

Grundlagen, Strategien und Techniken der Anonymisierung von Transkripten in der qualitativen Forschung: eine praxisorientierte Einführung

Werner, Cosima; Meyer, Frank; Bischof, Susann

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Werner, C., Meyer, F., & Bischof, S. (2023). Grundlagen, Strategien und Techniken der Anonymisierung von Transkripten in der qualitativen Forschung: eine praxisorientierte Einführung. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 24(3). <https://doi.org/10.17169/fqs-24.3.4067>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

Grundlagen, Strategien und Techniken der Anonymisierung von Transkripten in der qualitativen Forschung: eine praxisorientierte Einführung

Cosima Werner, Frank Meyer & Susann Bischof

Keywords:

Anonymisierung;
personen-
bezogene Daten;
Datenschutz;
Transkript;
Vertraulichkeit;
Datenschutz-
Grundverordnung;
DSGVO

Zusammenfassung: In der qualitativen Forschungspraxis müssen beim Datenschutz die gesetzlichen Rahmenseetzungen beachtet werden; damit einhergehende Fragen der Anonymisierung gehen für qualitativ Forschende jedoch darüber hinaus. Denn will man Beforschte vor negativen Folgen durch eine Teilnahme schützen und dabei zugleich die eigenen Erkenntnismöglichkeiten ausschöpfen, müssen fall- und kontextspezifische Textveränderungen und andere Maßnahmen in den verschiedenen Phasen empirischer Praxis abgewogen werden. Dabei stellt das konkrete Vorgehen des Anonymisierens selbst jedoch häufig eine Blackbox dar, die in wissenschaftlichen Veröffentlichungen eher rudimentär Beachtung findet. Mit diesem Artikel stellen wir daher nicht nur die aktuellen rechtlichen Rahmenbedingungen dar, sondern fokussieren insbesondere die praktischen Aspekte des Anonymisierens von Transkripten. Mit dem Ziel, einen grundlegenden, zugänglichen und praktisch orientierten Einstieg in das Thema zu geben, systematisieren wir zunächst Datenarten, Anonymisierungsstrategien, und -techniken. Anschließend diskutieren wir praktische Umsetzungsmöglichkeiten und geben Interessierten die Möglichkeit, ein fiktives Transkript selbstständig zu bearbeiten bzw. in Workshops zum Thema einzubinden. Abschließend präsentieren wir grundlegende Schlussfolgerungen für die Methodenausbildung und plädieren für einen offenen kollegialen Austausch zu Anonymisierungspraktiken.

Inhaltsverzeichnis

- [1. Zur Anonymisierung qualitativer Forschungsdaten](#)
- [2. Rechtliche Grundlagen und offene Fragen](#)
- [3. Datenarten](#)
- [4. Die Praxis der Anonymisierung](#)
 - [4.1 Anonymisierung im Forschungsprozess](#)
 - [4.2 Anonymisierungsstrategien](#)
 - [4.3 Anonymisierungstechniken](#)
 - [4.4 Darstellung in Veröffentlichungen](#)
 - [4.5 Dokumentation](#)
 - [4.6 Zielgruppen](#)
- [5. Anwendungsbeispiel](#)
 - [5.1 Aufgabenstellung](#)
 - [5.2 Schritt 1: Markieren unterschiedlicher Datenarten](#)
 - [5.3 Schritt 2: Eruiieren möglicher Anonymisierungstechniken](#)
- [6. Schlussfolgerungen: Anonymisierung lehren und lernen](#)
 - [6.1 Programmatische Schlussfolgerungen](#)
 - [6.2 Methodische Schlussfolgerungen](#)

[Anhang](#)

[Literatur](#)

[Zu den Autorinnen und zum Autor](#)

[Zitation](#)

1. Zur Anonymisierung qualitativer Forschungsdaten

Die Anonymisierung personenbezogener Daten stellt in der Sozialforschung ein essenzielles Mittel dar, um Beforschte vor negativen Folgen ihrer Teilnahme an Erhebungen zu schützen. Durch Gesetze wie die europäische Datenschutz-Grundverordnung (DSGVO)¹, das Bundesdatenschutzgesetz (BDSG)² und entsprechende Landesdatenschutzgesetze (LDSG)³ werden u.a. Forschende zum Schutz personenbezogener Daten verpflichtet. Die wichtigsten Instrumente des Datenschutzes sind hierbei die Einverständniserklärung der Teilnehmenden und die Anonymisierung personenbezogener Daten. [1]

In diesem Beitrag fokussieren wir den Umstand, dass Anonymisierung nicht bei einer Änderung von Namen aufhört, sondern in ihrer Komplexität und Bedeutung darüber hinaus geht. Anonymisierung stellt ein "unproblematisches Problem" (LOCHNER 2017, S.284) dar, also einen Sachverhalt, der oberflächlich simpel erscheint, jedoch bei näherer Beschäftigung blinde Flecken und Komplexitäten aufweist. Besonders bei qualitativen Studien mit wenigen Teilnehmenden, die ggf. in kleinen Orten oder überschaubaren Gemeinschaften durchgeführt werden, ist Anonymisierung aufwendig (HOPF 2016 [2010], S.200-201; VAINIO 2013, S.687). Durch sie wird ein "Dilemma [eröffnet] zwischen einer möglichst präzisen Darstellung der erhobenen Daten und einem Schutz der beforschten natürlichen oder juristischen Personen" (KÜHL 2020, S.63). [2]

Dieses Dilemma wird von einer Theorie-Praxis-Lücke begleitet: Zwar ist für viele qualitativ Forschende Vertraulichkeit ein selbstverständlicher ethischer Standard (SURMIAK 2018, §54); jedoch gleicht Anonymisierung praktisch oftmals einer Blackbox. In gesetzlichen Regelwerken und allgemeinen Grundsätze der Forschungsethik werden keine konkreten Umsetzungsanweisungen geliefert. Aus Forschungsperspektive werden zudem mitunter Möglichkeit und Nutzen einer "generalisierbare[n] Methode der Anonymisierung von Personendaten" (LOCHNER 2017, S.293) bezweifelt. Gleichzeitig bilden Forschende intuitiv situationsbezogene Praktiken aus, die – obwohl sie von einem transdisziplinären Austausch profitieren würden – selten ausführlich in Methodenlehrbüchern oder Journalartikeln Platz finden. [3]

1 Verordnung (EU) 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG. *Amtsblatt der Europäischen Union*, <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679> [Datum des Zugriffs: 31. März 2023].

2 Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680, 30. Juni 2017. *Bundesgesetzblatt*, <https://dsgvo-gesetz.de/bdsg/> [Datum des Zugriffs: 31. März 2023].

3 Bspw. das Gesetz zum Schutz personenbezogener Daten im Land Brandenburg, 15. Mai 2018, <https://bravors.brandenburg.de/gesetze/bbgdsg> [Datum des Zugriffs: 31. März 2023].

Die Blackbox hat also zwei Dimensionen: Einerseits kann, paradigmatisch, Anonymisierung kaum vereinheitlichend dargestellt werden, weil die damit verbundenen praktischen Fragen des Umganges mit kontextspezifischen Daten zu komplex sind. Andererseits wird Anonymisierung aktuell vor allem aus rechtlicher Sicht betrachtet, während sich für ihre Praxis kaum dedizierte Foren finden, in denen sie diskutiert, abgewogen und geübt wird. [4]

Aus diesem Grund wollen wir mit diesem Artikel einen (gerne für die Lehre einsetzbaren) Einstieg in Anonymisierungspraktiken geben, mit dem wir dazu einladen, selbst ein Beispieltranskript zu bearbeiten. Der vorliegende Beitrag ist dementsprechend als eine theoretisch unterfütterte *Hands-on*-Anleitung zum Anonymisieren von Transkripten zu verstehen. Dabei greifen wir insbesondere auf Studien oder Berichte zurück, in denen Anonymisierung aus der Sicht qualitativ arbeitender Wissenschaftler:innen behandelt wird. Auf teilautomatisierte Anonymisierungstools (MOZYGEMBA 2022) gehen wir nicht ein. Aufbauend auf dem Workshop "Anonymisierung in qualitativen Forschungsarbeiten" anlässlich der [6. Jahrestagung](#) des Arbeitskreises [Qualitative Methoden in der Geographie und raumsensiblen Sozial- und Kulturforschung](#) werden wir aktuelle Diskussionsstränge zu Anonymisierung zusammenzufassen, Kernaspekte systematisieren und Umsetzungsweisen diskutieren. Forschungsbezogene Anonymität ist die Abwesenheit von erkennbaren Identitätsmerkmalen von Personen in wissenschaftlichen Daten. Den Begriff Anonymisierung verwenden wir hier als Nutzung datenverändernder Techniken zur Erschwerung der Identifizierung dieser Personen (SAUNDERS, KITZINGER & KITZINGER 2015). [5]

In einem ersten Schritt fassen wir die für den deutschen Kontext maßgeblichen rechtlichen Grundlagen der Anonymisierung zusammen (Abschnitt 2). Im Anschluss werden verschiedene Datenarten unterschieden: Durchführungs-, Identifikations-, Kontext- und sensible Daten wie Gesundheitsdaten oder sexuelle Orientierung, die in Interviews oder Gruppendiskussionen geäußert werden. Danach verorten wir das Anonymisieren im Forschungsprozess, skizzieren zentrale Strategien, diskutieren konkrete Techniken, die Darstellung von Anonymisierungsmaßnahmen in Veröffentlichungen sowie ihre Dokumentation als Teil des Datenmanagements und loten aus, an wen sie sich richtet (Abschnitt 3). Nachfolgend bieten wir interessierten Leser:innen die Möglichkeit, anhand eines fiktiven Transkripts diese Grundlagen in die Praxis zu übertragen (Abschnitt 4). Wir schließen mit Schlussfolgerungen zum Lehren und Lernen von Anonymisierung als einer Forschungspraxis (Abschnitt 5). [6]

2. Rechtliche Grundlagen und offene Fragen

Mit der im Mai 2018 in Kraft getretenen DSGVO wird der Datenschutz für die Mitgliedstaaten der EU vereinheitlicht. Dies führt allerdings zu Herausforderungen für mit qualitativen Forschungsdesigns arbeitende Wissenschaftler:innen, denn nach der DSGVO (Art.9) wird die Verarbeitung personenbezogener Daten verboten:

"Die Verarbeitung personenbezogener Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person ist untersagt". [7]

In vielen sozialwissenschaftlichen Erhebungen sind ethnische Zugehörigkeit, sexuelle Orientierung, Geschlechtsidentität oder politische Haltungen jedoch Untersuchungsgegenstände oder zentrale Analysekatoren. Hierfür gibt es nach der DSGVO keine Sonderregeln im Umgang mit personenbezogenen Daten. Stattdessen wird die Umsetzungskompetenz vonseiten der EU an die Mitgliedsstaaten übertragen. Die DSGVO wird um das deutsche Bundesdatenschutzgesetz ergänzt; demnach dürfen öffentliche Einrichtungen für den Zweck der Forschung personenbezogene Daten verarbeiten (BDSG, Art.3). Somit handelt es sich um ein "Verbot mit Erlaubnisvorbehalt", durch das Möglichkeiten für wissenschaftliche Studien eröffnet werden. Forscher:innen müssen sich zudem an das für sie geltende Landesdatenschutzgesetz halten, welches wiederum durch das BDSG präzisiert wird. [8]

Der hier relevante Fokus in den Gesetzen bezieht sich auf den Prozess der Verarbeitung personenbezogener Daten. Laut DSGVO (Art. 4) umfasst die Verarbeitung deren Sammeln, Erfassen, Organisieren, Speichern, Anpassen, Abfragen, Verwenden, Übermitteln, Verknüpfen, Einschränken, Löschen oder Zerstören. In der Forschungspraxis finden die Erfassung und Verarbeitung allerdings nicht immer im selben rechtlichen Kontext statt: So führt bspw. die Erhebung im Ausland zu Unklarheiten. In solchen Fällen kann es schwierig sein, zu bestimmen, welche Rechtsgrundlage gilt, da in der DSGVO die Kontexte der Erhebung und Verarbeitung nicht differenziert werden. Zwar gilt für die Verarbeitung das Gesetz am Standort der Universität. Dennoch können Gesetze in anderen Ländern möglicherweise unterschiedlich ausgelegt werden. [9]

Für die Datenverarbeitung benötigen Forschende eine Rechtsgrundlage: Diese besteht nach Art.6 der DSGVO entweder dann, wenn Teilnehmende einwilligen, dass ihre Daten verarbeitet werden dürfen, wenn ein öffentliches Interesse an den Forschungsdaten gegeben ist, oder wenn damit die Wahrnehmung einer Aufgabe im öffentlichen Interesse verbunden ist. Im BDSG (Art.27.1) wird konkretisiert, dass eine Verarbeitung

"auch ohne Einwilligung für wissenschaftliche oder historische Forschungszwecke oder für statistische Zwecke zulässig [ist], wenn die Verarbeitung zu diesen Zwecken erforderlich ist und die Interessen des Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen". [10]

Eine informierte Einwilligung ist damit von Rechts wegen nicht pauschal notwendig, aber forschungsethisch wünschenswert. An gleicher Stelle wird zudem auf die Pflicht verwiesen, dass personenbezogene Daten "zu anonymisieren [sind], sobald dies nach dem Forschungs- oder Statistikzweck möglich ist, es sei denn, berechnete Interessen der betroffenen Person stehen dem entgegen" (BDSG, Art.27.3). Wenn diese Daten nicht anonymisiert wurden, müssen die betreffenden Personen der Verwendung von Klarnamen spätestens zum Zeitpunkt einer Veröffentlichung zustimmen. [11]

Aufgrund der abstrakten gesetzlichen Formulierungen plädieren wir für eine dokumentierte Einwilligungserklärung und ermuntern dazu, die in der DSGVO vorgeschlagenen Maßnahmen umzusetzen. Dazu zählen u.a., den Zugang zu Daten zu limitieren, Gebote der Datenminimierung zu beachten sowie Verantwortlichkeiten der Verarbeitung zu regeln und zu dokumentieren. Im Folgenden werden wir hierauf näher eingehen und die damit verbundenen Herausforderungen für qualitative Forschungsdesigns diskutieren. [12]

Eine datenschutzrechtliche Maßnahme für Forschende betrifft die Einwilligung seitens Beforschter. Im Zuge der informierten Einwilligung werden Teilnehmende darüber aufgeklärt, wer die datenverarbeitenden Personen sind und welche Daten zu welchem Zweck verarbeitet werden. Außerdem werden sie über Möglichkeiten des Widerrufs informiert (EUROPEAN DATA PROTECTION BOARD 2020, S.15). Sollten die entsprechenden Daten später für andere Zwecke verwendet werden, erfordert dies eine neue Zustimmung der Teilnehmenden. Der Zweck kann dabei nicht durch eine beliebig weit gefasste Zweckbestimmung geregelt werden, sondern muss eindeutig identifizierbar und abgrenzbar sein (RAT FÜR SOZIAL- UND WIRTSCHAFTSDATEN [RATSWD] 2020, S.11). Damit die Einwilligung auch rechtlichen Bestand hat, müssen Teilnehmende nicht nur über diese Aspekte informiert sein, sondern die Studienteilnahme muss freiwillig (BDSG, Art.4a) und die Person muss einwilligungsfähig sein (DSGVO, Art.8). Nach DSGVO Art. 8 ist diese Einwilligungsfähigkeit ab 16 Jahren gegeben. Jüngere Personen benötigen das Einverständnis der Erziehungsberechtigten. Jedoch wies die ZENTRALE DATENSCHUTZSTELLE DER BADEN-WÜRTTEMBERGISCHEN UNIVERSITÄTEN (ZENDAS) darauf hin, dass dieser Passus nicht für Forschung gelten müsse (ZENDAS 2018a). Diesbezüglich wird im Landesdatenschutzgesetz Bayerns erläutert, dass eine Einwilligungsfähigkeit individuell beurteilt werden muss, aber diese schon bei 14-Jährigen durchaus gegeben sein kann (BAYERISCHER LANDESBEAUFTRAGTER FÜR DEN DATENSCHUTZ 2021, S.20). [13]

