

# Aufbereitung und Anonymisierung qualitativer Forschungsdaten

Nadeshda Jung

Reihe „Forschungsdatenmanagement in  
der empirischen Bildungsforschung“

# Kernziele dieses Vortrages

- » Was ist die Besonderheit von qualitativen Daten?
- » Was sollte bei der Durchführung der Anonymisierung beachtet werden?
- » Warum ist es wichtig, qualitative Daten zu kontextualisieren?
- » Hilfreiche Tipps und was man beachten sollte

Besonderheiten qualitativer Daten

# **Datenpluralität und Informationsdichte**

# Beispiele von qualitativen Daten der Bildungsforschung

- » Audiovisuelle Forschungsdaten (z. B. Videoaufzeichnungen von beobachteten Unterrichtssituationen)
- » Textuelle Forschungsdaten (z. B. Transkripte von Interviews und Gruppendiskussionen)
- » Analysematerialien (z. B. Codierungen für videografierte Situationen)
- » Dokumentationsmaterialien (z. B. Interviewleitfäden, Manuale der Videografie und Transkription)



# Besonderheit qualitativer Daten: Informationsdichte

» Hohes Analysepotenzial der Daten für Sekundärforschungen

aber auch

» hoher Personenbezug innerhalb der Daten

➔ Die Anonymisierung qualitativer Daten kann sehr aufwendig sein und teilweise auch nicht sinnvoll, wenn Informationsgehalt und Analysepotenzial der Daten dadurch zerstört werden (z. B. bei Audio- und Videodateien).

Anonymisierung von qualitativen Daten

# Hintergrund und Vorgehen

# Definition Anonymisierung

Anonymisierung bezeichnet das Entfernen personenbezogener oder personenbeziehbarer Informationen aus den Forschungsdaten.

In der Regel gilt es persönliche Daten zu anonymisieren, sobald es der Forschungszweck erlaubt.

# Unterschiede Pseudonymisierung und Anonymisierung

## » **Pseudonymisierung** nennt man...

„die Verarbeitung personenbezogener Daten in einer Weise, dass [diese] ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können [...]“ (Art. 4.5 DSGVO).

## » **Anonymisierung:**

„[...] das Verändern personenbezogener Daten derart, dass [diese] nicht mehr oder nur mit einem unverhältnismäßig große[m] Aufwand an Zeit, Kosten, Arbeitskraft einer [...] Person zugeordnet werden können“ (Art. 36.6 DSG NRW)

# Was muss man anonymisieren?

- » Direkte personenbezogene und personenbeziehbare Merkmale:  
Personennamen, Ortsangaben, Straßennamen, Bundesländer, Institutionen und Organisationen (z. B. Firmen, Schulen, Institute)
  
- » Indirekte, aber spezifische personenbezogene/-beziehbare Merkmale:  
Merkmale der Probanden wie auch solche von Dritten, in den Interviews erwähnten Personen
  
- » Sensible Informationen:  
Informationen zur ethnischen Herkunft, politischen Meinung, religiösen oder philosophischen Überzeugung, Gewerkschaftszugehörigkeit, Gesundheit oder zum Sexualleben

# Beispiel für Anonymisierung bei Interviewtranskripten

» Löschung der Merkmale und Ersetzung durch Platzhalter

Originalangabe, Identifikator		(1) starke Abstraktion
Name	Aishe	PERSON
	Christian	
Alter	27	ALTER
Ort	Gelsenkirchen	ORT
	Berlin-Neukölln	
Bundesland	Sachsen	BUNDESLAND
Arbeitgeber	Bäckerei Schmidt	ARBEITGEBER
Beruf	Friseur	BERUF
Schule	Gymnasium Hochstadt	SCHULE
Sonstige Organisationen / Vereine	Johannes-Kirche	ORGANISATION / VEREIN
Datumsangaben	12. 01. 1998	TAG

