

Kati Mozygamba & Betina Hollstein

In Zusammenarbeit mit: Jan-Ocko Heuer, Elisabeth Huber, Hai Ha Nguyen

**Anonymisierung und Pseudonymisierung
qualitativer textbasierter Forschungsdaten –
Eine Handreichung**

University of Bremen – Qualiservice Working Papers

QS-WP-5-2023

University of Bremen – Qualiservice Working Papers

Empfohlene Zitierung / Suggested Citation:

Mozygamba, Kati & Betina Hollstein (2023): Anonymisierung und Pseudonymisierung qualitativer textbasierter Forschungsdaten – eine Handreichung. Qualiservice Working Papers 5-2023, Bremen, <https://doi.org/10.26092/elib/2525>

© FDZ Qualiservice, Oktober 2023

Impressum

Forschungsdatenzentrum (FDZ) Qualiservice
Universität Bremen
SOCIUM – Forschungszentrum Ungleichheit und Sozialpolitik
Mary-Somerville-Str. 7
UNICOM - Gebäude
D-28359 Bremen
Germany



Webseite: <https://www.qualiservice.org>

E-Mail: info@qualiservice.org

Inhalt

1 Einleitung	4
1.1 Hintergrund	4
1.2 Zielgruppe und Anwendungsbereiche der Handreichung	6
1.3 Begriffsklärung	6
2 Forschungsethik, Datenschutz und Forschungsinteresse abwägen	8
2.1 Datenschutzaspekten nachkommen.....	8
2.2 Forschungsethische Aspekte berücksichtigen.....	9
2.3 Forschung im Spannungsfeld zwischen Datenschutz, Forschungsethik, Forschungsinteresse und guter wissenschaftlicher Praxis.....	11
3 Das Anonymisierungskonzept des FDZ Qualiservice	11
3.1 Die informierte Einwilligung als Grundlage für die Datenaufbereitung	11
3.2 Zweck der Anonymisierung	11
3.3 Anonymisierung als Bestandteil des Forschungsprozesses und kooperative Aufgabe	12
3.4 Die Strategie der „flexiblen Anonymisierung“	13
3.5 Löschen personenbezogener Informationen	13
3.6 Listenverwendung und Teilstandardisierung	14
3.7 Weitere Bestandteile des Datenschutzkonzeptes von Qualiservice.....	15
4 Personenbezogene Informationen Schritt für Schritt ersetzen	15
5 Zentrale Kategorien für die Ersetzung personenbezogener Information	21
5.1 Ersetzung von Namen.....	24
5.2 Ersetzung von Zeitangaben	26
5.3 Ersetzung von Ortsangaben.....	27
5.4 Ersetzung von Informationen zu Ausbildung und Beruf	32
5.5 Ersetzung von Organisationen und Institutionen	35
5.6 Informationen zu Wirtschaftszweigen	36
5.7 Ersetzung von Gesundheitsinformationen	37
5.8 Ersetzung von Positionen und Rollen.....	39
5.9 Weitere Informationen, die einen Personenbezug enthalten können	40
5.10 Kumulation von personenbezogenen Informationen verschiedener Kategorien	41
6 QualiAnon - Das Qualiservice Anonymisierungstool	42
7 Zusammenfassung und Ausblick	46
8 Quellenverzeichnis	48
9 Anhang: Beispiele	51

1 Einleitung

1.1 Hintergrund

Bei der Anonymisierung werden Informationen über Forschungsteilnehmer/-innen, Forscher/-innen und etwaige Dritte, die in den Materialien enthalten sind, so verändert, dass ein Personenbezug nicht mehr möglich ist und die Beteiligten nicht mehr (oder nur mit einem unverhältnismäßig großen Aufwand an Ressourcen – faktische Anonymisierung) identifiziert werden können. Sie dient damit dem Schutz von Persönlichkeitsrechten beteiligter Akteure. Die rechtliche Basis dafür, Forschungsdaten zu anonymisieren, sobald es der Forschungszweck erlaubt, resultiert aus der europäischen Datenschutzgrundverordnung (DSGVO) (Europäische Union (EU), 2018).¹ Sind die Daten anonymisiert, fallen sie nicht mehr in den Anwendungsbereich der DSGVO. Darüber hinaus ist die Anonymisierung als Schutzstrategie Bestandteil forschungsethischer Anforderungen an die gute wissenschaftliche Praxis (RatSWD 2017b) und in verschiedene fachspezifische Ethikkodizes sowie Positionspapiere zum Umgang mit Forschungsdaten eingegangen (z.B. Deutsche Forschungsgemeinschaft (DFG), 2019; Deutsche Gesellschaft für Volkskunde (DGV), 2018; Deutsche Gesellschaft für Soziologie (DGS), 2017; Deutsche Gesellschaft für Erziehungswissenschaft (DGfE), 2005, 2016).

Gleichzeitig geht mit der Anonymisierung von Forschungsmaterialien immer auch ein Informationsverlust einher. Es gilt deshalb Strategien zu entwickeln, die den Schutzrechten von Forscher/-innen und Beforschten gerecht werden und gleichzeitig den wissenschaftlichen Nachnutzungswert erhalten. Der Zweck, für den anonymisiert wird, ist dabei für die Entwicklung eines passenden Anonymisierungskonzeptes von besonderer Bedeutung (Medjedovic, 2011; Opitz & Mauer, 2005) (vgl. Punkt 3.2).