In der Forschungspraxis zeigen sich große Herausforderungen sowohl bei der Darstellung des Zwecks als auch bei dem Zeitpunkt, wann man die Einwilligung schriftlich einholt. Beispielsweise muss der Zweck der Datenerhebung so formuliert sein, dass er für Teilnehmende unmissverständlich ist, aber zugleich ihr Antwortverhalten nicht beeinflusst wird (RATSWD 2014, S.12). Manche Forschende empfinden ggf. eine schriftliche informierte Einwilligung als hinderlich für die Entwicklung von Vertrauen. BENNER und LÖHE (2019) plädierten deswegen für eine mündliche Einwilligung, die mithilfe eines Diktiergerätes aufgenommen wird. Damit diese den Datenschutzbestimmungen gerecht wird, muss die Aufzeichnung mit der mündlichen Einwilligung von der Tonaufnahme des Interviews getrennt und für Dokumentationszwecke sicher aufbewahrt werden. Generell hat die Einwilligung vor der eigentlichen Datenerhebung zu erfolgen und kann nicht nachträglich eingeholt werden (BAYERISCHER LANDESBEAUFTRAGTER FÜR DEN DATENSCHUTZ 2021, S.28). [14]

Bei der Erhebung personenbezogener Daten stellt sich die Frage, ob dies notwendig ist und dem Forschungszweck dient. Um das Risiko des Missbrauchs zu verringern, gilt grundsätzlich das Gebot der Datensparsamkeit und Datenminimierung (DSGVO, Art.5, Abs.3). Um dieses umzusetzen, sollten nur jene Daten gesammelt werden, die dem festgelegten Zweck dienlich sind. Folglich sollten Forschende im Vorfeld überlegen, welche Informationen erhoben werden sollen. Daten, die dem Forschungszweck nicht dienlich sind, müssen aus dem erhobenen Material entfernt und dürfen nicht verarbeitet werden (ZENDAS 2018b). Das Gebot der Datenminimierung gilt auch bei Aussagen über Dritte. Eine Datenminimierung wird hier durch das Anonymisieren und Pseudonymisieren personenbezogener Daten erreicht. [15]

Anonymisierung und Pseudonymisierung stellen also zentrale Maßnahmen für den Schutz personenbezogener Daten dar: Im Falle von *pseudonymisierten* Daten werden Klarnamen durch alternative Namen oder Kennkürzel ersetzt, sodass Dritte nicht auf die betroffene Person schließen können. Eine Kennkürzeltabelle ermöglicht in diesem Zusammenhang Forschenden die Zuordnung; ist diese jedoch nicht dokumentiert, gelten pseudonymisierte Daten als anonymisiert (RATSWD 2017a, S.15). Die Beschränkungen zur Identifizierung der Teilnehmenden sind durch die Verantwortlichen so zu gestalten, dass weder Hilfskräfte noch nicht-gelistete Dritte Personenbezüge nachvollziehen können (SACHS 2010, S.547-548). All Jene, die im Rahmen der Forschung Zugang zu nicht unkenntlich gemachten Personendaten haben (z.B. Verbundpartner:innen, Hilfskräfte, Transkriptionsfirmen), müssen in einer Liste der Verantwortlichkeiten geführt werden. Das heißt nicht nur, dass innerhalb eines Forschungsprojektes der Zugang zu personenbezogenen Daten zu limitieren ist, sondern auch, dass Universitäten eine dafür geeignete Infrastruktur bereitzustellen haben. [16]

Bei anonymisierten Daten existiert hingegen kein Bezug zu einer bestimmten Person. Damit dies zutrifft, muss eine sog. objektive Anonymität bestehen, d.h., dass weder die datenverarbeitende noch eine andere Instanz über Zuordnungsmöglichkeiten zwischen Personen und deren Aussagen verfügt

(RATSWD 2017a, S.14). Sollte eine Anonymisierung vollständig gelingen, hat dies datenschutzrechtliche Konsequenzen, da Maßnahmen des Datenschutzes in diesem Fall nicht mehr angewendet werden müssen (a.a.O.). So wird in statistischen und klinischen Studien häufig auf Anonymisierung zurückgegriffen, um mit einem veränderten Datensatz weiterarbeiten zu können und um den technischen und organisatorischen Aufwand für die Verhinderung einer Deanonymisierung zu umgehen. [17]

In der DSGVO wird dabei nicht geregelt, *wie* Anonymisierung umzusetzen ist. Stattdessen wird in den LDSG abstrakt ausgeführt, dass Anonymisierung erreicht wurde, wenn Personen "nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft" reidentifiziert werden können (z.B. Brandenburgisches Datenschutzgesetz⁴, Art.3.3.1). Sie bleiben damit jedoch offen für Interpretationen. Gerade durch kontinuierlich wachsende technische Möglichkeiten wie digitale Bildersuchen oder das Entzerren technisch audioverzerrter Interviews und Ähnliches werden die Grenzen der Verhältnismäßigkeit von Aufwand kontinuierlich verschoben. Ein weiterer Aspekt der Anonymisierung betrifft die Frage, für wen Daten anonym sind. Im LDSG Baden-Württembergs wird bspw. gefordert, dass Personen für andere öffentliche Einrichtungen anonym bleiben sollen, sodass Matrikelnummern Studierender von Kindergeldkassen nicht dechiffriert werden können. Die Daten der Person sind entsprechend für andere Institutionen anonymisiert, aber innerhalb eines Systems pseudonymisiert (ZENDAS 2016). [18]

Nach der DSGVO sind zudem auch andere Datenarten schützenswert: So werden bspw. Weltanschauungen, sexuelle Orientierung oder politische Haltungen als personenbezogene Daten bezeichnet (Art.9.1); diese sind in einem Maße zu verändern, dass kein Personenbezug hergestellt werden kann. Persönliche Einstellungen, biografische Schlüsselmomente oder die Perspektive von Expert:innen auf soziale Phänomene sind jedoch häufig Gegenstand sozialwissenschaftlicher Forschung und müssen für den Zweck der Verarbeitung in unveränderter Form vorliegen (LOCHNER 2017). Es stellt sich folglich die Frage, ob qualitativ erhobene Daten überhaupt im rechtlichen Sinne *anonymisiert* werden können oder es sich dabei nicht um Pseudonymisierung handelt, womit die Datenschutzverordnungen grundsätzlich weiterhin gelten. Daher argumentierten Datenschutzbeauftragte nicht selten, dass in qualitativen Forschungsdesigns eine informierte Einwilligung vorzuliegen habe (RATSWD 2017a, S.15, 2020, S.19). Folglich müssen Forschende Maßnahmen umsetzen, um den Personenschutz zu gewähren, und zudem, um sich rechtlich abzusichern. [19]

Im Rahmen der Auseinandersetzung mit aktuellen rechtlichen Rahmenbedingungen kann der Eindruck entstehen, dass vorwiegend distinkte Daten wie Matrikelnummern oder Namen im Fokus des Datenschutzes stünden. In qualitativ angelegten Studien sind die erhobenen Materialien aber ebenso durchdrungen von Informationen mit Personenbezug, in Transkripten bspw. im

4 Gesetz zum Schutz personenbezogener Daten im Land Brandenburg, 15. Mai 2018, <https://bravors.brandenburg.de/gesetze/bbgdsg> [Datum des Zugriffs: 31. März 2023].

Fall von Namensnennungen, Aussagen über Dritte oder persönlichen Einstellungen. Dementsprechend ist es gängige Praxis, Transkripte zu pseudonymisieren. Ob nach diesem Schritt ein "erheblicher Aufwand" zur Deanonymisierung von Befragten seitens Dritter notwendig ist, hängt bei qualitativen Forschungsmethoden jedoch insbesondere von der Präsentation der Daten und ihrer Erhebungskontexte ab. Da es in qualitativer Sozialforschung oft darum geht, Zusammenhänge und Handlungsmöglichkeiten in komplexen Situationen zu verstehen, sind diese Kontexte besonders relevant, können damit aber gleichzeitig Datenschutzbestrebungen zuwiderlaufen. [20]

3. Datenarten

In gängigen Datenschutzrichtlinien liegt der Hauptfokus auf "personenbezogenen Daten". Als personenbezogen werden dabei nicht nur Namen, Orte, biografische und Gesundheitsdaten, sondern ebenso politische Meinungen oder weltanschauliche Überzeugungen verstanden (s. Abschnitt 2). In der Praxis ist eine Anonymisierung erst dann erreicht, wenn die verarbeiteten Daten keiner Person mehr zugeordnet werden können. Entscheidend ist, inwieweit diese Daten auf sich im empirischen Material äußernden Personen beziehen oder auf Personen, über die sich Teilnehmende im Material äußern. [21]

Die Gewinnung von Daten erfolgt häufig mithilfe von Aufnahmegeräten, Videokameras oder Fotoapparaten (RATSWD 2017b, S.10). Durch diese Instrumente können verschiedene Datenformate miteinander kombiniert werden. Zum Beispiel enthalten Audioaufnahmen Informationen zu Stimmen, während in Videodateien oder Fotos Gesichter mit räumlichen Kontexten verbunden sind, was eine Anonymisierung erschwert (RATSWD 2017a, S.23). Bei der Erhebung von Forschungsdaten ist es daher wichtig, das Erhebungsinstrument so zu wählen, dass das Prinzip der Datenminimierung beachtet wird. Weiterhin erfordern unterschiedliche Datenformate ggf. spezifische Strategien des Datenschutzes. Der Fokus liegt im Folgenden auf Transkripten – einem Datenformat, bei dem einige Informationen (bspw. Stimmen) bereits aus dem Material entfernt wurden – und die für qualitativ Forschende ein weit verbreitetes Werkzeug der Dokumentation sind. [22]

MEYERMANN und PORZELT (2019) unterschieden für die empirische Sozialforschung vor allem zwischen direkten und indirekten Identifikatoren sowie sensiblen Merkmalen: "Direkte Identifikatoren" stellen personenbezogenen Daten dar und umfassen bspw. Namen, Adressen und Kontaktdaten, aber auch Geburtstage, Stimmen, Bildnisse, Beschreibungen physischer Merkmale oder andere Aspekte wie spezifische Orte (S.6-7). Wenngleich scheinbar eindeutig, so ist innerhalb dieser Kategorie prinzipiell ein Kontinuum erkennbar: So ist es mittels Namen, Bildnissen oder der Stimme unmittelbar möglich, Personen zu identifizieren. Adressen und Kontaktdaten können eine Kontaktherstellung und damit direkte Folgen für die Genannten bedeuten. Beschreibungen physischer Merkmale hingegen treffen auf viele Menschen zu, sodass sie allein nicht zwangsläufig zur Identifikation einer Person führen müssen. Werden jedoch einzigartige physische Merkmale genannt (spezifische Tattoos, Muttermale etc.),

oder mehrere Merkmale mit einer Person verbunden, können diese u.U. eindeutig zugeordnet werden. [23]

"Indirekte Identifikatoren" sind nach MEYERMANN und PORZELT "personenbeziehbare" Daten, d.h., dass keine unmittelbare und spontane Deanonymisierung möglich ist, weil es sich nicht um Daten handelt, die einen direkten Bezug auf die sich äussernden Personen haben (S.7). Vielmehr entsteht ein Datennetz, in dessen Rahmen Beziehungen zwischen den Teilnehmenden, dritten Personen und Institutionen expliziert werden, die dann durch Rückschlüsse und Kombinationen potenziell eine Deanonymisierung erlauben. Diese Daten sind also vor allem als Kontextinformationen zu verstehen. Zu ihnen gehören die Namen Dritter, Verwandtschaftsbeziehungen, Zugehörigkeiten zu Gruppen, Orte (bspw. Städte), Institutionen oder Rollen, Kompetenzen sowie geografische Bezüge. Inwieweit bestimmte Informationen als personenbeziehbar gelten, ist, so MEYERMANN und PORZELT (a.a.O.), abhängig "von den Angaben [...], die vorliegen, und den sonstigen Informationen, die zugänglich sind". Diese Kategorie von Daten ist also weniger eindeutig bestimmbar. Auch kann das Wissen darüber, dass bestimmte Personen mit Forschenden gesprochen haben (ohne zu wissen, was besprochen wurde) als Kontextdatum gewertet werden, denn "the presence of the researcher in the field for an extended time can make it easy for others, especially in a small community, to locate individuals and places" (MOOSA 2013, S.484). [24]

"Sensible Daten" sind nach MEYERMANN und PORZELT (2019, S.24) bezüglich der Frage wichtig, "ob den Betroffenen durch Reidentifizierung ein Schaden entstehen könnte", wobei dieser Begriff eher diffus bleibt. Jedoch bemerkten MEYERMANN und PORZELT, dass es in der DSGVO das Konzept der "besonderen Kategorien von Daten" (S.7) gebe, worunter neben der Herkunft, politischen Überzeugungen und vielem mehr auch Gesundheitsdaten und Daten zur sexuellen Orientierung gehörten. Wir schließen daraus, dass diesen Aspekten in der Praxis der Anonymisierung aus ethischer und rechtlicher Sicht besondere Aufmerksamkeit zukommen muss. [25]

Die Unterscheidung zwischen direkten und indirekten Identifikatoren trifft jedoch auf praktische Hürden:

- Anonymisierung ist nicht nur ein Aspekt in Daten, sondern kann im Zuge von Forschung auch bezogen auf Metadaten Bedeutung erlangen (BISCHOF, LENGERER & MEYER 2022).
- Weiterhin können Teilnehmende selbst Quelle personenbezogener Daten sein, bspw. wenn sie über Dritte reden. Die Begriffe "direkt" und "indirekt" könnten dann verschwimmen und verwirren, wenn der Bezugsrahmen nicht expliziert wird.
- Dies ist besonders relevant, da Forschende nicht mit letzter Sicherheit abschätzen können, mittels welchem Aufwandes eine Deanonymisierung für welche Personengruppen möglich ist. Personenbezogenheit, d.h. die

vermeintliche Direktheit, kann jenseits rechtlicher Definitionen ein graduelles Phänomen sein.