Meyermann & Porzelt 2014, S. 7

# Beispiel für Anonymisierung bei Interviewtranskripten

## » Aggregation von Informationen

Originalangabe, Identifikator		(3) Aggregation
Name	Aishe	Frau
	Christian	Mann
Alter	27	20 bis 30 Jahre alt
Ort	Gelsenkirchen	Ruhrgebiet
	Berlin-Neukölln	Migrantenbezirk einer Großstadt
Bundesland	Sachsen	Ostdeutschland
Arbeitgeber	Bäckerei Schmidt	Handwerk (Branchenangabe)
Beruf	Friseur	Handwerker (z. B. Angabe des ISCO-Codes)
Datumsangaben	12. 01. 1998	Zeitraum zwischen 1990 und 2000; 01/1998

Meyermann & Porzelt 2014, S. 8

# Beispiel für Anonymisierung bei Interviewtranskripten

- » Ersetzen durch Merkmale mit vergleichbarer Bedeutung

Originalangabe, Identifikator		(4) Ersetzung
Name	Aishe	Freundin
	Christian	Bruder
Alter	27	Person der gleichen Alterskohorte/Generation
Ort	Gelsenkirchen	Wohnort, Arbeitsort, Ausbildungsort
	Berlin-Neukölln	
Bundesland	Sachsen	Bundesland des Geburtsorts

Meyermann & Porzelt 2014, S. 8

# Beispiel für Anonymisierung bei audiovisuellen Daten

## » Akustische Veränderung:

- Nennung von Namen mit einem Piepton überdecken
- Rauschsignal, um das Wiedererkennen von Stimmen zu verhindern
- Tonhöhenveränderung (Senken oder Heben von Stimmen)

## » Visuelle Manipulation:

- In Videodateien werden Gesichter mit Hilfe von Verpixelung oder Balkenüberlagerung verfremdet.

Optionen bei der Anonymisierung

# **Hinweise und Best Practices**

# Problematik in der Durchführung der Anonymisierung

- » Kosten- und zeitaufwendig
- » Zerstörung des Analysepotenzials für eine Sekundäranalyse



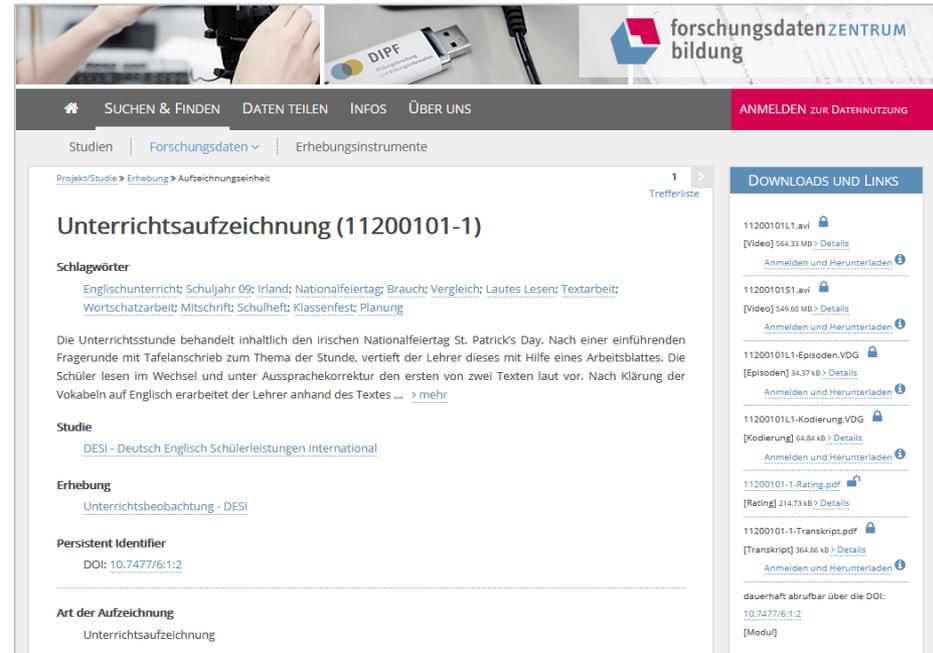
art\_inthecity (o.D.): Diane Gougeon, Les iris, Art souterrain 2014, Montréal. licensed under CC BY 2.0  
<https://ccsearch.creativecommons.org/photos/1c93c849-29fc-42ce-8964-1e3f33c7bd21>