Die wissenschaftliche Nachnutzung von Daten aus der qualitativen Sozialforschung stellt darüber hinaus besondere Ansprüche an eine solche Strategieentwicklung. Am Beispiel von biographischen Erzählungen oder ethnographischen Feldbeobachtungen kann leicht nachvollzogen werden, dass qualitative Forschung mit dem Ziel, untersuchte Phänomene in ihrer Komplexität zu verstehen, auf ein besonders tiefes Eindringen und mögliches Verständnis des untersuchten Phänomens zielt (von Unger, 2014). Des Weiteren richtet sich das Erkenntnisinteresse in den meisten Fällen gerade auf sensible Themenbereiche, die zudem aus der subjektiven Sicht von Befragten rekonstruiert werden (Kretzer, 2013).² Begründet in der Offenheit des Forschungsprozesses, der hohen Informationsdichte sowie der

¹ Sind die Daten anonymisiert, fallen sie nicht mehr in den Anwendungsbereich der DSGVO. Solange die Zuordnung von Ersetzungen zu Originalinformationen noch möglich ist, spricht man von pseudonymisierten Daten. Pseudonymisierte Daten sind personenbezogene Daten (vgl. Punkt 1.3).

² Bei der Erhebung personenbezogener Daten sollte immer reflektiert werden, ob bei der Durchführung einer Studie den Beteiligten Anonymität tatsächlich versprochen werden kann oder ob nicht auch die Grenzen des Vorgehens verdeutlicht werden sollten Kelly (2009). Eine Überlegung, die z.B. auch Stein, A. (2010) reflektiert – eine Forscherin, die selbst Erfahrungen mit der Re-Identifikation von Forschungsteilnehmer/-innen gemacht hat.

Kontextgebundenheit der Daten findet sich in qualitativen Forschungsmaterialien in der Regel eine Vielzahl miteinander verwobener personenbezogener Informationen.

Sollen qualitative Daten wissenschaftlich nachgenutzt werden, muss der sozialwissenschaftlich relevante (Entstehungs-)Kontext von zu anonymisierenden Äußerungen so weit wie möglich erhalten bleiben, um die im Material repräsentierten Sinnzusammenhänge für eine wissenschaftliche Nachnutzung zu erhalten. Entsprechende Aufbereitungsstrategien sollten auf verschiedenen Ebenen ansetzen. Bestimmte Vorgehensweisen wie die Abstraktion von Informationen eignen sich dabei mehr für das Ziel der Sekundäranalyse als z.B. das Löschen von Informationen (Medjedovic, 2014).³ Die Anforderungen an die Anonymisierung qualitativer Forschungsmaterialien variieren dabei von Studie zu Studie, wenn z.B. in einer Studie politisch brisante Themen verhandelt werden; oder auch von Datensatz zu Datensatz innerhalb einer Studie, wenn z.B. begründet in der Verwendung verschiedener Interviewformen unterschiedlich komplexe Daten gewonnen werden. Das Anonymisierungskonzept muss für jede Studie entwickelt bzw. angepasst werden; ein Vorgehen im Sinne von "one size fits all" (Stam & Kleiner, 2020) wäre mit einer (unnötigen) Reduktion des Nachnutzungswertes und/oder einer etwaigen Erhöhung des Re-Identifikationsrisikos verbunden.

Um ein gutes datenspezifisches Schutzkonzept mit Blick auf die Nachnutzung qualitativer Daten zu entwickeln, gilt es Fragen des Vorgehens und der Tiefe von Abstraktionen ebenso wie Fragen der Notwendigkeit zusätzlicher organisatorisch-technischer Schutzmaßnahmen zu berücksichtigen. Ein Beispiel hierfür ist die Einschränkung der Nachnutzung auf besonders geschützte Arbeitsplätze im Forschungsdatenzentrum (FDZ) vor Ort (onsite-Nutzung). Neben technisch-organisatorischen Maßnahmen kann hier auch eine Option sein, die Forschungsteilnehmer/-innen erneut zu kontaktieren und in die Abwägungs- und Entscheidungsprozesse mit einzubeziehen (von Unger, 2014).

Bisher finden qualitativ Forschende nur wenig praktische Anleitung, um ein für ihr Vorhaben adäquates Anonymisierungsprozedere zu entwickeln. Informationen zum Vorgehen finden sich zum Teil in Texten, die sich eher generisch mit den besonderen Schutzmaßnahmen und Bedarfen von qualitativen Daten auseinandersetzen (z.B. Bishop, 2005; Corti, Day, & Backhouse, 2000). Manche Infrastruktureinrichtungen bieten auch kurze Hinweispapiere oder Beispiele für die Anonymisierung qualitativer Forschungsdaten an wie (z.B. das Finnish Social Science Data Archive (2021) und das FDZ Bildung (Meyermann & Porzelt, 2014) oder der UK Data Service (2021). Darüber hinaus finden sich Berichte von Forschern/-innen, die ihre Erfahrungen mit der Anonymisierung qualitativer Forschungsdaten reflektieren (z.B. Richter et al., 2021; Laudel & Bielick, 2019; Saunders et al., 2015a, 2015b; Thomson et al., 2005). Dennoch – das zeigen u.a. die Erfahrungen aus der Beratungspraxis beim FDZ Qualiservice

³ Ist eine Bereitstellung für die wissenschaftliche Sekundärnutzung von Forschungsdaten anvisiert, sollten Fragen der Pseudonymisierung und Anonymisierung von Beginn an mitgedacht werden, um z.B. zeitliche und personelle Ressourcen zu planen, datenschutzrechtlichen und forschungsethischen Erfordernissen zu entsprechen sowie das Nachnutzungspotenzial eines Datensatzes so weit wie möglich zu erhalten.

– gibt es viele offene Fragen bzgl. der Anonymisierung qualitativer Forschungsmaterialien, die Forscher/-innen Unterstützung suchen lässt.