- Weiterhin können Diskriminierungen nicht nur rechtlich relevant sein, sondern auch eine forschungsethische Dimension haben. Gleichsam hat jedoch nicht jeder potenziell forschungsethisch relevante Sachverhalt ein diskriminierungsbezogenes rechtliches Pendant. Trotzdem müssen auch dann Maßnahmen der Anonymisierung abgewogen werden. [26]

Hier wird das praktisch fragile Verhältnis zwischen den Konzepten "personenbezogen" und "personenbeziehbar" deutlich, welches auch in der Begrifflichkeit von "direkt" und "indirekt" nicht transparent problematisiert, sondern eher metaphorisch verschleiert wird: Anonymisierung beinhaltet das reflexive Ringen mit der Gefahr einer Deanonymisierung von Teilnehmenden auf Basis von bestehendem oder antizipiertem (und damit aus Forschendensicht nicht vollständig erfassbarem) Hintergrundwissen. Dieses Ringen ist nicht allein in einen rechtlichen Kontext (als Ergebnis institutionalisierter gesellschaftspolitischer Aushandlungen) eingebunden, sondern auch in einen forschungsparadigmatischen Kontext im Sinne guter wissenschaftlicher Praxis, und in einen ethischen interpersonellen Kontext zwischen Forschenden und Beforschten. Während MEYERMANNs und PORZELT's (2014, 2019) Unterscheidung von direkten und indirekten Identifikatoren bzw. personenbezogenen und personenbeziehbaren Daten vor allem rechtlichen Ursprungs ist, erachten wir es für wichtig, Datenarten zu unterscheiden, bei denen die rechtlichen Bedingungen zwar anerkannt werden, jedoch eine stärker forschungspraktische Orientierung wesentlich ist. So kann die Komplexität einer anonymisierenden qualitativen Forschungspraxis berücksichtigt werden. [27]

Wir schlagen eine Unterscheidung von vier für die Praxis der Anonymisierung bedeutsamen Datenarten vor, die sich sowohl auf Teilnehmende selbst als auch auf durch Teilnehmende genannte Dritte beziehen können:

- *Durchführungsdaten* sind alle Daten, die bei der Vorbereitung und praktischen Umsetzung von Erhebungen als Metadaten anfallen. Darunter zählen bspw. Namen von Teilnehmenden, Telefonnummern, Adressen oder auch die Orte, an denen die Erhebung stattfindet. Ebenso zählen dazu vergebene Kennkürzel und Zuordnungstabellen.
- *Identifikationsdaten* sind alle Daten, die eindeutig auf konkrete Personen bezogen oder beziehbar sind (MEYERMANN & PORZELT 2019) – sowohl von Teilnehmenden als auch von Dritten, über die Teilnehmende sprechen. Das können z.B. Namen von Personen, Orte, Organisationen, Vereine, Geburtstage, Arbeitgeber:innen u.v.m. sein. Insbesondere können hier die von MEYERMANN und PORZELT benannten "seltene[n] Merkmalskombinationen" (S.25) relevant sein, bei denen durch die Kombination von zwei nicht direkt auf Personen beziehbaren Merkmalen nahezu eindeutig zuordenbare Kombinationen entstehen, die de facto Identifikationsdaten darstellen, z.B. "Rektor der örtlichen Grundschule,

Autobauer in Niedersachsen, Familie mit fünf Kindern in Ort XY, weiblicher Feuerwehr'mann'" (S.25).

- *Kontextdaten* sind jene Daten, die bspw. Auskunft über den Kontext Sprechender bzw. Dritter geben. Zum Beispiel können das Informationen über Verwandtschaftsbeziehungen zu nicht zwingend namentlich genannten Personen, unspezifische Ortsangaben, biografische Begebenheiten oder Orte sein, die dann relevant sind, wenn dabei Verbindungen zu anderen Informationen hergestellt werden können (z.B. 27 km südlich von Stadt X; 12,3% als Zweitwahlergebnis der Partei Y bei der letzten Landtagswahl). Kennzeichnend ist hier die mangelnde Eindeutigkeit der Beziehbarkeit auf spezifische Personen.
- *Sensible Daten* und Bewertungen bedürfen besonderer Feinfühligkeit bei der Anonymisierung. Entsprechend Art.9 der DSGVO sind bspw. Konfession, Sexualität, politische Meinung etc. geschützt, weil Menschen auf ihrer Grundlage keine Diskriminierung erfahren sollen. Ebenso können öffentlich abgegebene Bewertungen Teilnehmende schädigen. Wenngleich sich die Erweiterung auf Bewertungen damit von der rein rechtlichen Begründung entfernt, umfasst sie ebenso mögliche negative Konsequenzen für Teilnehmende und Dritte, die fallspezifisch abzuwägen sind. [28]

4. Die Praxis der Anonymisierung

4.1 Anonymisierung im Forschungsprozess

Dazu, wann Anonymisierung im Forschungsprozess umgesetzt werden soll, herrschen unterschiedliche Auffassungen: BACHER und HORWARTH (2011, S.9) begriffen bspw. die Anonymisierung von Namen, Orts- und eventuell Zeitangaben als einen der "zentrale[n] Bestandteile des Transkriptionsvorganges". ROSENTHAL (2015, S.104-105) verwies hingegen darauf, dass biografische Daten erst dann anonymisiert werden können, wenn durch die Analyse herausgefunden wurde, welche Aspekte für die Interpretation und theoretischen Generalisierungen relevant seien. [29]

Fragen nach Anonymisierungspraktiken kommen oft bereits während des Rekrutierungsprozesses auf. TAYLOR (2015, S.286) berichtete davon, dass es für einige an einer Studie teilnehmende Organisationen zentral war, ob sie namentlich genannt würden, weil sie sich von der Teilnahme eine positive Außenwahrnehmung erhofften. Im Kontext größerer Projekte sind dementsprechend bereits früh der Internetauftritt, die Nutzung sozialer Medien oder erste Veröffentlichungen von besonderer Bedeutung (S.287), auch wenn es zunächst so scheinen mag, als wäre Anonymisierung ein Thema, das erst bei Veröffentlichung besondere Relevanz erlangt. [30]

Mit Blick auf die Frage, wann zu anonymisieren ist, plädieren wir dafür, sich vor der Datenerhebung für eine Anonymisierungsstrategie zu entscheiden. Wie VAINIO (2013, S.695) argumentierte, ist dies ein zentraler Bestandteil des Forschungsdesigns, weil es ggf. mit der Modifikation von Daten einhergeht und

Auswirkungen auf jeden Schritt des Forschungsprozesses hat. Dabei ist zugleich zu beachten, dass eine frühzeitige Entscheidung für eine Anonymisierungsstrategie nicht bedeutet, dass damit alle diesbezüglichen Fragen endgültig geklärt wären (TAYLOR 2015, S.288-289). Anonymität als Prozess und Verhältnis zu verstehen bedeutet, die verschiedenen Akteur:innen zu berücksichtigen (Teilnehmende, Wissenschaftler:innen, Beobachter:innen) und im Zeitverlauf flexibel auf entstehende Veränderungen in ihren Ansprüchen und praktischen Auslegungen von Anonymität zu reagieren. [31]

4.2 Anonymisierungsstrategien

In der Forschungsliteratur sind zumindest drei Anonymisierungsstrategien zu unterscheiden (s. Tabelle 1). Von jeder Strategie können unterschiedliche methodische Konsequenzen abgeleitet werden. Die Wahl der Strategie orientiert sich dabei an der *Analyse*, der *Publikation* oder den *Teilnehmenden*. Durch die folgende Systematisierung wollen wir es Forschenden ermöglichen, die eigene Vorgehensweise zu dokumentieren und in Teams zu diskutieren – letzteres ist aufgrund der Heterogenität von Zugängen zu Anonymität unter Forschenden (SURMIAK 2018) eine zentrale Herausforderung.

	Analyseorientiert	Publikationsorientiert	Teilnehmendenorientiert
Umfang	Umfassende Anonymisierung des gesamten Transkripts	Selektive Anonymisierung ausgewählter Textstellen	Variabel
Zeitpunkt	Früh (Aufarbeitungsphase)	Spät (Veröffentlichungsphase)	Kontinuierlicher Austauschprozess
Durchführende Person	Forschende, Erstanonymisierung durch Transkribierende	Forschende	Forschende im Austausch mit Teilnehmenden
Anlass	Primär- und Sekundärnutzung	Wissenschaftliche und andere Publikationen	Partizipative Forschung; anwendungsorientierte Forschung

	Analyseorientiert	Publikationsorientiert	Teilnehmendenorientiert
Vorteile	Teil der Analyse Bei der Arbeit mit unerfahrenen Teammitgliedern, großen Gruppen etc. kann Datenschutz nachhaltig sichergestellt werden Bereitstellung für Sekundärnutzung einfacher möglich Einheitlichkeit bei verschiedenen Veröffentlichungen	Ökonomisch günstiger Keine Verluste in der analytischen Interpretationstiefe Änderungen bei Gesetzeslage können zeitaktuell berücksichtigt werden	Kommunikative Validierung kann Qualität der Anonymisierung und der Interpretation steigern Vertrauen Teilnehmender in Forschungsprozess wird gestärkt
Nachteile	Sehr zeitaufwendig und teuer Gefahr der Verfälschung der Daten Gefahr von Fehlinterpretationen in der Analyse	Vorsicht während der Verarbeitung und Auswertung der Daten notwendig Bei Datenleck Verletzung der Datenschutzversprechen Inkonsistente Anonymisierungsweisen bei unterschiedlichen Veröffentlichungen von Teammitgliedern	Hoher Kommunikations- und Organisationsaufwand Widersprechende Meinungen unter einzelnen Teilnehmenden

Tabelle 1: Anonymisierungsstrategien [32]

Eine *analyseorientierte Anonymisierungsstrategie* umfasst die vollständige und durchaus aufwendige Anonymisierung möglichst früh im Prozess, sodass in der Analysephase mit Daten gearbeitet wird, die möglichst wenig Rückschlüsse auf die Identität Teilnehmender zulassen (CORTI, DAY & BACKHOUSE 2000). Umgesetzt wird Anonymisierung dabei durch Forschende und ggf. Transkribierende. Für jede Anonymisierungstechnik müssen sowohl die spezifischen Erfordernisse der jeweiligen Auswertungsmethode als auch mögliche Verwertungsweisen antizipiert werden. CORTI et al. (§24) empfehlen, die Anonymisierung bereits bei der Transkription durchzuführen und verwiesen auf die Notwendigkeit eines Systems der Ersetzung von Namen und Orten, welches im Team geteilt und über Publikationen hinweg konsistent sein sollte. Als den ressourcenintensivsten Schritt bezeichneten CORTI et al. das aufmerksame Lesen jedes Transkripts, um subtilere Hinweise auf Personen, Orte und Institutionen zu identifizieren. In der Literatur wird ein solcher Ansatz

insbesondere im Zusammenhang mit Sekundärnutzungen diskutiert (s. auch MEDJEDOVIĆ & WITZEL 2010; THOMSON, BZDEL, GOLDEN-BIDDLE, REAY & ESTABROOKS 2005). [33]

Die *publikationsorientierte Anonymisierungsstrategie* zeichnet sich dadurch aus, dass nicht ganze Transkripte vor der Analyse anonymisiert werden, sondern lediglich einzelne Passagen für Präsentationen oder schriftliche Veröffentlichungen (REICHERTZ 2015; TAYLOR 2015). Die Veränderung von Daten geschieht also spät im Forschungsprozess und wird vor allem durch die publizierenden Forschenden umgesetzt. REICHERTZ (2015, S.195-196) schlug diesbezüglich ein zweistufiges Vorgehen vor, bei der eine grobe Anonymisierung während der Datenaufarbeitung erfolgt und Klarnamen Teilnehmender ersetzt werden, um eine sofortige Identifikation zu vermeiden. Er betonte jedoch, dass dieser erste Schritt nur strukturell unbedeutende Merkmale erfassen dürfe, und möglichst alle Merkmale des untersuchten Falles mit dem Ziel der Interpretierbarkeit zu erhalten seien. Schwierig ist es gleichwohl, zu diesem Zeitpunkt bereits abzuschätzen, welche Merkmale für die Analyse relevant sind und erhalten werden müssen. Selbst eine grobe erste Anonymisierung kann zudem äußerst zeitaufwendig sein (SAUNDERS et al. 2015, S.628). Erst in der Phase vor der Publikation schlug REICHERTZ (2015) eine tiefergreifende Anonymisierung mit dem Ziel vor, "den Fall in seiner *Einzigartigkeit* zu verschleiern und dennoch seine Typik beizubehalten" (S.196). [34]

Mit einer publikationsorientierten Anonymisierungsstrategie geht eine gewisse Flexibilisierung einher, weil für jede Publikation neu entschieden wird, wie vorzugehen ist. Da sich bspw. durch Pseudonymisierung die Logik des veröffentlichten Textes insgesamt ändern kann (s. VAINIO 2013, S.689 für ein anschauliches Beispiel), ist es für jede Publikation sinnvoll, möglichst früh eine Entscheidung darüber zu treffen. TAYLOR (2015, S.287) wies darauf hin, dass sich Anonymisierungspraktiken einzelner Teammitglieder im Zeitverlauf individualisieren können, nicht zuletzt, weil unterschiedliche Journale und Zielgruppen dies notwendig machten. Auch kann es gerade bei großen Projektteams schwierig sein, die Publikationen von Kolleg:innen mitzudenken und deren Anonymisierungsentscheidungen praktisch mittragen zu müssen. Eine besondere Bedeutung kommt dabei dem Projektende zu, wenn Teammitglieder zwar weiter mit den Daten arbeiten, sich das Team aber zerstreut und weniger einfach für Absprachen zusammengebracht werden kann, oder Anonymisierungen zu Archivierungszwecken eventuell nicht von den an der Erhebung beteiligten Wissenschaftler:innen, sondern von technischem Personal durchgeführt werden (S.288). [35]

Quer zu den beiden oben genannten Strategien liegt die *teilnehmendenorientierte Anonymisierungsstrategie*. Sie kann sowohl früh, umfassend und analyseorientiert oder spät, selektiv und publikationsbezogen angewendet werden und zeichnet sich dadurch aus, dass die konkreten Schritte der Anonymisierung im Austausch mit Teilnehmenden festgelegt werden, also nicht allein durch die Wissenschaftler:innen (und Transkribierenden). Entsprechend wird dieses Vorgehen in der Literatur vor allem im Zusammenhang mit

partizipativen Methoden genannt (WILES, CROW, HEATH & CHARLES 2008): Forschende sprechen die Nutzung der Daten mit Teilnehmenden ab, löschen ggf. Daten oder Interpretationen, wenn Teilnehmende dies wünschen und ermöglichen zum Teil die Auswahl von Pseudonymen (S.425). WILES et al. betrachteten diese Strategie als Konsequenz gesellschaftlicher Individualisierung und Digitalisierung und gaben zu bedenken, dass diese Herangehensweise auch im Wunsch Teilnehmender resultieren könnte, Anonymität abzulehnen, falls sie durch die Chance, ihre Geschichte zu erzählen, motiviert seien. Ein Verzicht auf Anonymisierung bedeute jedoch nicht automatisch weniger Arbeit (REICHERTZ 2015, S.192), sondern eine Erhöhung des kommunikativen Aufwands. Zudem zeigte SURMIAK (2018), dass mit der Wahl einer teilnehmendenorientierten Strategie keineswegs die Verantwortung von den Forschenden auf die Teilnehmenden verlagert wird. In der Forschungspraxis entschieden sich auch partizipativ orientierte Wissenschaftler:innen zum Teil, gegen den Willen Beforschter zu anonymisieren und besprachen die Transkripte teils nicht mit den Beforschten (§18). Grundsätzlich gilt, dass auch bei der teilnehmendenorientierten Anonymisierungsstrategie bspw. Aussagen über Dritte vor Weitergabe oder Veröffentlichung genau geprüft werden müssen (s. für eine kritische Diskussion VAINIO 2013).