# Alternative zur Anonymisierung

- » Informierte Einwilligung im Vorfeld einholen, damit die Daten auch nicht anonymisiert genutzt werden können.
- » Während der Erhebung schon vermeiden, personenbezogene Informationen aufzuzeichnen (z. B. durch Umsetzen der Schüler\*innen im Klassenzimmer).
- » Im Vorfeld mit den Proband\*innen vereinbaren, identifizierbare Merkmale nicht zu erwähnen – sofern dies nicht die Aussagefähigkeit beeinträchtigt.
- » **Weitere Maßnahme:** Zugang zu diesen Daten über kontrollierte Zugangswege und nur für bestimmte Nutzer\*innengruppen über ein Forschungsdatenzentrum

# Beispiel: Sichere Zugangswege

- » Daten liegen in einem **geschützten Verzeichnis**
- » **Nach Registrierung zusätzlich Nutzungsantrag** (Online-Formular)
- » **Einzelprüfung** jedes Antrags, PostIdent-Verfahren
- » **Nutzungsvertrag:** für max. drei Jahre mit Verlängerungsmöglichkeit, strikte Bindung an Vertragsauflagen
- » **Berechtigte** bei nicht-anonymisierten Daten:
  - nur für ausgewiesene qualifizierte Wissenschaftler\*innen
  - nur für definierten Forschungszweck
  - Bedingung: Forschung ist auf spezifische Daten angewiesen
  - **NICHT** für die Lehre



The screenshot shows the 'forschungsdatenZENTRUM bildung' website. The main content area displays details for 'Unterrichtsaufzeichnung (11200101-1)'. It includes sections for 'Schlagwörter', 'Studie' (DESI - Deutsch Englisch Schülerleistungen International), and 'Erhebung' (Unterrichtsbeobachtung - DESI). A 'Persistent Identifier' (DOI: 10.7477/6:1:2) is provided. On the right, a 'DOWNLOADS UND LINKS' sidebar lists several files for download, including video files (11200101L1.avi, 11200101S1.avi), an episode (11200101L1-Episoden.VDG), a coding file (11200101L1-Kodierung.VDG), a rating file (11200101-1-Rating.pdf), and a transcript (11200101-1-Transkript.pdf). Each file entry includes its size and a 'Details' link. At the bottom of the sidebar, it states 'dauerhaft abrufbar über die DOI: 10.7477/6:1:2 [Modul]'.

Zugriffsstufen des FDZ Bildung

# Qualitative Daten der Bildungsforschung am FDZ Bildung des DIPF (Partner im VerbundFDB)

Bereitstellung aktuell

- » **Datenbestände aus 12 Studien** (z. B. bildungshistorisch „Audiovisuelle Aufzeichnungen von Schulunterricht in der DDR“ oder große, nationale Studien wie „Pythagoras: Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“ oder „DESI – Deutsch Englisch Schülerleistungen International“)
- » **2100 Video- und Audio-Dateien**
- » **2500 Textdokumente** (Transkripte von Unterrichtsaufzeichnungen/Interviews, Codierungen, Episoden, Ratings)

# Praktische Tipps bei der Durchführung der Anonymisierung

- » Anonymisieren während des Prozesses der Auswertung durchführen (z. B. beim Transkribieren)
- » Anonymisierte und sensible Passagen kennzeichnen
- » Anonymisierungsprotokoll anfertigen



Besonderheiten qualitativer Daten

# **Kontextsensitivität**

# Nachnutzbarkeit und Interoperabilität von anonymisierten Daten

- » Zur Aufbereitung der Daten gehören sowohl Anonymisierungsprinzipien als auch Kontextualisierungen
- » Um wichtige Informationen für die Verwendung der Daten und die Qualität der Daten nicht zu verlieren, sollte – neben der eigentlichen Erhebung und Aufbereitung des Materials sowie der Erhebung von soziodemografischen Daten – ein zusätzliches Dokument zur Kontextualisierung verfasst und als essenzieller Bestandteil des Prozesses gesehen werden.
- » Qualitativer Daten sind situational gerahmt und das Ergebnis von Interaktionen und wechselseitigen Interpretationen.