Mit der vorliegenden Handreichung möchten wir Antworten auf diese Fragen geben und Vorschläge machen, wie qualitative textgebundene Forschungsmaterialien so aufbereitet - d.h., anonymisiert oder pseudonymisiert – werden können, dass einerseits die Persönlichkeitsrechte der Beteiligten gewahrt werden und andererseits die Daten wissenschaftlich (nach-)genutzt werden können. Wir ziehen dabei ethische und datenschutzrechtliche Aspekte in Betracht und präsentieren praktische Anleitungen und Anwendungsbeispiele für die Anonymisierung und/ oder Pseudonymisierung qualitativer textgebundener Forschungsmaterialien. Wir zeigen verschiedene Wege der Ersetzung personenbezogener Informationen⁴ und werben für ein sensibles Vorgehen, welches dazu anhält, den Nachnutzungswert eines Datensatzes möglichst groß zu erhalten und einer Unter- sowie einer Überanonymisierung vorzubeugen. Für die effiziente Umsetzung hat Qualiservice das Anonymisierungstool QualiAnon für textgebundene Forschungsdaten (Nicolai et al., 2021) zusammen mit Forscher/-innen für die Anonymisierung und Pseudonymisierung textgebundener Forschungsmaterialien entwickelt, welches kostenfrei verwendet werden kann (vgl. Punkt 6). Im Anhang finden sich verschiedene Anwendungsbeispiele.

1.2 Zielgruppe und Anwendungsbereiche der Handreichung

Die Handreichung richtet sich an Forschende, die qualitative textgebundene Daten und Forschungsmaterialien anonymisieren/ pseudonymisieren und für die Archivierung und weitere wissenschaftliche Nutzungen (Data Sharing) bereitstellen möchten. Im Zentrum stehen konzeptionelle Überlegungen, die vordergründig an qualitativen Interviewstudien erprobt wurden, aber sich auch auf andere textbasierte Forschungsdaten wie Transkripte von Fokusgruppen oder Beobachtungsprotokolle anwenden lassen.⁵

1.3 Begriffsklärung

Für die vorliegende Handreichung zentral sind die Begriffe „personenbezogene Daten“, „Pseudonymisierung“ und „Anonymisierung“. Diese werden im Folgenden kurz erläutert.

Personenbezogene Daten: In Artikel 4 Abs. 1 DSGVO werden Informationen als „personenbezogene Daten“ bezeichnet, „die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen“; Zu diesen Informationen gehören z.B. Namen,

⁴ Wir sprechen in dieser Handreichung von der „Ersetzung personenbezogener Informationen“ und beziehen uns damit sowohl auf die Anonymisierung von Forschungsdaten, nach der die Re-Identifikation einer Person nicht bzw. nur mit unverhältnismäßig großem Aufwand möglich ist sowie auf die Pseudonymisierung von Informationen, in der die Originalinformation und die zugehörige Ersetzung dokumentiert sind und zur Re-Identifikation von Personen -z.B. für Längsschnittstudien – herangezogen werden können.

⁵ Die Handreichung zur Anonymisierung qualitativer Forschungsdaten fügt sich in einen Kanon von weiteren Handreichungen und Mustervorlagen ein, die Qualiservice kostenfrei über seine Website zur Verfügung stellt (<https://www.qualiservice.org/de/helpdesk.html#downloads>). Dazu gehören: Eine Handreichung zur Kontextualisierung von qualitativen Forschungsdaten (Heuer et al., 2020) sowie eine Handreichung und rechtlich geprüfte Mustervorlagen für das Einholen der informierten Einwilligung für Primärforschung, Archivierung und Sekundärnutzung (Kretzer et al., 2020).

Kenntnissen oder Standortdaten ebenso wie „besondere Merkmale, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person (...) sind“ (ebd.).

Besondere Kategorien personenbezogener Daten: Die DSGVO stellt in Art. 9 besondere Kategorien von personenbezogenen Daten heraus, die besonders geschützt sind. Dazu zählen die „rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen, die Gewerkschaftszugehörigkeit, genetische Daten, biometrische Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung“. Anders als für personenbezogene Daten generell müssen Personen in die Verarbeitung dieser Daten explizit einwilligen, d.h. die in Art. 9 DSGVO aufgelisteten besonderen Kategorien müssen, wenn sie verarbeitet werden sollen, in der Einwilligungserklärung gelistet sein (Erlaubnisvorbehalt).

Pseudonymisierung: Art. 4 Abs. 5 DSGVO definiert Pseudonymisierung als „die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden.“ Pseudonymisierte Daten sind weiterhin personenbezogen und fallen in den Anwendungsbereich der DSGVO. Inwieweit personenbezogene Daten erhalten bleiben können, bestimmt sich in der Regel nach den Vereinbarungen, die in der Einwilligung zur Forschungsteilnahme von Forschenden und Forschungsteilnehmern/-innen getroffen wurden.⁶ In Einzelfällen kann auch eine Interessensabwägung nach Art. 6 Abs. 1 lit. f DSGVO über die Zulässigkeit einer Verarbeitung personenbezogener Daten zu Forschungszwecken ohne Einwilligung herangezogen werden. Dies setzt eine entsprechende Interessensabwägung voraus.

Anonymisierung: Anonymisierung meint die „Tilgung des Personenbezuges“ (RatSWD, 2020, S. 18). Die DSGVO bietet keine eigene Begriffsbestimmung von Anonymisierung an; sie bezieht sich aber auf das Konzept der Anonymisierung (vgl. ebd.). Erwägungsgrund 26 der DSGVO macht deutlich, dass die DSGVO keine Anwendung auf anonyme Daten findet. Weitere Ausführungen zur Anonymisierung finden sich in der DSGVO nicht (vgl. ebd.).