Forschungsphase. Anonymisierungsstrategie	Erhebung	Aufbereitung	Auswertung	Veröffentlichung
Analyseorientiert	Datenschutz- erklärung Vergabe von Kennkürzeln zur formalen Anonymisierung bei Rekrutierung und Erhebung	Faktische Anonymisierung des gesamten Transkripts während oder nach der Transkription und vor Weiter- verarbeitung Dokumentation der faktischen Anonymisierung in Protokoll	Auswertung anhand faktisch anonymisierter Transkripte	Vor Veröffentlichung lediglich Prüfung der Wirksamkeit vorher vorgenommener Anonymisie- rungsschritte

Forschungsphase. Anonymisierungsstrategie	Erhebung	Aufbereitung	Auswertung	Veröffentlichung
Publikationsorientiert	Datenschutz- erklärung	Vergabe von Kennkürzeln zur formalen Anonymisierung bei Rekrutierung und Erhebung	Auswertung anhand nicht anonymisierter Transkripte Besondere Zugangsbe- schränkungen zu den Daten notwendig	Faktische Anonymisierung selektiv nur für zitierte Interviewstellen und tatsächlich verwendete Fälle vor Veröffentlichung Dokumentation in Protokoll
Teilnehmerorientiert	Datenschutz- erklärung Variable Absprache mit Teilnehmenden Dokumentation der Absprachen	Variabel anhand der Absprachen mit Teilnehmenden Separate Abwägung der Bewertungen und Aussagen Dritter durch Teilnehmende Protokollierung der gewählten Technik	Separate Abwägung der Bewertungen und Aussagen Dritter durch Teilnehmende Protokollierung der gewählten Technik	Separate Abwägung der Bewertungen und Aussagen Dritter durch Teilnehmende Darstellung der gewählten Technik

Tabelle 2: Anonymisierungsstrategien im Forschungsprozess [36]

In Tabelle 2 fassen wir zusammen, welche Maßnahmen in welchen Phasen des Forschungsprozesses jeweils Anwendung finden können. Was hier idealtypisch getrennt dargestellt wird, ist in der Forschungspraxis teils wesentlich durchmischter. Anonymität kann nicht nur theoretisch als fragiles Verhältnis verstanden werden, das permanent neu hergestellt werden muss (BACHMANN, McHARDY, KNECHT & ZURAWSKI 2021), sondern ist es auch praktisch. Dies gilt sowohl aufseiten Forschender, die ihre Ergebnisse je nach Journal oder Publikum anders aufbereiten, als auch für Teilnehmende: TAYLOR (2015, S.286) schilderte bspw. Fälle, in denen Teilnehmende auf ihrer persönlichen Webseite das Forschungsprojekt verlinkten (ähnliche Problematiken diskutieren BISCHOF et al. 2022). Eine flexible Nutzung von Anonymisierungstechniken wirft zudem Fragen nach der Vergleichbarkeit zwischen anonymisierten und nicht anonymisierten Fällen sowie nach der Anonymität einzelner innerhalb benannter Fälle auf (TAYLOR 2015, S.289). [37]

4.3 Anonymisierungstechniken

Mit den bisherigen Ausführungen zeigen wir, dass es nicht die eine richtige Form der Anonymisierung gibt, sondern dass darüber kontextspezifisch zu entscheiden ist, wobei der gesetzliche Rahmen und die Fachstandards berücksichtigt werden müssen. Im Folgenden werden wir insbesondere auf die publikationsorientierte Strategie fokussieren und um die konkrete Umsetzung der in der Einwilligungserklärung versprochenen Anonymität:

- *Kennkürzel verwenden*: Diese Technik findet früh im Forschungsprozess Anwendung: Mittels eines Kürzels werden einzelne Datensätze einem Fall zugeordnet oder Relationen zwischen Fällen notiert (anstatt: "Toni ist Nachbarin von Marianne", x7 ist Nachbarin von y13). In Verbindung mit der Datenart können Kürzel dazu dienen, Daten zu managen, z.B. "A43int" für Region A, Fall Nr. 43, Datenart: Interview. Gegebenenfalls können durch Kürzel Informationen jedoch auch unerwünscht offenbart werden, insbesondere wenn sie in Texten genutzt werden, um auf Interviewpassagen zu verweisen. So können bspw. KfZ-Kennzeichen in Kürzeln eine ungewollte Re-Anonymisierung zur Folge haben (RATSWD 2020, S.19).
- *Daten verändern*: Gerade durch die Veränderung von Daten unterscheidet sich die Anonymisierung von anderen forschungsethischen Praktiken wie der Einwilligungserklärung (VAINIO 2013, S.685-686): Während im Forschungsprozess sonst versucht wird, exakte Daten zu erhalten (z.B. durch die Transkription von Pausen), werden beim Anonymisieren teils inhaltliche Änderungen der Daten vorgenommen, indem gestrichen, ersetzt, verallgemeinert, verfälscht oder zusammengeführt wird. Aufgrund der dabei getroffenen Vielzahl praktischer und theoretischer Entscheidungen können verändernde Anonymisierungen als Teil der Datenanalyse begriffen werden (S.690). Einen ersten Überblick über die im Folgenden dargestellten Techniken gibt Abbildung 1:

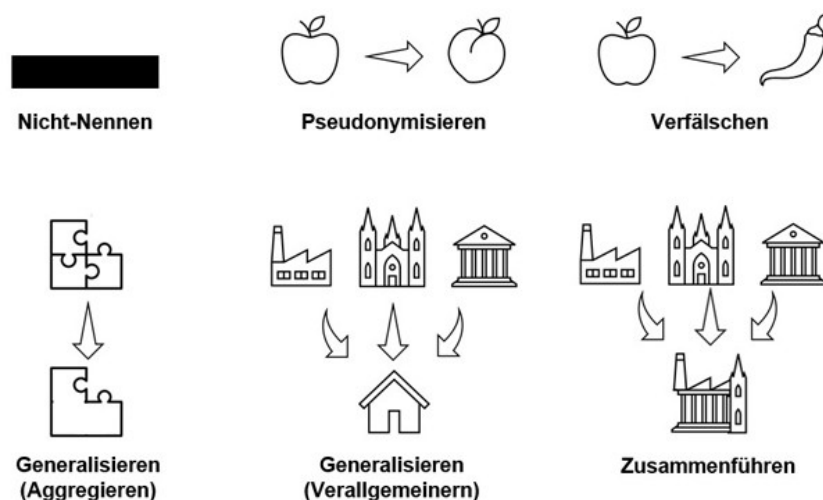


Abbildung 1: Anonymisierungstechniken (s. auch MEYERMANN & PORZELT 2014) [38]

4.3.1 Nicht-Nennen

Daten nicht zu nennen bzw. zu streichen, zu löschen oder zu schwärzen ist eine zentrale Anonymisierungstechnik. Dass Forschende sich häufig mit der Datenschutzvereinbarung verpflichten, auch anonymisierte Interviews nicht vollständig zu veröffentlichen oder an Dritte weiterzugeben, ist bereits eine Form der Nichtnennung. Auch wenn aus Transparenzgründen teilweise explizit die Veröffentlichung von Daten gefordert wird, sollte dies mit Blick auf Datenschutz zurückgewiesen. Denn gerade durch die digitale Datenverknüpfbarkeit kann dies im Kontext qualitativer Forschung ein erhöhtes Risiko der Deanonymisierung nach sich ziehen (RATSWD 2017b, S.38). [39]

Nur kurze Zitate zu nennen kann dazu beitragen, Identitäten zu schützen (SURMIAK 2018, §28). Auch innerhalb von Zitaten ist Streichen eine weit verbreitete Technik. Was konkret nicht genannt wird, hängt von der Fragestellung ab und kann nicht pauschal bestimmt werden. SURMIAK (§29) berichtete bspw. von einer Studie über Veteran:innen, die darum baten, verlorene Körperteile nicht zu benennen. THOMSON et al. (2005, §21) rieten dazu, problematisches Material wie verleumderische oder beleidigende Kommentare über (anonymisierte) Dritte zu löschen (s. Abschnitt 3). Obacht ist zudem bei der technischen Umsetzung insbesondere hinsichtlich der Durchsuchbarkeit von Dokumenten geboten, damit bspw. textverdeckende Balken nicht einfach gelöscht werden können. [40]

4.3.2 Pseudonymisieren bzw. gleichwertiges Ersetzen

Das gleichwertige Ersetzen von Eigennamen durch Pseudonyme ist ebenfalls eine weit verbreitete Technik. Pseudonymisiert werden können nicht nur Personennamen, sondern z.B. auch Orte, Organisationen oder Straßen. Pseudonyme sind jedoch nicht neutral. So kann es bspw. einen Unterschied machen, ob ein Teilnehmer "Kevin oder Paul" heißt (LOCHNER 2017, S.287ff.): Lesende assoziieren mit Pseudonymen bestimmte Zugehörigkeiten zu Altersgruppen, Klassen oder kulturellen Identitäten, was nicht in der Kontrolle der Autor:innen liegt (CORDEN 2007, S.28). Alternativ zu Namen können eindeutige Kürzel oder Nummern (S.22) eingesetzt werden; das verhindert zugleich, versehentlich Namen anderer Teilnehmender als Pseudonyme zu nutzen (THOMSON et al. 2005, §24). THOMSON et al. empfahlen zudem, dieselben Pseudonyme und Ortsbezeichnungen über verschiedene Publikationen hinweg zu verwenden (§21). Dies erhöht jedoch ggf. die Kombinationsmöglichkeiten durch Dritte und könnte eine Deanonymisierung erleichtern. [41]

Pseudonyme sind jenseits des Datenschutzes als eine wesentliche Darstellungsweise wissenschaftlicher Veröffentlichungen relevant, mit der Forschende diese Art der Textproduktion von anderen Genres wie z.B. dem Journalismus abgrenzen (NESPOR 2000, S.546). NESPOR diskutierte kritisch, dass durch die Nutzung von Pseudonymen bestimmte Repräsentationen von Wirklichkeit und Wissenschaftlichkeit produziert würden, die dazu beitragen, die Verbindungen zwischen den Orten, den Autor:innen, den Teilnehmer:innen und den Leser:innen zu verdecken oder auszublenden (S.555). Dagegen erlaube es

erst die konkrete Benennung von Orten, eben solche Verbindungen zwischen hervorzuheben (S.555-556). [42]

4.3.3 Generalisieren

Generalisieren meint das Ersetzen eines Wortes (Name, Ort, Beruf usw.) durch eine verallgemeinernde Bezeichnung. Eine Möglichkeit der Generalisierung ist das Aggregieren, d.h., man wählt höhere Beschreibungsebenen und fasst damit untere Ebenen zu Klassen zusammen, z.B. indem man Menschen Altersgruppen zuordnet, anstatt ihr genaues Alter zu nennen (MEYERMANN & PORZELT 2014, S.8). Eine andere Form der Generalisierung ist das Wählen von Oberbegriffen, z.B. anstelle von Grundschullehrerin die Formulierung "im Bereich Bildung und Erziehung tätig". Generalisierungen werden oft eingesetzt, um eine Dekontextualisierung der Daten im Zuge einer Anonymisierung zu vermeiden, jedoch mit reduzierter Deanonymisierungsgefahr (SURMIAK 2018, §30). [43]

Generalisierungen haben weitreichende erkenntnistheoretische Bedeutungen, denn dadurch wird von der konkreten Untersuchungssituation abstrahiert und die Zitate werden je nach Umsetzung in gewissem Maße entbettet (z.B. wird aus dem Scholl-Gymnasium in Chemnitz ein Gymnasium in (Ost-)Deutschland oder eine weiterführende Schule in Sachsen; NESPOR 2000, S.551). NESPOR argumentierte daher, dass Anonymisierung eine Repräsentationstechnik sei, durch die zusätzlich zum Datenschutz auch der wissenschaftliche Anspruch auf Generalisierbarkeit sprachlich realisiert werde (S.552). Dies habe zudem eine politische Dimension, wenn hierbei gesellschaftliche Verhältnisse verschleiert würden. [44]

VAINIO (2013, S.690) hingegen plädierte dafür, bei der Anonymisierung getroffene Entscheidungen als Teil des Analyseprozesses zu verstehen, bei denen Einzelfälle als beispielhafte Illustrationen generalisierbarer Phänomene dargestellt würden. So sind z.B. Generalisierungen in interdisziplinären Teams voraussetzungsvoll: Bei THOMSON et al. (2005, §19) wurden für Mitglieder die Berufsbezeichnungen "nurse" und "physician" durch "health care professional" ersetzt, was in anderen Kontexten einen entscheidenden Datenverlust bedeuten könnte. [45]

4.3.4 Verfälschen

Durch datenverändernde Eingriffe in den Text können sogar bewusst Angaben verfälscht werden (KÜHL 2020). Im Unterschied zur Generalisierung besteht das Ziel der Verfälschung nicht allein in der Reduktion einer spezifischen Qualität der Daten, sondern in der Konstruktion anderer Qualitäten, die nicht in den ursprünglichen Daten existieren, wobei dies bewusst nicht transparent gemacht wird. Hier ist besondere Vorsicht geboten, damit der Schutz Teilnehmender der Maßstab bleibt und Daten nicht auf eine Art verändert werden, die den eigenen Thesen und Interpretationen entgegenkommt. [46]

Verfälschende Techniken wurden insbesondere im Zusammenhang mit der Präsentation und Publikation von Daten genannt, da hier das Dilemma zwischen präziser Darstellung und dem Schutz Teilnehmender besonders deutlich wird. So berichtete KÜHL (S.62) bspw. davon, mit dem Ziel der Anonymisierung Orts- und Akteur:innenangaben verfälscht zu haben. ROSENTHAL (2015, S.105) wies zudem darauf hin, dass bei Teilnehmenden mit hohem Bekanntheitsgrad auch biografische Angaben verfälscht werden sollten (s. auch WILES et al. 2008, S.423-424). [47]

Ein anderes Beispiel für eine Verfälschung ist das Ersetzen des Namens eines Untersuchungsortes durch den Namen einer Stadt ähnlicher Größe, falls dies nicht die Bedeutung des Inhalts beeinträchtigt (SURMIAK 2018, §31). Im Kontext einer qualitativen Längsschnittstudie über Organisationen änderte bspw. TAYLOR (2015, S.285) unter anderem deren Standort, Schwerpunkt, Finanzierung und Größe. Andere Forschende nutzten zwei verschiedene Pseudonyme für eine Person, um zu vermeiden, dass Lesende verstreute Informationen aus dem Text einer einzelnen Person zuordnen könnten (SAUNDERS et al. 2015, S.621) oder veränderten das Geschlecht (S.628). METSCHKE und WELLBROCK (2002, S.22) nannten auch das "Einstreuen von Zufallsfehlern" als Möglichkeit. [48]