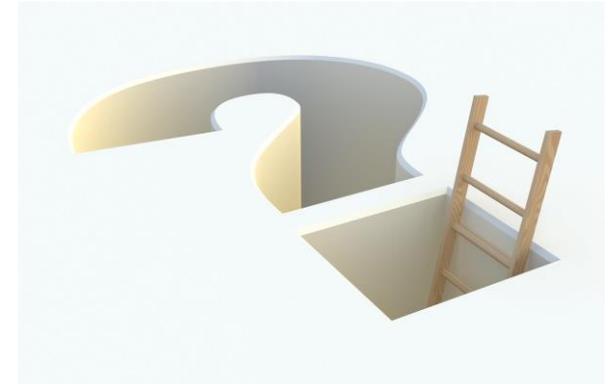
 Erst eine umfangreiche und ausführliche Dokumentation ermöglicht Sekundärnutzer\*innen, die Perspektive der Primärforschenden nachzuvollziehen.

# Problematik der Kontextsensitivität

» Oftmals wird das für die Primärforschenden aus dem Forschungszusammenhang abrufbare und daher selbstverständliche Kontextwissen nicht dokumentiert.

➔ Daten- und Projektbeschreibungen oft mangelhaft

➔ Rekonstruktion des Forschungsprozesses und Nachnutzung der Daten erschwert oder nicht möglich



3D Problem Solving (2012): Chris Potter licensed under CC BY 2.0 .  
<https://ccsearch.creativecommons.org/photos/01da9e81-d484-4ded-8f91-18ac0a1c7d3c>

Ebene Studie (Makro)	Ebene Setting (Meso)	Ebene Objekt (Micro)
<p><b>Hintergrund</b> der Studie (Projektleiter*in, Mitarbeiter*innen etc.)</p> <p><b>Beschreibung der Methodik</b> und des <b>Prozesses der</b> <b>Datenerhebung</b> (Erhebungsinstrumente, Stichprobenauswahl/Feld- zugang, Ort und Zeitpunkt der Erhebung etc.)</p> <p><b>Rechtliche Aspekte</b> (Vertraulichkeit, Zugangsregelungen und Nutzungsbedingungen, Einverständniserklärungen)</p>	<p><b>Metainformationen</b> über das Gespräch als soziale Situation oder das <b>Setting</b> (z. B. Interviewsituation)</p> <p><b>Informationen zu den</b> <b>untersuchten</b> <b>Einheiten/</b>Probanden (Alter, Geschlecht, Bildung etc.)</p>	<p><b>Datenaufbereitung</b> (z. B. Vorgehen beim Transkribieren und Anonymisieren)</p> <p><b>Modifikationen</b> an den Daten, Versionierungen</p> <p><b>Anzahl und Beziehungen</b> der Dateien (z. B. Aufzeichnungen) untereinander</p> <p><b>Informationen zur Datei</b> selbst (Dateiname, Format, Größe etc.)</p>

# Beispiel für die Kontextualisierung von Interviewtranskripten

Kontextualisierungsbogen	
<b>Allgemeine Daten</b>	
Code der Daten	<i>1234_ABC</i>
Projekt/Studie/Teilstudie	<i>Teilprojekt 1 - Frankfurt</i>
Dateien	<i>Audiodatei, Transkript</i>
Datum	<i>01.01.2018</i>
Länge der Aufnahme	<i>1:23.45 min</i>
Ort	<i>Büro der Schulverwaltung (den Teilnehmenden nicht bekannt)</i>
Thema	<i>Erfahrungsberichte der internen Personalentwicklung in der Grundschule</i>
<b>Aufnahmesituation</b>	
Typ/ Methode	<i>Gruppendiskussion</i>
Aktive Teilnehmende	<i>Drei Schulleiter, zwei Schulleiterinnen; eine Diskussionsleiterin (Projekt)</i>
Passive Teilnehmende	<i>Ein Protokollant (Projekt)</i>
Anmerkung zu den Teilnehmern	<i>Die Diskussionsteilnehmenden sind untereinander nicht bekannt/ arbeiten nicht zusammen</i>
Aufnahmestatus	<i>Offene Aufnahme</i>
Verwendete Materialien während des Gesprächs	<i>Leitfaden Teilprojekt Frankfurt (Diskussionsleiterin); Vorlage: Personalentwicklungsbogen (alle Teilnehmenden)</i>