Das frühere Bundesdatenschutzgesetz unterschied drei Varianten der Anonymisierung, die unterschiedliche Anonymisierungsgrade beschreiben. Auch wenn die gesetzlichen Regelungen nunmehr ausschließlich von Anonymisierung sprechen, wird sich in der Praxis weiterhin an der differenzierteren Betrachtung dieser drei Anonymisierungsgrade orientiert. Bei der *formalen Anonymisierung* werden die direkten Identifikatoren wie Namen und Adressen

⁶ Hinweise und juristisch geprüfte Mustervorlagen zur informierten Einwilligung in Forschung, Archivierung und Nachnutzung qualitativer Forschungsdaten bietet Qualiservice in der Handreichung für das Einholen der informierten Einwilligung für Primärforschung, Archivierung und Sekundärnutzung (Kretzer et al. 2020).

verändert. Mit Blick auf qualitative Daten muss nach einer formalen Anonymisierung i.d.R. weiterhin von personenbezogenen Daten gesprochen werden. Die *absolute Anonymisierung*, die auch unter Einsatz von Ressourcen eine Re-Identifikation ausschließt, arbeitet mit Paraphrasierung, Maskierung oder Löschung und nimmt eine starke Reduktion des wissenschaftlichen Nachnutzungspotenzials in Kauf (Kretzer, 2013). Bei der *faktischen Anonymisierung* (Anonymisierung im Sinne des § 3 Abs. 6 BDSG alte Fassung) werden personenbezogene Daten derart verändert, „dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbarer natürlichen Person zugeordnet werden können“. Die faktische Anonymisierung bietet die Möglichkeit, Datenschutz und Nachnutzungspotenzial auszubalancieren. Sie ist die für die Anonymisierung qualitativer Daten i.d.R. gewählte Form der Anonymisierung.

Der RatSWD führt in der Handreichung zum Datenschutz aus: „Das Datenschutzrecht enthält besondere Anonymisierungsgebote für die Forschung. Art. 89 Abs. 1 Satz 4 DSGVO verlangt, die Weiterverarbeitung von Daten zu Forschungszwecken, wenn möglich in einer Form durchzuführen, bei der die Identifizierung von betroffenen Personen nicht oder nicht mehr möglich ist. Daraus ist ein Gebot abzuleiten, Daten bei ihrer Verarbeitung zu Forschungszwecken nach Möglichkeit zu pseudonymisieren oder zu anonymisieren (Golla 2019: 658 f.). Ein besonderes Gebot der Anonymisierung enthält auch § 27 Abs. 3 Satz 1 BDSG für besondere Kategorien personenbezogener Daten (...), die zu Forschungszwecken verarbeitet werden“ (RatSWD 2020, S. 18f.). Dem RatSWD folgend, gelten Daten als anonym, wenn sie faktisch anonymisiert sind, d.h., „wenn ein Personenbezug nur mit unverhältnismäßig hohem Aufwand wiederhergestellt werden kann“ (ebd.).

Die Anonymisierung von Daten stellt selbst eine Verarbeitung personenbezogener Daten dar und bedarf somit einer Rechtsgrundlage.⁷ Diese können sich z.B. aus den Erlaubnistatbeständen des Artikel 6 DSGVO ergeben. In diesem Zusammenhang ist aber auch der Art. 5 Abs. 1 DSGVO wichtig, der Ausnahmen für z.B. die wissenschaftliche Forschung definiert.

2 Forschungsethik, Datenschutz und Forschungsinteresse abwägen

2.1 Datenschutzaspekten nachkommen

Im Datenschutz geht es um den Schutz von Persönlichkeitsrechten und hier insbesondere um Schutzaspekte, die sich aus dem Recht auf informationelle Selbstbestimmung – also dem Recht jedes Einzelnen, über die eigenen personenbezogenen Daten und deren Nutzung zu bestimmen – ableiten. Die rechtliche Grundlage für den Schutz dieser Daten bildet die DSGVO. Sie trat am 25. Mai 2018 in Kraft und ist als Verordnung in allen Mitgliedsstaaten der Europäischen Union unmittelbar bindend, d.h., sie muss nicht erst in nationales Recht

⁷ MLS Legal (2020) Rechtsfragen zur Nutzung von Daten in Forschung, Archivierung und Sekundärnutzung. <https://youtube/vC8ucG9MJSw> (Zugriff: 01.09.2023)

umgesetzt werden. Nationale Regelungen wie z.B. das 2018 neuformulierte Bundesdatenschutzgesetz (BDS neu) – und ggf. subnationale Regelungen wie die Datenschutzgesetze der Bundesländer treten nur ergänzend zur DSGVO hinzu, sofern in der DSGVO entsprechende Öffnungsklauseln vorgesehen sind.

Die DSGVO ist die verbindliche Rechtsgrundlage für Forschende aus den EU-Ländern. Ihr folgend, bedarf jeder Verarbeitungsschritt personenbezogener Daten einer Rechtsgrundlage – unabhängig davon, ob es sich um die Erhebung, Speicherung, Auswertung oder Aufbereitung personenbezogener Daten für die wissenschaftliche Nachnutzung handelt. Die Rechtmäßigkeit der Verarbeitung ist in Art. 6 DSGVO geregelt. Hier steht u.a., dass die personenbezogenen Daten verarbeitet werden dürfen, wenn die Person z.B. in die Verarbeitung eingewilligt hat.

Die Verarbeitung personenbezogener Daten für die wissenschaftliche Forschung wird in der DSGVO besonders privilegiert. Dies zeigt sich an mehreren Stellen (RatSWD 2020, S. 8), wie an den Ausnahmen von den Grundsätzen der Zweckbestimmtheit und Zweckbindung der Verarbeitung in Art. 5 oder der Möglichkeit besondere Kategorien von personenbezogenen Daten für Forschungszwecke zu verarbeiten (RatSWD 2020, 8).⁸

Der Zeitpunkt der Anonymisierung wird in der DSGVO vom Zweck bestimmt, für den die Informationen verarbeitet werden (Art. 5 DSGVO). Solange die Verarbeitung personenbezogener Daten notwendig ist, um den Verarbeitungszweck (z.B. das Forschungsziel) zu erreichen, müssen sie nicht anonymisiert, aber – z.B. durch Pseudonymisierung und/ oder Verschlüsselung – besonders geschützt werden.