4.3.5 Zusammenführen

Bei der Bildung von Komposita, Idealtypen oder Fiktionalisierungen zeigt sich erneut, wie eng Analysemethoden, Anonymisierungsstrategien und Darstellungstechniken miteinander verwoben sind. Hierbei sind Komposita als zusammengesetzte Personen zu verstehen, wobei eine fiktive Person anhand von Merkmalen verschiedener Personen erfunden und mit einem Pseudonym versehen wird (INTER-AGENCY STANDING COMMITTEE 2014, S.37). Eng verwandt damit sind Idealtypen mit dem Ziel, typische Vertreter:innen für Merkmalskombinationen als Anker für Lesende zu präsentieren. Ergebnisse können so mit Blick auf die Forschungsfrage dargestellt werden, ohne persönliche Details Teilnehmender überhaupt zu nennen (ROSENTHAL 2015, S.105). Wer nicht ganz auf sozialstrukturelle Daten verzichten will, kann im Text für die Gesamtheit der Teilnehmenden deskriptive statistische Merkmale aufführen (SURMIAK 2018, §26). Beide Techniken stellen empirisch fundierte Fiktionalisierungen dar, da fiktive Personen mit dem Ziel der Anonymisierung reeller Personen kreiert werden. Von SURMIAK (§28) interviewte Wissenschaftler:innen setzten bspw. von den Befragten erzählte Geschichten zu einer zusammen und fiktionalisierten sie zum Teil. Der in diesem Zusammenhang genutzte Begriff des Kompilierens bezeichnet hierbei, Zitate und Einsichten basierend auf den Daten verschiedener Interviewpartner:innen zusammenzuführen. [49]

4.4 Darstellung in Veröffentlichungen

Spätestens bei der Publikation stellt sich die Frage nach der Transparenz der Anonymisierung. Zwar riefen einige Wissenschaftler:innen und Einrichtungen dazu auf, Anonymisierungsentscheidungen in Publikationen zu beschreiben (MEYERMANN & PORZELT 2014, S.10; VAINIO 2013, S.694); wie diese Transparenz zu erreichen ist, wurde jedoch selten dezidiert verhandelt (eine Ausnahme ist CORDEN & SAINSBURY 2006a). Häufig wird nur in einem Satz erwähnt, dass Daten anonymisiert wurden und alle Namen im Text Pseudonyme sind. KÜHL (2020, S.68) befürchtete, dass mit einer genauen Beschreibung der Anonymisierungsmaßnahmen auch die Gefahr steige, genau auf das hinzuweisen, was man verbergen wollte. TAYLOR (2015, S.285) hingegen plädierte dafür, ausführlicher über die Gründe, den Abwägungsprozess und die gewählten Techniken der Anonymisierung Auskunft zu geben. Flankierend sei zudem transparent zu machen, in welchem Maße Teilnehmenden spätere Anonymisierungen bewusst waren bzw. diese darauf Einfluss hatten. [50]

Besonders relevant sind textverändernde Maßnahmen in Publikationen im Zusammenhang mit direkten Zitaten: So zeigten bspw. WILES et al. (2008, S.426), dass viele Forschende zwar Änderungen in publizierten wörtlichen Zitaten vornehmen, aber selten für Lesende transparent machen, ob solche Änderungen vorgenommen wurden und was genau diese umfassten. Dabei hatten sowohl die Rezipierenden von qualitativer Forschung als auch die Teilnehmenden hohe Erwartungen an Wissenschaftler:innen, die direkte Zitate in Veröffentlichungen einsetzten (CORDEN & SAINSBURY 2006b, S.28). Damit zusammenhängend stellt sich ebenfalls die Fragen, inwieweit Sprache für Lesende (bspw. durch Streichung von "ähms") mit besserer Lesbarkeit (bspw. durch Weglassen von Nebensätzen und Schlaufen) überarbeitet werden sollte, da besondere Sprachmerkmale ggf. deanonymisierend wirken können. Zudem muss beachtet werden, dass die Markierung von anonymisierten Textstellen die Lesbarkeit unterschiedlich beeinflusst. [51]

Neben der Darstellung direkter Zitate stellt sich die Frage ihrer Einbettung in den Text. In vielen Veröffentlichungen werden Zitate mit einem Hinweis auf die äussernde Person versehen. Dies kann durch Pseudonyme geschehen, durch Kürzel bzw. Zahlen oder die Nutzung allgemeiner Deskriptoren (CORDEN 2007, S.22-23). CORDEN und SAINSBURY (2006b, S.26-27) dokumentierten in diesem Zusammenhang, dass Lesenden teils methodologische Gründe für die Nutzung von Pseudonymen erwarteten und diese in Forschungsreports zu nennen seien. Dass Zitate überhaupt einzelnen Fällen (bspw. Gesprächspartner:innen, Organisationen, Regionen) zugeordnet werden, ist nicht zwingend. Es kann jedoch für Forschende und Rezipierende hilfreich sein, um Äußerungen Teilnehmender kontextuell einzubetten, die Verteilung der verwendeten Zitate im Blick zu behalten und eine eventuell einseitige Berücksichtigung bestimmter Fälle sicht- und kritisierbar zu machen (CORDEN & SAINSBURY 2006a, S.22). CORDEN (2007, S.22-23) berichtete zudem von Autor:innen, die zwar grundsätzlich ein solches Verweissystem nutzten, jedoch bei besonders sensiblen Zitaten überhaupt keine Verweise setzten, um Lesenden

die Möglichkeit zu nehmen, andere Zitate oder Beschreibungen aus dem Text mit diesen sensiblen Stellen zu verknüpfen. [52]

CORDEN und SAINSBURY (2006a, S.21) erläuterten zudem, dass es insbesondere Forschende in politikbezogenen Kontexten gewöhnt seien, Zitate mit beschreibenden Kategorien wie "parent", "carer" usw. zu versehen und dies von Fördermittelgeber:innen zunehmend detaillierter verlangt werde. Dabei stellten sich jedoch nicht nur Fragen nach Lesbarkeit und Vertraulichkeit, sondern auch nach damit eröffneten Interpretationsspielräumen, nach der jeweiligen Relevanz der Informationen und nach den Schlussfolgerungen, die Lesende aus solchen Angaben wirklich ziehen könnten (a.a.O.). Festzuhalten ist, dass Lesende bei der Nutzung allgemeiner Deskriptoren erwarteten, dass die angegebenen Charakteristika relevant für das dargestellte Thema seien (CORDEN & SAINSBURY 2006b). Welche Deskriptoren für die jeweilige Veröffentlichung relevant sind, sollte entsprechend kritisch reflektiert werden. [53]

Direkte Zitate sind jedoch nur ein Beispiel für Daten in Publikationen; vielfach werden bei vorwiegend auf Interviewdaten basierenden Veröffentlichungen ergänzend andere Datenformate genutzt. Insbesondere zur Beschreibung des Kontextes und zur geografischen wie sozialstrukturellen Einbettung werden oft numerische Daten oder Karten zur näheren Charakterisierung herangezogen. Doch durch die Angabe von Zustimmungsraten in Wahlbezirken, von Bevölkerungsgrößen von Landkreisen, der Schuldenquote von Kommunen oder der Sozialhilfequoten in Quartieren können eindeutige Raumbezüge hergestellt werden. Selbst wenn Forschende drei allgemeine Merkmale wie Unternehmensgröße, Firmensitz und Branchenzugehörigkeit angeben, kann dies bestimmten Personengruppen offenbaren, welche Organisation erforscht wurde (KÜHL 2020, S.64). Somit ist es auch für qualitativ Forschende angebracht, sich mit Anonymisierungsinstrumenten quantitativ Forschender vertraut zu machen, z.B. der Vergrößerung ortsbezogener Daten (KINDER-KURLANDA & WATTELER 2015, S.6; WATTELER & KINDER-KURLANDA 2015, S.517). Ein Beispiel wäre hier, nicht die Zensusdaten der Untersuchungsregion zu nutzen, sondern deren Durchschnitt zusammen mit den umliegenden Gebietseinheiten zu bilden (WERNER 2023/im Druck, S.12). [54]

Auch Textdaten kommen in verschiedenen Formen und aus verschiedenen Quellen vor: So nutzt man ggf. neben Interviewäußerungen noch ein besonders treffendes Zitat aus dem Internet, Quellen aus Archiven oder Fotografien, um die Analyse anzureichern oder die Beschreibung zu verdichten: So bettete MEYER (2018a) bspw. zur Kontextualisierung seiner Erhebungen zur Wahrnehmung von Abwanderung durch Jugendliche in seiner Untersuchungsregion ein Zitat der Landrätin von einer öffentlichen Veranstaltung ein, die auf der Website des Landratsamtes dokumentiert worden war. Eine Anonymisierung der Region stellte deswegen keine realistische Option mehr dar. ROSENTHAL (2015, S.105) empfahl diesbezüglich, spezifische Referenzen in den eigenen Daten zu vermeiden, die Identitäten offenbaren könnten. [55]

4.5 Dokumentation

Anonymisierung als einen Prozess zu verstehen, heißt auch, dass Forschende Entscheidungen über einen längeren Zeitraum hinweg treffen. Um sich diese zu merken, sie konsistent zu halten, später darstellen und rechtfertigen zu können, müssen sie dokumentiert und vereinheitlicht werden (CORTI et al. 2000, §24). REICHERTZ (2015, S.195-196) riet dazu, für jeden Fall ein "Datenblatt" anzulegen, "auf dem notiert wird, welche fiktiven Daten den Klardaten und Klarnamen" entsprechen. CORTI et al. (2000, §24) schlugen im Kontext der analyseorientierten Anonymisierung vor, jedes bearbeitete Transkript sorgfältig zu lesen und auf subtile Hinweise auf Personen, Orte oder Institutionen zu prüfen. THOMSON et al. (2005, §24) machten darauf aufmerksam, dass dies auch bei Teilautomatisierung z.B. durch Suchen und Ersetzen notwendig sei, um Rechtschreibfehler oder Namensvariationen (bspw. "Steffi" statt "Stefanie") nicht zu übersehen. Neben einem Pilottest, bei dem Kolleg:innen versuchen, auf Basis anonymisierter Daten eine Deanonymisierung vorzunehmen, regten sie an, eine Legende in die Kopfzeile des Transkripts zu integrieren, um den aktuellen Stand der Anonymisierung festzuhalten (z.B. ein Zeichensystem für bereits anonymisierte Daten, Namen und Orte). [56]

Bei einer publikationsorientierten Strategie empfiehlt es sich, ein Datenblatt pro Artikel, Vortrag oder Buch anzulegen. Selbst wenn das Textprojekt kurz erscheint, ist eine Dokumentation sinnvoll. In einem solchen Anonymisierungsprotokoll kann festgehalten werden, welche Fälle direkt zitiert, welche Pseudonyme für Namen oder Orte verwendet oder welche falschen Fährten warum gelegt wurden. Bei teilnehmendenorientierten Ansätzen muss gegebenenfalls "mit den Erforschten Rücksprache gehalten werden, ob aus ihrer Sicht die Anonymisierung hinreichend ist" (REICHERTZ 2015, S.197). Und bei allen Strategien sollte der Text dediziert mit Blick auf Anonymisierung gegengelesen werden, bevor er an Dritte gesendet wird. [57]

4.6 Zielgruppen

Rechtlich ist eine ausreichende Anonymität dann gegeben, wenn ein unverhältnismäßiger Aufwand (DSGVO, Art. 27) zur Deanonymisierung notwendig wäre. Forschungspraktisch stellt sich die Frage, wessen Aufwand hier gemeint ist. Diesbezüglich unterschied REICHERTZ (2015, S.196) zwischen einer Anonymisierung nach innen, die sich an die untersuchte Gruppe selbst richtet bzw. an deren Angehörige, Bekannte und Kolleg:innen, und einer Anonymisierung nach außen, d.h. die allgemeine Öffentlichkeit, Behörden usw. Dabei stellen die internen Dynamiken in beforschten Populationen sowie bestehende Kenntnisse über eine untersuchte Gruppe seitens des Zielpublikums besondere Herausforderungen dar. [58]

4.6.1 Anonymisierung nach innen

Es ist ein zentrales Dilemma von Anonymisierungsbestrebungen, dass Anonymisierung gerade dort am schwierigsten ist, wo sie am wichtigsten wäre: auf der lokalen Ebene, wo Teilnehmende von Angehörigen, Kolleg:innen oder Kolleg:innen erkannt werden könnten (NESPOR 2000, S.548). In ihrer Untersuchung befand SURMIAK (2018, §26), dass viele Forschende externer Anonymität mehr Wichtigkeit zusprachen als interner Anonymität ihrer Teilnehmenden. Auch ROTH und VON UNGER (2018) stellten im Kontext von Forschung zu Geflüchteten fest, dass die Herstellung interner Vertraulichkeit Grenzen habe, z.B. weil die Intervieworte in überfüllten Unterkünften dazu führten, dass Familienmitglieder, Freund:innen und Nachbar:innen von den Interviews und ihren Inhalten wüssten (s. auch AKESSON, HOFFMAN, EL JOUEIDI & BADAWI 2018, §35). Selbst eine schon oft – mit Kolleg:innen und auch im Interview – geteilte Erzählung kann auch ohne Personen oder Ortsnamen einen Hinweis auf die Identität der erzählenden Person darstellen (SURMIAK 2018, §51). [59]

Aussagen über Dritte sind in diesem Kontext besonders sensibel, z.B. mit Blick auf Geheimnisse und Vertraulichkeit im Kontext von Paar- und Familienstudien oder bei negativen Aussagen über Dritte (WILES et al. 2008, S.424). KÜHL (2020, S.67-68) deutete zusätzlich darauf hin, dass die Unterscheidung zwischen innerer und äußerer Anonymisierung auch Grenzen habe, da nie klar sei, wer Publikationen lese. Er betonte daher, die Möglichkeit, dass Ergebnisse gefunden und im Feld diskutiert würden, grundsätzlich zu berücksichtigen und entsprechend sorgfältig zu anonymisieren. [60]

4.6.2 Anonymisierung nach außen

Mit Blick auf die Öffentlichkeit und insbesondere Behörden als Publikum von anonymisierten Texten sind drei sensible Punkte zu unterscheiden: *Erstens* ist besondere Sorgsamkeit mit Blick auf vulnerable Personengruppen gefordert. ROTH und VON UNGER (2018) nannten bspw. Geflüchtete, bei denen mangelhafte Anonymisierung im Zweifel dazu führen könnte, dass kein Asyl gewährt würde, wenn Behörden bestimmte Aussagen zurückverfolgen könnten. Ähnlich betonte REICHERTZ (2015), ein zentraler Maßstab für die Qualität von Anonymisierung sei, "dass in den Daten dokumentierte kleine wie große Übertretungen der rechtlichen Normen nicht bestimmbar Personen zugeordnet werden können. Arbeitsämter dürfen also nicht informell arbeitende Leistungsempfänger erkennen und verfolgen können und Polizisten keine Straftäter" (S.197). [61]

