10.fisuzida2015.2>.

Quellen & Bilder

- European Values Study (EVS) 2022: Integrated Dataset (EVS 2017). GESIS, Köln. ZA7500 Datenfile Version 5.0.0, Fragebogen. <https://doi.org/10.4232/1.13897>.
- Pixabay (2023). Verfügbar unter: www.pixabay.com. (Pixabay Content License).
- Statistische Bundesamt (2022). Ausländerstatistik. Verfügbar unter: <https://www-genesis.destatis.de/genesis/online?sequenz=statistikTabellen&selectionname=12521#abreadcrumb>.
- The Comparative Study of Electoral Systems (2015). CSES Module 3 Full Release. GESIS Datenarchiv, Köln. ZA5181 Datenfile Version 5.0.0. <https://doi.org/10.7804/cses.module3.2015-12-15>.

Webseiten

- Datenmanagement Expert Guide von CESSDA: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>.
- FDZ Qualiservice: <https://www.qualiservice.org/de/>.
- GESIS - Forschungsdaten managen: <https://www.gesis.org/datenservices/daten-managen>. KonsortSWD: <https://www.konsortswd.de>.
- Verbund Forschungsdaten Bildung: <https://www.forschungsdaten-bildung.de/datenmanagement>.