2.2 Forschungsethische Aspekte berücksichtigen

Forschungsethik umfasst die ethischen Prinzipien und Regeln, „in denen mehr oder minder verbindlich und mehr oder minder konsensuell bestimmt wird, in welcher Weise die Beziehungen zwischen den Forschenden auf der einen Seite und den in sozialwissenschaftliche Untersuchungen einbezogenen Personen auf der anderen Seite zu gestalten sind (Hopf, 2004, S. 589-590)“.⁹ Forschungsethische Aspekte beinhalten die Reflektion und Aushandlung der Werte forschenden Vorgehens/ Ergründens, die dem Forschungshandeln als sozialem Handeln innewohnen. „Eine forschungsethische Perspektive besteht darin, ‚kritisch zu reflektieren, inwiefern bestimmte ethische Grundsätze für das

⁸ Der RatSWD führt hierzu weiter aus: „Die wissenschaftliche Forschung ist ein berechtigtes Interesse in diesem Sinne, das gegen die Interessen der Betroffenen abzuwägen ist. Auf der Seite der Betroffenen ist unter anderem zu berücksichtigen wie sensibel die Daten sind (...), ob sie aus allgemein zugänglichen oder anderen Quellen stammen und ob der Betroffene mit einer (weiteren) Verarbeitung der Daten rechnen konnte. Es ist auch die Beziehung zwischen dem forschenden Datenverarbeitenden und dem Betroffenen zu berücksichtigen. Bei den Forschungsinteressen kommt es darauf an, welche Bedeutung die Verarbeitung spezifischer personenbezogener Daten für die Ausführung eines Forschungsvorhabens hat“ (RatSWD, 2020).

⁹ Eine Orientierung für in sozialwissenschaftlichen Studien relevanten forschungsethischen Aspekten bietet das vom RatSWD veröffentlichte Working Paper Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften (RatSWD 2017a), aber auch die Arbeiten von Hella von Unger (2014, 2020) bieten hilfreiche Hinweise und Informationen für einen Einstieg in die Thematik. Darüber hinaus bietet der RatSWD eine Best-Practice-Sammlung zum Thema an: <https://www.konsortswd.de/ratswd/best-practice-forschungsethik/> (Zugriff: 01.08.2023)

Forschungshandeln gelten und in der Praxis realisiert werden [können]‘ (von Unger et al. 2014: 2)“.

Verschiedene Fachgesellschaften haben Ethikkodizes verabschiedet, die für ihr Handeln und das ihrer Mitglieder handlungsleitend sind¹⁰ und auch Forschungsförderer wie die Deutsche Forschungsgemeinschaft (DFG) gehen in Leitlinien wie denen zu einer guten wissenschaftlichen Praxis auf ethische Aspekte forschenden Tuns ein (DFG 2022). Die Wahrung der Anonymität von an der Forschung beteiligten Akteuren findet sich in diesen Kodizes als ein forschungsethisches Prinzip neben anderen.

Dabei können forschungsethische Aspekte den datenschutzrechtlichen Schutzvorschriften für forschersches Tun entsprechen. Sie können aber auch darüber hinausgehen, wie das Vorgehen in einem konkreten Beispiel eines Projektes, welches qualitative Daten über Qualiservice zur Nachnutzung zur Verfügung stellen lässt, zeigt: In diesem Fall hatten sich Forscher/-innen entschieden, ihre Interviewdaten zu archivieren und für die Nachnutzung bereit zu stellen. Die Forschungsteilnehmer/-innen hatten in die Archivierung und Weiterverarbeitung für die wissenschaftliche Nachnutzung eingewilligt. Damit wäre datenschutzrechtlichen Verpflichtungen genüge getan gewesen. Das Forscherteam entschied sich aber, die Einwilligung zur Weitergabe der personenbezogenen Daten zur Archivierung und Nachnutzung unter Vorlage des fertig für die Nachnutzung aufbereiteten Interviewtranskriptes noch einmal zu validieren und legte diese den Teilnehmer/-innen erneut zur Bestätigung Ihrer Entscheidung vor, bevor die Daten an Qualiservice übergeben wurden. Darüber hinaus gelangte das Forschungsteam, im Verlauf ihres Forschungsprojektes zu der Entscheidung, die nicht anonymisierten Interviewtranskripte aus forschungsethischen Überlegungen heraus durch weitere technisch-organisatorische Maßnahmen schützen zu wollen und entschied, sie ausschließlich im Gastwissenschaftlerraum vor Ort zugänglich zu machen – auch wenn rein datenschutzrechtlich betrachtet in diesem Fall eine nichtanonymisierte Weiternutzung als Scientific-Use-File (SUF) zum externen Download möglich wäre.

Die Auseinandersetzung mit forschungsethischen Aspekten ist immer ein studien- und datensatzbezogener Prozess. Die Abwägung für mögliche Schutzmaßnahmen betreffen die methodische Anlage einer Studie ebenso wie deren Durchführung, Publikation und Aufbereitung für die Nachnutzung. Die Entscheidung aus forschungsethischen Gründen die Nachnutzung auf einen stark geschützten und dadurch (evtl. auch nur für einzelne Interviews eines Datensatzes) auch stärker reglementierten Datenzugang im Gastwissenschaftlerraum vor Ort zu beschränken, um die Teilnehmer/-innen zu schützen, könnte in einem anderen Fall als problematisch oder auch paternalistisch angesehen werden - z.B. in einer Studie mit Zeitzeugeninterviews in der Oral-History-Forschung (Thomson et al., 2005) oder in der partizipativen Forschung (von Unger, 2014), wo Forschungsteilnehmer/-innen unter Umständen auch genannt werden wollen.

¹⁰ Z.B. Deutsche Gesellschaft für Soziologie (DGS) (2017) oder Deutsche Gesellschaft für Erziehungswissenschaft (DGfE) (2005, 2016).