Zweitens sind kleine Populationen eine besondere Herausforderung. Während es bei großen Grundgesamtheiten (bspw. Arbeitslose in Deutschland, Einwohner:innen Sachsens) oft schon genügt, mit Pseudonymen und Generalisierungen zu arbeiten, sind bei kleinen Samples invasivere Anonymisierungstechniken wie das Verfälschen in Betracht zu ziehen. Auch wenn Letzteres dazu führen könne, "dass die publizierte Datenanalyse an

Plausibilität verliert, da die Datenlage ja nicht mehr stimmt" (S.196), müsse dies hingenommen werden, "da der Schutz der Person oder Institution in diesem Fall Vorrang" (S.197) habe. [62]

Ähnlich verhält es sich *drittens* in kleinen Forschungskontexten (bspw. Dörfern, Unternehmen): Wenn bekannt wird, wo geforscht wurde, könnte eine Deanonymisierung leichter anhand spezifischer Äußerungen, Äußerungsweisen oder nur bestimmten Personengruppen bekannten Fakten gelingen. Entsprechend berichteten WILES et al. (2008, S.423), dass es für Forschende, die ihre Daten kontextspezifisch aufarbeiteten (z.B. für Evaluations-, Organisations- und Gemeindestudien) schwieriger war, Anonymisierungsfragen handzuhaben als für diejenigen, die eher allgemeine soziale Phänomene behandelten. Auch hier zeigt sich, dass Forschungsfrage und Anonymisierungsdesign eng zusammenhängen. [63]

4.6.3 Innen und Außen aus Forschendenperspektive

Aus Teilnehmendensicht ist die Öffentlichkeit ein externes Publikum und die eigenen Bekannten bilden interne Lesende. Jedoch können die Anonymisierungszielgruppen auch von den Forschenden aus gedacht es kann und ebenfalls zwischen einer inneren und einer äußeren Anonymisierungsrichtung unterschieden werden. [64]

Ersteres wirft die Frage nach Vertraulichkeit im sozialen oder akademischen Umfeld der Forschenden auf. WILES et al. (2008, S.426) wiesen darauf hin, dass Forschende in Alltagsgesprächen gegen ihr Anonymisierungsversprechen verstießen – sei es aus Versehen oder weil sie das auf der Suche nach methodischem Rat, inhaltlichem Austausch oder emotionaler Unterstützung hinnehmen. Sie betonten Notwendigkeit, sich untereinander bzgl. des "letting something slip" (S.426) zu sensibilisieren und zu fragen, was Vertraulichkeit tatsächlich in einer wissenschaftlichen Praxis bedeute, in der die Diskussion empirischen Materials essenziell sei. Ähnlich kam auch SURMIAK (2018, §39-40) zu dem Schluss, dass bezüglich Unterhaltungen mit anderen Wissenschaftler:innen tlw. unklar sei, was in diesem Kontext als Wahrung von Vertraulichkeit gelten könne und ob bspw. die Datenschutzvereinbarungen Alltagsgespräche Forschender mit Kolleg:innen oder Freund:innen umfassten. Die von SURMIAK befragten Forschenden hatten diesbezüglich weit auseinanderliegende Standards: Während einige beiläufige Bemerkungen über Teilnehmende in Form von anekdotischen Felderzählungen selbst ohne Namen oder Orte als unethisch verurteilten, redeten andere sogar mit Gatekeeper:innen über Beforschte und schätzten das als unproblematisch ein (§42). [65]

Nach außen wäre mit Blick auf die Forschenden zum Beispiel zu fragen, wie und ob es möglich ist, anderen Wissenschaftler:innen Zugang zu bereits bestehenden Erkenntnissen zu öffnen. KÜHL (2020) brachte es auf den Punkt: "Eine Frage, die ich für ebenso relevant wie die Sekundäranalyse qualitativer Daten halte, ist, wie bei strikter Anonymisierung anderen Forscherinnen und Forschern anschließende Folgestudien über die gleichen Organisationen ermöglicht werden

können" (S.65). Solche Fragen des Austauschs zu geteilten Fällen wurden bisher kaum diskutiert. [66]

Beachtenswert sind hier neuere Entwicklungen bei der Sekundärnutzung von qualitativen Daten. So wurde im Forschungsdatenzentrum [Qualiservice](#) bspw. ein Konzept der "flexiblen Anonymisierung" (MOZYGEMBA 2022, S.8) entwickelt, bei dem je nach Forschungsfrage unterschiedliche Anonymisierungsgrade für Orte, Berufe oder individuelle Charakteristika im Transkript angegeben werden können. Wenn kleinteilige Nennungen von Berufen für die Forschungsfrage relevant sind, können dann Ortsbezeichnungen oder andere Befragtenmerkmale stärker anonymisiert werden. Ähnlich wiesen CORTI et al. (2000, §30) darauf hin, dass auch ein selektiver Zugang zu Informationen ein gängiger Weg sein könne, um zumindest forschungsbezogen die Balance zwischen Anonymisierung und Analyse zu halten. [67]

5. Anwendungsbeispiel

5.1 Aufgabenstellung

Im Rahmen dieses Beitrags laden wir dazu ein, grundlegende Aspekte der Anonymisierung durch die Bearbeitung eines fiktiven Transkripts selbst praktisch nachzuvollziehen. Hierzu haben wir anhand eigener Forschungserfahrungen mit Wegzugmotiven Jugendlicher in ländlichen Regionen ein Beispieltranskript einer Gruppendiskussion erstellt, das hinsichtlich der genannten Orte, Personen, Einstellungen und Beziehungen fiktiv ist, in dem jedoch Aspekte aus bestehenden Erhebungen zu Abwanderungskulturen exemplarisch illustriert werden (LEIBERT 2016; MEYER 2018a; MEYER & LEIBERT 2021; WIEST 2016). Die fiktiven Personen sind Mitglieder eines Breitensportvereins, haben unterschiedliche Geschlechter und nehmen freiwillig an der Gruppendiskussion teil. Sie sind einander vertraut, teils freundschaftlich verbunden; zwischenmenschliche Konflikte innerhalb der Gruppen treten nicht zutage. [68]

Die Fragestellung, vor deren Hintergrund das Transkript betrachtet werden soll, lautet: Welche Motive und Entscheidungsfindungsprozesse führen Jugendliche in ländlichen Regionen dazu, nach dem Ende ihrer schulischen Ausbildung aus ihrer Heimatregion fortzuziehen? Wir schlagen zwei Bearbeitungsschritte vor und empfehlen, das Transkript selbst zu bearbeiten, sich nach dem jeweiligen Schritt den dazugehörigen Diskussionsabschnitt durchzulesen und dann das eigene Resultat mit unseren Ergebnissen zu vergleichen. [69]

5.2 Schritt 1: Markieren unterschiedlicher Datenarten

Hierzu können die Leser:innen das im [Anhang 1](#) hinterlegte unbearbeitete Beispieltranskript nutzen und Textstellen mit Markup-Werkzeugen markieren. Ziel des ersten Schrittes ist weniger die trennscharfe Zuordnung jedes einzelnen Wortes zu einer Datenkategorie als vielmehr die Identifikation von anonymisierungsrelevanten Daten, die je nach Art Gegenstand unterschiedlicher Anonymisierungstechniken sein können. Die Zuordnung von Datenkategorien ermöglicht es, Entscheidungen über spezifische Anonymisierungsschritte zu treffen. Im Beispieltranskript sollen somit Identifikationsdaten, Kontextdaten und sensible Daten identifiziert werden (s. Abschnitt 3). Wir haben in [Anhang 2](#) eine Version des Beispieltranskripts hinterlegt, in der diese exemplarisch markiert wurden. Wichtig ist, dass abhängig von dem Vorwissen über die Details der Erhebungssituation auch mehr Aspekte des Transkripts als Kontextdaten dienen können. Wir sind bei der Erstellung der Vorlage davon ausgegangen, dass der anonymisierenden Person nur begrenzte Detailinformationen bekannt sind. [70]

Im angehängten Beispiel für markierte Datenarten ist ersichtlich, dass zuerst die durch das Transkript angegebenen Namen der einzelnen Teilnehmenden Identifikationsdaten darstellen. Auch all jene Angaben zu Namen, Unternehmen, Orten etc. sind als solche zu kategorisieren, wenn hier spezifische Verbindungen zu anderen Daten hergestellt werden können. So ist die Angabe, dass man in einer unspezifizierten Vergangenheit in Wolfsburg gelebt hat (Z.16-18), ein Kontextdatum, da Wolfsburg eine Großstadt mit vielen Einwohner:innen ist. Wenn mit Orten aber zudem Arbeitgeber:innen oder Berufe verbunden werden (bspw. Wolfsburg – VW – Autoingenieur:in), können hier eindeutigere Zuordnungen entstehen und diese daher zu Identifikationsdaten werden. [71]

Hinsichtlich der Kontextdaten sind alle unspezifischen Orts-, Namens- und Beziehungsangaben relevant, die nicht unmittelbar zu einer Deanonymisierung führen können, aber Lesenden mit Zusatzwissen eine Deanonymisierung ermöglichen. So ist bspw. in Zeilen 42-45 der Beziehungsstatus zwischen den Praktikumsarbeitgeber:innen von "Ben" als "Freunde der Familie" angegeben. Jegliche Informationen, die also aus dem Ort des Praktikums folgen (bspw. Bens kritische Haltung hinsichtlich der Arbeitsinhalte und des Verhaltens Dritter), könnten also auf die Inhaber:innen des Betriebs rückwirken, oder die Wertungen und die Beziehung zu den Inhaber:innen könnten wiederum Rückschlüsse auf den spezifischen Arbeitsort erlauben. Sensible Daten, die potenziell Grundlage für Diskriminierung bieten oder aus anderen Gründen zu negativen Konsequenzen führen können, finden sich im vorliegenden Beispiel in drei Formen: Zu beachten sind hier erstens Wertungen, die die Teilnehmenden selbst vornehmen (bspw. Z.42-45 zu Praktikumsinhalten) oder für Dritte tätigen (bspw. Z.30-32 zu möglicher Korruption im ehemaligen Arbeitsumfeld des Vaters). Diese Wertungen sind sensibel, da sie bei Deanonymisierung zu negativen sozialen, arbeits- oder strafrechtlichen Konsequenzen für die Teilnehmenden und ihre Angehörigen führen können. Zweitens sind vereinzelt persönliche Informationen zum Gesundheitsstatus Dritter (bspw. Z.33 zum Burnout des Vaters) erkennbar, welche ggf. zu Problemen bei Wiederbeschäftigung, Verbeamtung oder

Versicherung führen können. Drittens ist der Diskussionsstrang um das Mobbing bezüglich Homosexualität (Z.53-59) als sensibel einzustufen, weil hier bei Deanonymisierung die sexuelle Identität von "Dennis" ohne dessen Zustimmung verbreitet werden könnte, woraus neue Mobbingverfahren resultieren könnten. [72]

Insgesamt können mögliche Diskrepanzen zwischen unseren Markierungsvorschlägen und den Markierungen der Lesenden hinsichtlich zweier Aspekte entstehen:

- *Das notwendige Loslösen von einer rein rechtlichen Definition personenbezogener Daten:* Wenngleich deutsche Datenschutzgesetze und -standards den Rahmen geben, so ist das Ziel des Artikels dennoch die Komplexitäten bei der Anonymisierung qualitativer Daten aufzuzeigen. So sind manche Markierungen nur dann relevant, wenn vorher keine Anonymisierung vorgenommen wurde. Viele Kategorien personenbezogener Daten in der rechtlichen Debatte kommen in der Praxis qualitativer Forschung selten vor (bzw. IP-Adresse, Heimatadresse, Körpergröße). Dafür ergeben sich durch die teils dichten Beschreibungen eines sozialen Milieus insbesondere durch die Verknüpfung von Äußerungen Datenfragmente, die allein genommen wenig relevant sind (bspw. Vater bei VW in Wolfsburg, Mutter Lehrerin in Wolfsburg), die durch Verknüpfungen mit Hintergrundwissen jedoch deanonymisierungsrelevant werden. Mit Blick auf Kennbeziehungen Teilnehmender könnten hier Fragen aufkommen wie: "Kommt Andreas nicht aus Wolfsburg, seine Mutter ist Lehrerin und sein Vater arbeitet in der Werkstatt?" Hierdurch sind wiederum die Erzählungen über die Einschätzung der Mutter, dass an ihrer Schule "nur Dorfdeppen" seien, von Bedeutung, da diese – wenn sie den geschützten Bereich des Privaten verließen – zu negativen Konsequenzen führen könnten.
- *Die Unterscheidung zwischen Identifikations- und Kontextdaten:* Im Prinzip bestehen Identifikationsdaten aus konkreten Orts-, Namens- oder Unternehmensnennungen und ggf. daran anschließende Daten (wie Distanzen). Kontextdaten sind jene nicht direkt auf Personen oder andere beziehbare Daten, die jedoch kontextuell Gewicht haben und zudem eine Deanonymisierung durch Verknüpfung ermöglichen können. Zu unterschiedlichen Einschätzungen kann es hier kommen, weil der Aufwand für eine Deanonymisierung nicht immer objektiv bestimmt werden kann und zur Bewertung des Kontexts ggf. Hintergrundwissen herangezogen wird. So kann es sein, dass eine Teilnehmende angibt, in einem kleinen Ort zu leben, in dem ihre Mutter Bäckerin ist. Wenngleich dies generell als Kontextdatum gewertet würde, kann in Abhängigkeit von der Größe des Ortes und der dortigen Anzahl der Bäckereien diese Angabe de facto so eindeutig sein, dass sie einem Identifikationsdatum gleichkommt. [73]

5.3 Schritt 2: Eruiieren möglicher Anonymisierungstechniken

In diesem Schritt werden anhand der markierten Datenarten für konkrete Textstellen spezifische Anonymisierungstechniken erörtert. Während wir in Abschnitt 4.3 einen systematisierenden Blick auf etablierte Anonymisierungsinstrumente bieten, haben wir in Tabelle 3 noch einmal die möglichen Anonymisierungstechniken auf die verschiedenen Datenarten bezogen. In dieser Tabelle gehen wir davon aus, dass Identifikationsdaten Dritter immer zu anonymisieren sind, während Identifikationsdaten Teilnehmender dann nicht anonymisiert werden müssen, wenn diese dem explizit zustimmen.

	Organisationsdaten	Identifikationsdaten	Kontextdaten	Bewertungen/ sensible Daten
Äußerung über sprechende Person	Getrennt aufbewahren, Kennkürzel statt Klarnamen	Beibehalten (bei informierter Einwilligung), pseudonymisieren, verfälschen, nicht nennen	Beibehalten (bei informierter Einwilligung), nicht nennen, generalisieren, verfälschen, nicht veröffentlichen	Beibehalten (bei vorheriger Anonymisierung), generalisieren, nicht veröffentlichen
Äußerung über Dritte	(Fallen nur beim Schneeballverfahren an)	Gleichwertig ersetzen, verfälschen, nicht nennen	Beibehalten (bei vorheriger Anonymisierung) nicht nennen, generalisieren, nicht veröffentlichen	Beibehalten (bei vorheriger Anonymisierung), generalisieren, nicht veröffentlichen

Tabelle 3: Anonymisierungstechniken in Bezug auf Datenarten [74]

In [Anhang 3](#) haben wir zwei mögliche Anonymisierungsergebnisse dokumentiert, die durch unterschiedliche Personen blind erstellt wurden und Auskunft über die möglichen und gewählten Anonymisierungstechniken gegeben. Grundlegend sind folgende Entscheidungen zu treffen:

1. *Entscheidung zwischen Kennkürzeln und Pseudonymen* (bspw. Z.9): Hier wägt man insbesondere zwischen Lesbarkeit, der Kontextualisierung und der Notwendigkeit einer Abstraktion ab. Während pseudonymisierte Manuskripte durch Beibehaltung von gleichwertig ersetzten Namen lesbarer sind, können durch die Namen bei Lesenden Fremdassoziationen hervorgerufen werden und damit deren Interpretationen beeinflussen (LOCHNER 2017). Dieses Problem kann durch die Verwendung von Kennkürzeln vermieden werden.