Materialien mit Hilfestellung finden Sie unter <https://www.forschungsdaten-bildung.de/publikationsreihen>

Leitfaden zur Kontextualisierung von qualitativen Befragungsdaten [https://www.forschungsdaten-bildung.de/files/fdbinfo\\_9\\_Kontextualisierung-von-qualitativen-Befragungsdaten\\_v1.0.pdf](https://www.forschungsdaten-bildung.de/files/fdbinfo_9_Kontextualisierung-von-qualitativen-Befragungsdaten_v1.0.pdf)

# Hinweise zum Vorgehen bei der Aufbereitung und Anonymisierung

- » Zeitpunkt: prozessbegleitend
- » Zuständigkeit: Es sollte zu Beginn der Forschung geklärt werden, wer für die Dokumentation verantwortlich ist.
- » Beantragung von Geldmitteln: Bei den Förderorganisationen können Gelder speziell für die Aufbereitung der Daten beantragt werden.
- » Nehmen Sie die Perspektive eines Sekundärnutzenden ein. Könnte diese\*r Ihre Daten auf Grundlage der Dokumentation interpretieren? Sind die Daten einfach zu de-anonymisieren?



Niederberger, U. (o.D.): Bulb Fiction. is licensed under CC BY 2.0  
<https://ccsearch.creativecommons.org/photos/71fc18e6-5eea-4706-b1bf-2654ab836d99>

# Literatur

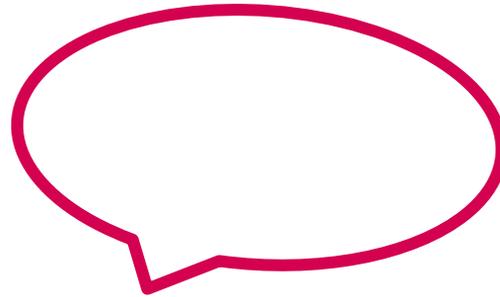
Metschke, R. & Wellbrock, R. (2002). Datenschutz in Wissenschaft und Forschung (Materialien zum Datenschutz.28). Verfügbar unter: <https://www.hu-berlin.de/de/datenschutz/einwilligung/datenschutz-in-wissenschaft-und-forschung>(abgerufen am 27.10.2014)

Meyermann, A. & Porzelt, M. (2014). Hinweise zur Anonymisierung von qualitativen Daten. Verfügbar unter: [https://www.forschungsdaten-bildung.de/get\\_files.php?action=get\\_file&file=fdb-informiert-nr-1.pdf](https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-1.pdf) (abgerufen am 18.02.2020)

Jung, N; in Zusammenarbeit mit Porzelt, M.; Kretzer, S. & Berdelmann, K. (2019). Kontextualisierungsbogen von qualitativen Daten. Leitfaden zur Kontextualisierung für Befragungsmethoden. Version 1.0, fdbinfo Nr. 9.

Vielen Dank für Ihre Aufmerksamkeit!

# Abschließende Fragerunde



# Abschluss

**Haben Sie weitere Fragen oder Anregungen zum Thema?**

Nadeshda Jung

E-Mail: [jung@dipf.de](mailto:jung@dipf.de)

Telefon: +49 (0)69 - 24708 - 309

**Haben Sie Fragen zum Forschungsdatenmanagement?**

Verbund Forschungsdaten Bildung

E-Mail: [verbund@forschungsdaten-bildung.de](mailto:verbund@forschungsdaten-bildung.de)

Telefon: +49 (0)69 - 24708 - 300