2.3 Forschung im Spannungsfeld zwischen Datenschutz, Forschungsethik, Forschungsinteresse und guter wissenschaftlicher Praxis

Datenschutz und Forschungsethik liefern einen Handlungsrahmen, in dem sich wissenschaftliche Forschung bewegt. Sollen Daten für wissenschaftliche Zwecke nachgenutzt werden, müssen datenschutzrechtliche und forschungsethische Anforderungen erfüllt sein. Zugleich gilt es die wissenschaftliche (Nach-)Nutzung in dem so bestimmten Rahmen so gut es geht zu ermöglichen. Die DSGVO beinhaltet aus diesem Grund für die Forschung besondere Befugnisse im Umgang mit personenbezogenen Daten (vgl. Punkt 2.1). Die Abwägung zwischen Datenschutz, Forschungsethik und Erkenntnisinteresse sollte studien- und ggf. auch fallspezifisch vorgenommen werden. Der Abwägungsprozess erstreckt sich oft über den gesamten Forschungsprozess und schließt auch mögliche Nachnutzungsszenarien mit ein. Mit Blick auf das Anonymisierungsverfahren gilt es demzufolge ein sicheres und dennoch flexibles Vorgehen zu entwickeln, welches für sozialwissenschaftliche Forschung relevante Information erhält, ohne gegen datenschutzrechtliche und ethische Handlungsmaximen zu verstoßen. Die Lösungen, die das FDZ Qualiservice hierfür erarbeitet hat, beschreiben wir im Folgenden.

3 Das Anonymisierungskonzept des FDZ Qualiservice

3.1 Die informierte Einwilligung als Grundlage für die Datenaufbereitung

In der Regel stellt die informierte Einwilligung der Forschungsteilnehmer/-innen bei Qualiservice die Grundlage für die Datenverarbeitung dar (vgl. Fußnote 11). In der Einwilligungserklärung ist geregelt, ob und wenn ja in welcher Form Forschungsdaten/-materialien genutzt und an Dritte weitergegeben werden dürfen – z.B. ob personenbezogene Informationen in den zu archivierenden Materialien erhalten bleiben dürfen, ob sie anonymisiert oder gar gelöscht werden müssen.¹¹

3.2 Zweck der Anonymisierung

Der Grad der notwendigen Abstraktion hängt neben datenschutzrechtlichen und forschungsethischen Aspekten auch davon ab, wofür die anonymisierten Materialien verwendet werden sollen. So stellt die Publikation von Forschungsergebnissen in wissenschaftlichen Fachjournals andere Ansprüche an das Vorgehen als die Aufbereitung für die wissenschaftliche Nachnutzung oder die akademische Lehre: Für die Publikation in Fachjournals werden i.d.R. Interviewausschnitte aus einem Datensatz ausgewählt, um Analyseergebnisse zu veranschaulichen und nachvollziehbar zu machen. Allein die

¹¹ Informationen zu den Vorlagen für die informierte Einwilligung finden Sie auf der Website von Qualiservice www.qualiservice.org oder in der Qualiservice-Handreichung zur informierten Einwilligung (Kretzer et al. 2020). Die Einwilligung der Forschungsteilnehmer/-innen sollte dokumentiert sein und im Idealfall schriftlich vorliegen. Aber auch andere Varianten der Einwilligung sind denkbar. Mit diesen alternativen Optionen beschäftigt sich z.B. das Qualiservice-Working Paper von Huber, E., & Imeri, S. (2021)

Konzentration auf Textausschnitte reduziert dabei ein mögliches Re-Identifikationsrisiko erheblich. Wird ein Datensatz wissenschaftlich nachgenutzt, verwenden Forscher/-innen vollständige Interviewtexte mit vielen kontextuellen Bezügen. Das Re-Identifikationsrisiko ist hier größer und die Anonymisierung wesentlich aufwendiger.

3.3 Anonymisierung als Bestandteil des Forschungsprozesses und kooperative Aufgabe

Um möglichst optimal aufbereitete Forschungsmaterialien zu erhalten, kooperiert Qualiservice eng mit den Datengeber/-innen in der Aufbereitung der Materialien für die Sekundärnutzung (Mozygemba & Kretzer, 2022), berät und stellt Handreichungen, Workshops, Webinare sowie das Anonymisierungstool QualiAnon zur Verfügung. Die Anonymisierung, resp. Pseudonymisierung übernehmen die Forscher/-innen im Projekt. Sie kennen die Daten genau und wissen, wo sich sensible Textbezüge finden. Die im Forschungsprojekt vorbereiteten pseudonymisierten/ anonymisierten Daten werden nach der Übergabe von Qualiservice geprüft. Wenn nötig werden weitere organisatorisch-technische Schutzvorkehrungen getroffen und ggf. die Anonymisierung vervollständigt.

Generell empfiehlt es sich, dass Forscher/-innen für die Aufbereitung von Forschungsmaterialien für die wissenschaftliche Nachnutzung, dass für die Langzeitarchivierung anvisierte FDZ von Beginn an mit einbeziehen. So bietet Qualiservice bereits während des Forschungsprozesses und an den insbesondere für die Archivierung wichtigen Anforderungen wie der Informierten Einwilligung und der Anonymisierung eine umfangreiche Beratung und Unterstützung an. Forscher/-innen realisieren so die Aufbereitung der Daten bereits schrittweise während des Forschungsprozesses. Dieses Vorgehen verhindert auch eine Ballung von Aufgaben am oder nach Projektende, wenn Mitarbeiter/-innen oft schon nicht mehr im Projekt arbeiten oder Publikationen fertig gestellt werden. Umgekehrt kommt die Integration der Datenaufbereitung in den Forschungsprozess oft auch der Qualität und Transparenz der Forschungsdokumentation und so der Forschung selbst zu Gute (vgl. Fielding, 2004). Darüber hinaus kann das FDZ bei der Antragstellung hinsichtlich der notwendigen personellen wie finanziellen Ressourcen für die Aufbereitung der Daten unterstützen (vgl. Punkt 6).