2. *Entscheidung zwischen verdeckter Anonymisierung und Markieren anonymisierter Textstellen:* In der Praxis wird dies uneinheitlich gehandhabt. So kann es für die Transparenz gegenüber Lesenden förderlich sein zu verdeutlichen, dass und in welchem Ausmaß anonymisiert wurde (bspw. durch Nutzung eckiger Klammern für Textstellen, die anonymisierend verändert wurden). Andererseits kann solches Markieren wiederum Hinweise geben, die einer Deanonymisierung dienlich sein können. Wir würden nicht per se empfehlen, die anonymisierten Textstellen immer kenntlich zu machen. Vielmehr ist es abhängig von der situativen Einschätzung des Deanonymisierungspotenzials, ob Anonymisierung transparent erkennbar sein sollte.
3. *Entscheidung zwischen Verallgemeinerungen und Weglassung* (bspw. Z.19-20): Bei Verallgemeinerungen im Kontext von Anonymisierungen wird anstelle eines spezifischen realweltlichen Verweises (eines Orts, einer Person, eines Berufs usw.) ein textliches Substitut mit einem höheren Abstraktionsgrad (ein Bundesland, die Rolle einer Person, eine Branche) genutzt. Im vorliegenden Beispiel ist erkennbar, dass man vielfältig verallgemeinern kann: durch Ersetzen oder Nicht-Nennung eines Wortes (hier: "Werkstatt" statt Werkstatt und Autohaus in der Legatstraße) ebenso wie durch neue Satzschöpfungen ("neue Firma (...) in derselben Branche").
4. *Entscheidung zwischen Verfälschung und Verallgemeinerung* (bspw. Z.21-25): Im vorliegenden Beispiel werden ehemalige und aktuelle Arbeitsorte der Mutter sowie ihr Beruf genannt. Vor dem Hintergrund der grundsätzlichen Fragestellung (s. Abschnitt 5.1) sind diese spezifischen Orte irrelevant, ihre Nennung birgt aber Deanonymisierungspotenzial. Bei einer Verfälschung werden faktische Angaben (z.B. ein Ort) durch inkorrekte ersetzt, um eine falsche Spur zu legen. Hier wird mit einer Verallgemeinerung, bei der bspw. die Ortsspezifität entfernt wird, unter Umständen weniger in das Manuskript eingegriffen. Eine Entscheidung für oder gegen Verfälschungen kann aus forschungsethischen Erwägungen geschehen. Werden systematische Verallgemeinerungen mit einzelnen Verfälschungen kombiniert, verringert dies Deanonymisierungsmöglichkeiten deutlich.
5. *Entscheidung zwischen Weglassung und Verallgemeinerung* (bspw. Z.29, 30-32, 42-45): Im vorliegenden Fall haben wir in Beispiel 1 den aktuellen Arbeitsort des Vaters vollständig weggelassen, da er inhaltlich nicht von Bedeutung ist (Z.29). Weggelassen wurde in beiden Beispielen in den Zeilen 30-32 auch die Äußerung zu korrupten Handlungen. Gleichzeitig erfolgte eine Verallgemeinerung bezüglich der damaligen wirtschaftlichen Lage, da der Grund der Migration in eine ländliche Region (Burnout des Vaters) und damit die Kontextualisierung von Ländlichkeit als bewusste Wahl inhaltlich relevant sind. Inwieweit eine Weglassung gerechtfertigt ist, ergibt sich aus der Abwägung der Größe des untersuchten sozialen Feldes und der Einzigartigkeit der Merkmalskombination (Großstadt – Firma – Burnout – Migration). Auch Verfälschung könnte hier eine Option sein.
6. *Gleichwertig ersetzen und Verallgemeinern:* In Z.56 erzählt eine Person von Mobbing Erfahrungen aufgrund eines unfreiwilligen Outings als homosexuell. Inhaltlich ist dieser Aspekt wichtig, weil er den Kontext des subjektiven

Abwanderungswissens darstellt. Gleichzeitig bilden sexuelle Identität und bestehende Mobbing Erfahrungen hier die Grundlage des besonderen Umganges mit solchen Textstellen: Während in Beispiel 1 entschieden wurde, die Ausrichtung der Sexualität zu ändern, wurde diese in Beispiel 2 beibehalten. In beiden Fällen wurde der Zeitraum des Outings verallgemeinert. Die Basis dieser Entscheidungen bildet die Abwägung, wie besonders und bekannt ein Merkmal potenziell lokal ist, und welche Gefahr für den Teilnehmenden durch dessen Nennung bestünde. In solchen Fällen gehen die Nutzung von Pseudonymen mit textstellenspezifischen Anonymisierungstechniken und manuskriptweiten Anonymisierungsabwägungen einher: Wenn das sensible Datum Outing/Mobbing inhaltlich relevant ist, die untersuchte Gruppe jedoch klein, kann ebenso eine Löschung bei Priorisierung des Teilnehmendenschutzes sinnvoll sein. Wenn es jedoch beibehalten wird, muss der gesamte Text mit höchster Genauigkeit auf lückenlose Anonymität geprüft werden.

7. *Entscheidung bzgl. Beibehaltungen bei vorheriger Anonymisierung* (bspw. Z.33-34, 57-59): Der Bezug auf das Burnout ist sensibel, da dies ein rechtlich relevantes Gesundheitsdatum darstellt und im Falle einer Deanonymisierung negative berufliche oder private Konsequenzen drohen könnten. Gleichzeitig ist die Burnout-Erkrankung bedeutsam für die Entscheidung, aus einer Großstadt in eine ländliche Region zu ziehen, und damit von inhaltlicher Relevanz. Dies trifft auch auf die Schilderung der Mobbing Erfahrung in den Zeilen 57-59 zu. Eine Priorität sollte also darauf liegen, den Sachverhalt beizubehalten und die Deanonymisierung selbst vollständig zu verunmöglichen. Jedoch ist als Alternative auch eine Verallgemeinerung im Sinne von "Krankheit eines Familienmitgliedes" denkbar. [75]

6. Schlussfolgerungen: Anonymisierung lehren und lernen

Eine Verwendung des *Begriffs* des "Anonymisierens" kann den Blick auf die praktische Komplexität verstellen, mit der die *Praxis* des Anonymisierens in qualitativer Forschung einhergeht. Aufgrund dieser Komplexität muss Anonymisierungsanlässen, -techniken und -bedingungen in der methodischen Ausbildung didaktische Aufmerksamkeit zukommen. Der Impuls, dies *jetzt* zu fordern, ergibt sich aus Unsicherheiten bezüglich der Umsetzung aktueller Gesetze, aus den Herausforderungen des immer häufiger geforderten Forschungsdatenmanagements, aus den zunehmenden Ansprüchen an Transparenz im Publikationsprozess und aus den technischen Möglichkeiten, große Datenmengen zu verknüpfen. Gleichzeitig besteht weiterhin ein Grunddilemma Sozialforschender: Diese haben ein großes Interesse, den jeweiligen Kontext widerzuspiegeln, möglichst viele Daten unverändert zu nutzen und diese dicht darzustellen. Dabei ist es notwendig, empathisch und opportun, die Beforschten vor negativen Konsequenzen durch Forschung zu schützen (SCHEPER-HUGHES 2000, S.128). Während auf regulatorischer Seite bzw. von Förderinstitutionen stärkere Vorgaben eingeführt werden, plädierten bspw. WILES et al. (2008, S.427) vor allem für eine angemessene Schulung und Unterstützung von Wissenschaftler:innen "so that they have the ethical literacy

that enables them to make appropriate ethical decisions that operate in the best interests of research participants within the context of a specific study".

Konsequenzen aus den dargestellten Techniken der Anonymisierung für die methodische Ausbildung können in programmatischer und praktischer Hinsicht gezogen werden. [76]

6.1 Programmatische Schlussfolgerungen

- *Anonymisierung als dialektische Praxis:* Im Zuge von Datentransparenzpostulaten, KI-gestützten Datenverarbeitungsweisen und wachsenden Möglichkeiten der Datenverknüpfung kommt der Frage der Datengewinnung, -bearbeitung und -darstellung kontinuierlich Bedeutung zu. Die Frage, wie und in welchem Maße zu anonymisieren ist, kann nur in Abwägung der aktuellen technischen Möglichkeiten beantwortet werden. Dies führt dazu, dass die *Debatte* um Anonymisierung als eine Daueraufgabe von Wissenschaft begriffen werden muss, und die Praktiken des Anonymisierens vor dem Hintergrund technischer Möglichkeiten fortwährend auf ihre Angemessenheit befragt werden müssen.
- *Anonymisierung als integrierte Praxis:* Leitende Aspekte und Zeitpunkte im Abwägungsprozess um Anonymisierung sind Fragestellung, Ausmaß und Kosten- bzw. Zeitaufwand, Absprachen und Vereinbarungen mit Beforschten, aktuelle rechtliche Rahmenbedingungen und die Form der Publikation der Ergebnisse. Beinhaltet Forschung bspw. die Präsentation der Daten im Rahmen einer öffentlichen Veranstaltung in der Erhebungsregion, über die womöglich in der Presse oder auf der Projektwebseite berichtet wird, ist der potenzielle Aufwand der möglichen Deanonymisierung sehr gering; bei Anonymisierungserwägungen zu früheren Zeitpunkten im Forschungsprozess sollten diese Möglichkeit beachtet werden.
- *Anonymisierung als zu fokussierende Praxis:* Anonymisierungstechniken werden häufig als Nebenaspekte qualitativer Sozialforschung behandelt. Praktisch finden wir Hinweise zu genutzten Anonymisierungstechniken in der Regel am Rande methodischer Darstellungen. Notwendig sind hingegen detaillierte Darstellungen, die Dokumentation und Diskussion von Missgeschicken (ohne Sanktion der Urheber:innen) und Metastudien über genutzte Techniken.
- *Anonymisierung als dialogische Praxis:* Aufgrund der Komplexität gerade von Kontextinformationen und der Problematik kleiner Gemeinschaften scheint es uns nicht sinnvoll, allein allgemeine technische Handreichungen als Mittel der Lehre einzusetzen. Vielmehr müssen die Untiefen von Anonymisierungen, Zweifelsfälle, Risiken und vieles mehr abgewogen werden, um die von WILES et al. (2008) geforderte Ausbildung und Unterstützung sowie letztlich einen sicheren und reflektierten Umgang mit Anonymisierung für Forschende gewährleisten zu können. Dementsprechend sind dialogische Formate im Sinne von Weiterbildungen und Arbeitsgemeinschaften besser geeignet, eine diskursive Reflexion zu ermöglichen. [77]

6.2 Methodische Schlussfolgerungen

- *Daten- und fallspezifische Anonymisierungstechniken nutzen:* Datenartenspezifische Anonymisierungsmethoden stellen eine Weise dar, nachvollziehbar und balanciert qualitative Daten zu verändern. Während bspw. Identifikationsdaten durch gleichwertige Substitute wie Namen ersetzt werden können, sind bei Kontextdaten andere Methoden notwendig, die fallspezifisch abgewogen werden müssen. In manchen Fällen können Daten bei vorliegender Einwilligung beibehalten werden, manchmal müssen sie jedoch verfälscht, verallgemeinert oder sogar gelöscht werden. Wir raten hier explizit dazu, der anonymisierungssensiblen Nachbearbeitung von Primärdaten genügend Zeit und Aufmerksamkeit im Forschungsprozess einzuräumen, um diese Schritte abwägen und durchführen zu können.
- *Anonymisierungspraxis protokollieren und diskutieren:* Mit dem Ziel, die Entscheidungen zu anonymisierungssensiblen Änderungen am Text nachvollziehbar zu gestalten, geht die Dokumentation dieser Entscheidungen einher. Dies kann bspw. in Form einer Art Regelbuchs als Anleitung zum Anonymisieren geschehen. Weiterhin sind Markierungen anonymisierter oder anderweitig veränderter Textstellen, das getrennte Aufbewahren von Kennkürzeltabellen oder die Formulierung systematischer Darstellungen, welche Datenarten entlang welcher Leitlinien mittels welcher Anonymisierungsmethode bearbeitet wurden, möglich. Ein wichtiger praktischer Aspekt dabei ist auch die publikationsspezifische Diskussion mit Kolleg:innen. Etwa könnte das Veröffentlichungsprotokoll vorsehen, dass jede Veröffentlichung, bevor sie das Haus verlässt, im Team unter dem Aspekt der Anonymisierung kritisch geprüft wird.
- *Anonymisierung ggf. als partizipativen Prozess praktizieren:* In Fällen, in denen kleine Gemeinschaften beforscht werden oder besonders ernste soziale, strafrechtliche oder politische Konsequenzen für Beforschte drohen, kann es angeraten sein, die gewählten Anonymisierungstechniken in Abstimmung mit den Beforschten in einer Art Feedback-Prozess durchzuführen (MEYER 2018b). In solchen Fällen kommt der rigorosen und wiederholten Aufklärung der Teilnehmenden über die Konsequenzen ihrer Wahl ethische Bedeutung zu. [78]