Für die Anonymisierung empfiehlt Qualiservice ein schrittweises Vorgehen, welches zunächst pseudonymisiert und/oder nur direkte Identifikatoren wie Namen und Adressen anonymisiert (formale Anonymisierung) und die Anonymisierung weiterer Informationen auf einen Zeitpunkt legt, an dem es der Forschungszweck erlaubt. Originalinformationen können so in die Auswertung einbezogen werden. Das schrittweise Ersetzen hilft auch, die Notwendigkeit von Ersetzungen noch einmal zu prüfen: Erfahrungsgemäß besteht anfangs eine Tendenz zur Überanonymisierung, die sich legt, wenn Forscher/-innen im Laufe des Projektes bzw. des Anonymisierungsprozesses besser einschätzen können, wo sich in den Daten Risiken zur Re-Identifikation finden. Das standardisierte Ersetzen in der Regel vorher festgelegter Informationen wie z.B. die generelle Ersetzung aller Ortsnamen in einem Interview weicht dann z.B. einer Abwägung, ob die Ortsangabe im jeweiligen Kontext einen Personenbezug enthält

oder nicht. Im Beispiel: Berlin als Geburtsort würde evtl. ersetzt, bei einer Urlaubsreise nach Berlin könnte Berlin erhalten bleiben (vgl. Punkt 5.3).

3.4 Die Strategie der „flexiblen Anonymisierung“

Um den Nachnutzungswert möglichst groß zu halten, empfiehlt Qualiservice die Strategie, sensible Informationen mit sozialwissenschaftlich relevanten Informationen zu ersetzen (Kretzer, 2013). Soll der Beruf des Rettungssanitäters ersetzt werden, böte sich z.B. als Ersetzung „ein Vertreter des Rettungsdienstpersonals“ an und weitere Informationen wie Tätigkeit im ambulanten oder stationären Bereich könnten hilfreiche Informationen für die Nachnutzung erhalten. Ersetzungen werden außerdem auf unterschiedlichen Abstraktionsebenen vorgenommen: Ein „Herzinfarkt“ könnte z.B. zum „akuten Koronarsyndrom“ oder auf einer höheren Abstraktionsebene zu einer „Herz-Kreislauf-Erkrankung“ oder noch abstrakter zu einer nicht näher bestimmten „Erkrankung“ werden (vgl. 5.7).

Die unterschiedlich stark von der Originalinformation abstrahierenden Ersetzungen – evtl. sogar unter Erhalt der Originalinformation (Pseudonymisierung) – sind die Grundlage der von Kretzer (2013) das erste Mal beschriebenen „flexiblen Anonymisierung“. Die flexible Anonymisierung ermöglicht durch das Auf- und Zuklappen von Informationen die Erstellung unterschiedlich stark anonymisierter Scientific-Use-Files für die Nachnutzung. D.h., es können je nach Forschungszweck bestimmte Informationen geöffnet und andere verdeckt werden.

Die Beschreibung von Kretzer (2013) bezieht sich auf die Verwendung pseudonymisierter Daten, bei denen Originalinformationen vorliegen. Sie kann aber darüber hinaus auch Anwendung finden, wenn Daten anonymisiert vorliegen (die Originalinformation also nicht mehr vorhanden ist) und Ersetzungen auf unterschiedlichen stark abstrahierenden Ebenen dokumentiert sind. Liegen Qualiservice wie im Beispiel des Herzinfarktes diese unterschiedlich stark von der Originalinformation abstrahierenden Ersetzungen vor, können diese verwendet werden, um unterschiedliche Versionen eines anonymisierten Datensatzes zu erstellen und somit den Nachnutzungswert für spezifische Forschungsfragen zu erhöhen. Grundlage für die flexible Anonymisierung sind somit unterschiedlich stark abstrahierte Ersetzungen – mit oder ohne Bezug zur Originalinformation (vgl. Schritt 6 auf Seite 18).

3.5 Löschen personenbezogener Informationen

Qualiservice empfiehlt auf das Löschen von Informationen ebenso wie auf das Schwärzen und das Verfälschen von Informationen (aus Berlin wird z.B. Hamburg oder aus der Kosmetikerin die Frisörin) zu verzichten. Müssen Daten gelöscht werden, weil ein etwaiger Personenbezug anders nicht auszuschließen ist, sollte der Löschvorgang im Text kenntlich gemacht werden. Für gelöschte Wörter können z.B. Buchstaben oder Sonderzeichen genutzt werden, die Auskunft über die Anzahl der Wörter geben, die gelöscht wurden. Somit wird zumindest die Information über den Umfang der Löschung, d.h., darüber ob z.B. nur einzelne Wörter oder ganze Absätze fehlen, erhalten.

3.6 Listenverwendung und Teilstandardisierung

Während in der quantitativen Sozialforschung oftmals auf standardisierte Ersetzungen für die Anonymisierung zurückgegriffen wird, sind in der qualitativen Sozialforschung i.d.R. studienspezifische Anonymisierungskonzepte geeignet, um personenbezogene Informationen zu ersetzen. Werden diese Ersetzungsschemata geteilt, können Forscher/-innen eines Themenbereichs diese weiterentwickeln und so peu á peu ausdifferenzierte Ersetzungsschemata mit etwaig sozialwissenschaftlich relevanten Informationen auf unterschiedlichen Abstraktionsebenen für bestimmte Themenbereiche entwickeln. Die bloße Verwendung standardisierter Ersetzungen von Kategorien wie Orten, Namen, Berufen, Zeitangaben etc. führt eher zu gering ausdifferenzierten Ersetzungsschemata, die Information recht stark abstrahiert und eher zu Standardisierung führt. Dennoch kann auch die Verwendung standardisierter Klassifikationssysteme für bestimmte Themenbereiche einen Mehrwert bieten. Neben z.B. Zeiteinteilungen in Monate, Jahre, Jahresabschnitte, Tage etc. können standardisierte Listen wie die International Classification of Occupations (ISCO), die in der quantitativen Forschung zur Codierung von Berufen genutzt wird (z. B. Züll, 2015), für diesen Zweck Verwendung finden. Sind standardisierte Klassifikationen wie die ISCO hierarchisch organisiert, können sie z.B. auch unterschiedliche Abstraktionsniveaus abbilden helfen (vgl. Punkt 5.4). Qualiservice hat gute Erfahrungen mit diesem Vorgehen.¹² Ob die Verwendung einer Liste einen Mehrwert für die Erhaltung sozialwissenschaftlicher Information bei der Abstraktion personenbezogener Informationen bietet, muss von Studie zu Studie entschieden werden.