Anhang

[Anhang 1:](#) Beispieltranskript unbearbeitet

[Anhang 2:](#) Beispieltranskript mit markierten Datenarten

[Anhang 3:](#) Vergleich der beiden Anonymisierungen

Literatur

- Akesson, Bree; Hoffman, David A. "Tony"; El Joueidi, Samia & Badawi, Dena (2018). "So the world will know our story": Ethical reflections on research with families displaced by war. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 19(3), Art. 5, <https://doi.org/10.17169/FQS-19.3.3087> [Datum des Zugriffs: 31. März 2023].
- Bacher, Johann & Horwarth, Ilona (2011). *Einführung in die Qualitative Sozialforschung. Teil 2*. Linz: Johannes Kepler Universität Linz, https://www.jku.at/fileadmin/gruppen/119/AES/Lehre/Bacc-Pruefung/SkriptTeil2ws11_12.pdf [Datum des Zugriffs: 17. November 2022].
- Bachmann, Götz; McHardy, Julien; Knecht, Michi & Zurawski, Nils (2021). Toward a kaleidoscopic understanding of anonymity. In Anon Collective (Hrsg.), *Book of anonymity* (S.16-34). Milky Way, Earth: punctum books.
- Bayerischer Landesbeauftragter für den Datenschutz (Hrsg.) (2021). *Die Einwilligung nach der Datenschutz-Grundverordnung. Orientierungshilfe*, <https://www.datenschutz-bayern.de/datenschutzreform2018/einwilligung.pdf> [Datum des Zugriffs: 28. März 2023].
- Benner, Angela & Löhe, Julian (2019). Die informierte Einwilligung auf Tonband: Analyse im Rahmen einer qualitativen Interviewstudie mit älteren Menschen aus forschungsethischer und rechtlicher Perspektive. *Zeitschrift für Qualitative Forschung*, 20(2), 341-356, <https://doi.org/10.3224/zqf.v20i2.08> [Datum des Zugriffs: 31. März 2023].
- Bischof, Susann; Lengerer, Franziska & Meyer, Frank (2022). Interaktionsspuren im digitalen Raum: Zum Spannungsfeld von Öffentlichkeit und Anonymität in qualitativen Forschungsprozessen. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 23(3), Art. 9, <https://doi.org/10.17169/fqs-23.3.3893> [Datum des Zugriffs: 31. März 2023].
- Corden, Anne (2007). Using verbatim quotations in reporting qualitative social research. A review of selected publications. *Working paper*, ESRC 2226, Social Policy Research Unit, University of York, UK, <https://www.york.ac.uk/inst/spru/pubs/pdf/verbpubs.pdf> [Datum des Zugriffs: 6. Februar 2023].
- Corden, Anne & Sainsbury, Roy (2006a). *Using verbatim quotations in reporting qualitative social research. Researchers' views*, <https://www.york.ac.uk/inst/spru/pubs/pdf/verbquoteresearch.pdf> [Datum des Zugriffs: 6. Februar 2023].
- Corden, Anne & Sainsbury, Roy (2006b). *Using verbatim quotations in reporting qualitative social research. The views of research users*, <https://www.york.ac.uk/inst/spru/pubs/pdf/verbusers.pdf> [Datum des Zugriffs: 6. Februar 2023].
- Corti, Louise; Day, Annette & Backhouse, Gill (2000). Confidentiality and informed consent: Issues for consideration in the preservation of and provision of access to qualitative data archives. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(3), Art. 7, <https://doi.org/10.17169/fqs-1.3.1024> [Datum des Zugriffs: 31. März 2023].
- European Data Protection Board (2020). Guidelines 05/2020 on consent under Regulation 2016/679. *Version 1.1*, 4. Mai, https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf [Datum des Zugriffs: 31. März 2023].
- Hopf, Christel (2016 [2010]). Forschungsethik und qualitative Forschung. In Wulf Hopf & Uwe Kuckartz (Hrsg.), *Schriften zu Methodologie und Methoden Qualitativer Sozialforschung* (S.195-206). Wiesbaden: Springer VS.
- Inter-Agency Standing Committee (2014). *Recommendations for conducting ethical mental health and psychological research in emergency settings*, https://publications.iom.int/system/files/pdf/iasc_mental_health_2014.pdf [Datum des Zugriffs: 17. Februar 2023].
- Kinder-Kurlanda, Katharina & Watteler, Oliver (2015). Hinweise zum Datenschutz: Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. *GESIS Papers* 2015-1, GESIS-Leibniz-Institut Für Sozialwissenschaften, <https://doi.org/10.21241/ssoar.43183> [Datum des Zugriffs: 19. April 2022].
- Kühl, Stefan (2020). Zwischen Präzision und Anonymisierung. Wie weit muss man bei der Verfälschung wissenschaftlicher Daten gehen?. *Soziologie*, 49(1), 62-71, https://soziologie.de/fileadmin/user_upload/zeitschrift/volltexte/Kuehl_Heft_1_2020_Soziologie.pdf [Datum des Zugriffs: 3. März 2023].
- Leibert, Tim (2016). She leaves, he stays? Sex-selective migration in rural East Germany. *Journal of Rural Studies*, 43, 267-279.

- Lochner, Barbara (2017). "Kevin kann einfach auch nicht Paul heißen". Methodologische Überlegungen zur Anonymisierung von Namen. *Zeitschrift für Qualitative Forschung*, 18(2), 283-296, <https://doi.org/10.3224/zqf.v18i2.07> [Datum des Zugriffs: 31. März 2023].
- Medjedović, Irena & [Witzel, Andreas](#) (2010). *Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Metschke, Rainer & Wellbrock, Rita (2002). *Datenschutz in Wissenschaft und Forschung*. Berlin: Berliner Beauftragter für Datenschutz und Informationsfreiheit, <https://www.forschungsdaten-bildung.de/files/metschkewellbrock2002.pdf> [Datum des Zugriffs: 28. März 2023].
- Meyer, Frank (2018a). Navigating aspirations and expectations: Adolescents' considerations of outmigration from rural eastern Germany. *Journal of Ethnic and Migration Studies*, 44(6), 1032-1049.
- Meyer, Frank (2018b). Yes, we can(?) Kommunikative Validierung in der qualitativen Forschung. In Frank Meyer, Judith Miggelbrink & Kirsten Beurskens (Hrsg.), *Ins Feld und zurück – Praktische Probleme qualitativer Forschung in der Sozialgeographie* (S.163-168). Berlin: Springer.
- Meyer, Frank & Leibert, Tim (2021). On the role of cultures of (out-)migration in the migration decisions of young people in shrinking regions of Central Germany. *Geographica Helvetica*, 76(3), 335-345, <https://doi.org/10.5194/gh-76-335-2021> [Datum des Zugriffs: 31. März 2023].
- Meyermann, Alexia & Porzelt, Maïke (2014). Hinweise zur Anonymisierung von qualitativen Daten. Version 1.0. *forschungsdaten bildung informiert 1*, Forschungsdatenzentrum Bildung, Deutsches Institut für Internationale Pädagogische Forschung, <https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-1.pdf> [Datum des Zugriffs: 12. November 2023].
- Meyermann, Alexia & Porzelt, Maïke (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung (2. Aufl.). *forschungsdaten bildung informiert 6*, Verbund Forschungsdaten Bildung, <https://doi.org/10.25656/01:21990> [Datum des Zugriffs: 3. August 2021].
- Moosa, Dheebea (2013). Challenges to anonymity and representation in educational qualitative research in a small community: A reflection on my research journey. *Compare: A Journal of Comparative and International Education*, 43(4), 483-495.
- Mozygemba, Kati (2022). QualiAnon – Das Qualiservice Anonymisierungstool für Textdaten. *Präsentation*, Workshop der NFDI-Sektion Ethical Legal & Social Aspects, Berlin, 12.-13. Oktober, https://www.nfdi.de/wp-content/uploads/2022/10/Mozygemba_ELSA-Workshop.pdf [Datum des Zugriffs: 17. Februar 2023].
- Nespor, Jan (2000). Anonymity and place in qualitative inquiry. *Qualitative Inquiry*, 6(4), 546-569.
- RatSWD (Rat für Sozial- und Wirtschaftsdaten) (Hrsg.) (2014). Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten. *RatSWD Working Paper 238*, https://www.konsortswd.de/wp-content/uploads/RatSWD_WP_238.pdf [Datum des Zugriffs: 28. März 2023].
- RatSWD (Rat für Sozial- und Wirtschaftsdaten) (Hrsg.) (2017a). Handreichung Datenschutz. *RatSWD Output 5*, <https://doi.org/10.17620/02671.6> [Datum des Zugriffs: 31. März 2023].
- RatSWD (Rat für Sozial- und Wirtschaftsdaten) (Hrsg.) (2017b). Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. *RatSWD Output 9(5)*, <https://doi.org/10.17620/02671.1> [Datum des Zugriffs: 10. August 2022].
- Rat SWD (Rat für Sozial- und Wirtschaftsdaten) (Hrsg.) (2020). Handreichung Datenschutz (2. Aufl.). *RatSWD Output 8(6)*, <https://doi.org/10.17620/02671.50> [Datum des Zugriffs: 10. August 2022].
- [Reichertz, Jo](#) (2015). Wie mit den Daten umgehen? Persönlichkeitsrechte und Datenschutz in der Qualitativen Sozialforschung. *Soziologie*, 44(2), 186-202.
- [Rosenthal, Gabriele](#) (2015). *Interpretative Sozialforschung. Eine Einführung* (5. Aufl.). Weinheim: Beltz Juventa.
- [Roth, Wolff-Michael](#) & [von Unger, Hella](#) (2018). Current perspectives on research ethics in qualitative research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 19(3), Art. 33, <https://doi.org/10.17169/fqs-19.3.3155> [Datum des Zugriffs: 10. August 2022].
- Sachs, Ulrich (2010). Datenschutzrechtliche Bestimmbarkeit von IP-Adressen. *Computer und Recht*, 26(8), 547-552.

Saunders, Benjamin; Kitzinger, Jenny & Kitzinger, Celia (2015). Anonymising interview data: Challenges and compromise in practice. *Qualitative Research*, 15(5), 616-632, <https://doi.org/10.1177/1468794114550439> [Datum des Zugriffs: 17. Februar 2023].

Scheper-Hughes, Nancy (2000). Ire in Ireland. *Ethnography*, 1(1), 117-140.

Surmiak, Adrianna D. (2018). Confidentiality in qualitative research involving vulnerable participants: Researchers' perspectives. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 19(3), Art. 12, <https://doi.org/10.17169/FQS-19.3.3099> [Datum des Zugriffs: 17. August 2022].

Taylor, Rebecca (2015). Beyond anonymity: Temporality and the production of knowledge in a qualitative longitudinal study. *International Journal of Social Research Methodology*, 18(3), 281-292.

Thomson, Denise; Bzdel, Lana; Golden-Biddle, Karen; Reay, Trish & Estabrooks, Carole A (2005). Central questions of anonymization: A case study of secondary use of qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(1), Art. 29, <https://doi.org/10.17169/fqs-6.1.511> [Datum des Zugriffs: 17. August 2022].

Vainio, Annukka (2013). Beyond research ethics: Anonymity as "ontology", "analysis" and "independence". *Qualitative Research*, 13(6), 685-698.

Watteler, Oliver & Kinder-Kurlanda, Katharina E. (2015). Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung. *Datenschutz und Datensicherheit*, 8, 515-519.

Werner, Cosima (2023/im Druck). *Convenience stores as social spaces—Trust and relations in deprived neighborhoods in the U.S.* Lanham, MD: Lexington Books.

Wiest, Karin (2016). Migration and everyday discourses: Peripheralisation in rural Saxony-Anhalt from a gender perspective. *Journal of Rural Studies*, 43, 280-290.

Wiles, Rose; Crow, Graham; Heath, Sue & Charles, Vikki (2008). The management of confidentiality and anonymity in social research. *International Journal of Social Research Methodology*, 11(5), 417-428.

ZENDAS (Zentrale Datenschutzstelle der baden-württembergischen Universitäten) (2016). *Der Begriff Anonymität bzw. Personenbezug: Relatives und absolutes Begriffsverständnis*, https://www.zendas.de/themen/anonymisierung/begriffsverstaendnis_anonymitaet.html [Datum des Zugriffs: 28. März 2023].

ZENDAS (Zentrale Datenschutzstelle der baden-württembergischen Universitäten) (2018a). *Einwilligung*, <https://www.zendas.de/themen/datenschutz-grundverordnung/einwilligung.html> [Datum des Zugriffs: 28. März 2023].

ZENDAS (Zentrale Datenschutzstelle der baden-württembergischen Universitäten) (2018b). *Löschen der Daten nach Durchführung wissenschaftlicher Umfragen*, https://www.zendas.de/themen/umfragen/umfragen_loeschung.html [Datum des Zugriffs: 31. März 2023].

Zu den Autorinnen und zum Autor

Cosima WERNER ist seit 2021 Mitglied der Arbeitsgruppe Kulturgeographie an der Universität Kiel. Vor ihrer aktuellen Tätigkeit hat sie ihre Promotion am Geographischen Institut und am Heidelberg Center for American Studies an der Ruprecht-Karls-Universität Heidelberg abgeschlossen. Davor absolvierte sie ihr Studium in Kulturgeographie an der Universität Erlangen und an der Sozialwissenschaftlichen Fakultät sowie am Geographischen Institut der Universität Göttingen. In ihrer Forschung setzt sie sich mit dem Verständnis von "Whiteness" im deutschen Kontext anhand von Hobbies auseinander. Des Weiteren beschäftigt sie sich mit stadtgeografischen Ansätzen und sozialen Ungleichheiten. Ihre praxistheoretischen Zugänge ergänzen das Ganze und bieten ihr eine theoretische Brille auf komplexe soziale Phänomene.

Kontakt:

Cosima Werner

Christian-Albrechts Universität zu Kiel,
Geographisches Institut
Ludewig-Meyn-Str. 8, 24118 Kiel

E-Mail: werner@geographie.uni-kiel.de
URL: <https://www.kulturgeo.uni-kiel.de/de/team/cosima-werner>,
<https://orcid.org/0000-0002-0324-8971>

Frank MEYER ist seit 2018 wissenschaftlicher Mitarbeiter an der TU Dresden. Zwischen 2010 und 2019 war er wissenschaftlicher Mitarbeiter am Leibniz-Institut für Länderkunde in Leipzig. Nach seinem Diplomabschluss der Geografie an der Universität Leipzig promovierte er im Jahr 2019 zu kirchlichen Restrukturierungsbemühungen in ländlichen Regionen. Seine Arbeitsschwerpunkte sind Religionsgeografie (als einer der Sprecher des AK Religionsgeographie), Globalisierungsforschung, Geografien ländlicher Regionen, politische Geografie, Migrationsforschung und die theoretischen und methodologischen Grundlagen geografischer Forschung. Er leitet zudem ein DFG-Netzwerk von Geograf:innen zur Visualisierung qualitativer Forschung (zusammen mit Kristine BEURSKENS und Francis HARVEY).

Kontakt:

Frank Meyer

TU Dresden, Institut für Geographie, Professur für Humangeographie
Helmholtzstr. 10, 01069 Dresden

E-Mail: frank.meyer1@tu-dresden.de
URL: <https://tu-dresden.de/bu/umwelt/geo/geographie/humangeo/die-professur/team/meyer>,
<https://orcid.org/0000-0001-9819-1914>

Susann BISCHOF ist wissenschaftliche Mitarbeiterin am Thünen-Institut für Lebensverhältnisse in ländlichen Räumen in Braunschweig und Promotionsstudentin an der Sozialwissenschaftlichen Fakultät der Universität Göttingen. Zuvor war sie wissenschaftliche Referentin am Institut für Demokratie und Zivilgesellschaft in Jena. Sie studierte Soziologie, Philosophie und Gesellschaftstheorie an der Universität Jena. Ihre Forschungsschwerpunkte sind ländliche Räume, Alltagsvorstellungen von Politik, soziale und räumliche Ungleichheiten und qualitative Sozialforschung.

Kontakt:

Susann Bischof

Johann-Heinrich von Thünen-Institut
Thünen-Institut für Lebensverhältnisse in ländlichen Räumen
Bundesallee 64, 38118 Braunschweig

E-Mail: susann.bischof@thuenen.de
URL: <https://orcid.org/0000-0003-0120-506X>

Zitation

Werner, Cosima; Meyer, Frank & Bischof, Susann (2023). Grundlagen, Strategien und Techniken der Anonymisierung von Transkripten in der qualitativen Forschung: eine praxisorientierte Einführung [78 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 24(3), Art. 12, <https://doi.org/10.17169/fqs-24.3.4067>.