Vier solcher Klassifikationen, die für die Ersetzung personenbezogener Information eingesetzt werden können, stellen wir im Rahmen der folgenden Ausführungen vor. Die vorgestellten Klassifikationen sind frei zugänglich und werden regelmäßig aktualisiert. Dazu gehören die ISCO (vgl. Punkt 5.4), die International Classification of Diseases (ICD) (vgl. Punkt 5.7), die Klassifikation der Wirtschaftszweige (vgl. Punkt 5.6) und die Gemeindelisten des Bundesinstituts für Bau-, Stadt- und Raumplanung (BBSR) (vgl. 5.3). Die Gemeindeliste des BBSR bietet sogar die Möglichkeit über eine Kennziffer weitere Listen zu verknüpfen und so weitere z.B. ortsspezifische Informationen wie Angaben zur Infrastruktur oder Arbeitslosigkeit in der Region für Ersetzungen zu verwenden. Bis auf die Gemeindelisten des BBSR sind die genannten Listen hierarchisch strukturiert. Wie die Listeninformationen in die Ersetzung aufgenommen werden, ob man z.B. den zugehörigen Listen-Code zu einem Wirtschaftszweig eines Unternehmens nur einmal dokumentiert, um die Lesbarkeit des Transkriptes nicht zu beeinträchtigen oder ob man sich entscheidet, diesen bei jeder Ersetzung mitzuführen, muss zu Beginn der Arbeit entschieden werden und sollte dann für die Studie bzw. den Datensatz konsistent gehandhabt werden.

¹² Forschende können Klassifikationen auch in das Anonymisierungstool QualiAnon einlesen sowie die erstellten (themen-, disziplinspezifischen) Ersetzungskategorien exportieren, wiederverwenden, weiterentwickeln, teilen und so auch als Projektoutput publizieren (vgl. Punkt 6).

3.7 Weitere Bestandteile des Datenschutzkonzeptes von Qualiservice

Bevor wir die einzelnen Schritte beschreiben, die Qualiservice bei der Ersetzung personenbezogener Daten vornimmt, möchten wir darauf hinweisen, dass die Anonymisierung nur ein Element des Datenschutzkonzeptes von Qualiservice neben weiteren technisch-organisatorischen Schutzmaßnahmen ist (vgl. z.B. Heuer, Hollstein, & Mozygemba, 2021). Beispiele für weitere Maßnahmen sind die getrennte Speicherung von Identifikationsdaten und Forschungsmaterialien, die Beschränkung der Nachnutzung auf wissenschaftliche Zwecke, der kontrollierte Zugang zu Forschungsdaten sowie vertragliche Vereinbarungen für die Nutzung von Forschungsdaten. Letztere verbieten u.a. Re-Identifikationsversuche oder die Veröffentlichung ganzer Interviews.¹³ Forscher/-innen können darüber hinaus weitere individuelle Schutzmaßnahmen wie z.B. zeitliche Embargos oder Bedingungen für die Zitation festlegen.

4 Personenbezogene Informationen Schritt für Schritt ersetzen

Abbildung 1 zeigt im Überblick welche Schritte notwendig sind, um die Ersetzung personenbezogener Informationen vorzubereiten, durchzuführen und zu dokumentieren, sodass das Vorgehen von potenziellen Nachnutzer/-innen gut nachvollzogen werden kann und das Material immer noch ausreichend Potenzial für die Nachnutzung bietet.

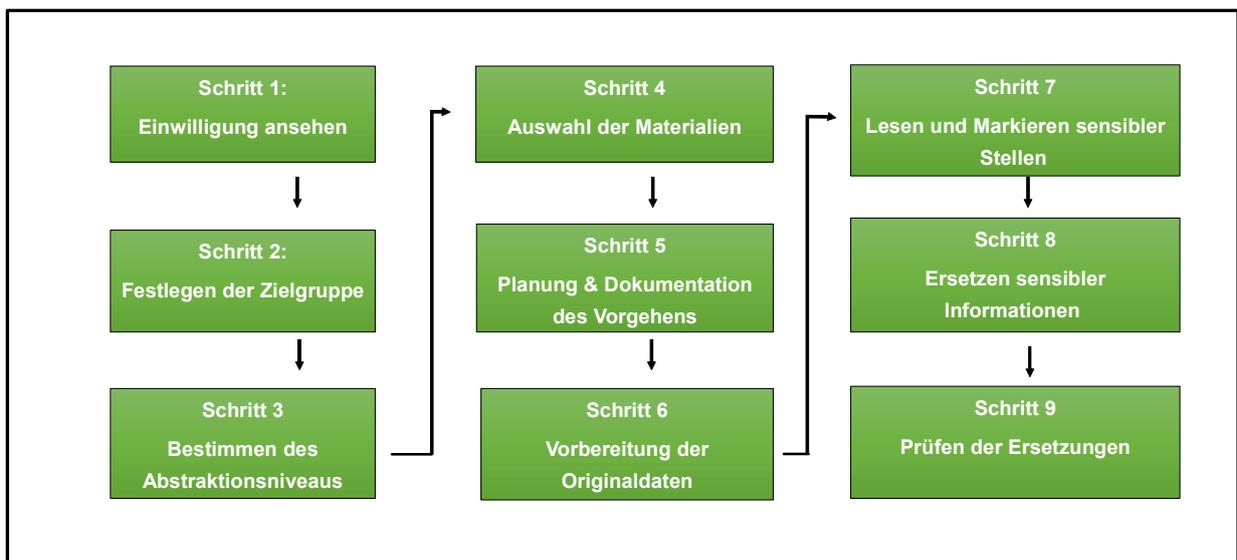


Abbildung 1: Das Vorgehen im Überblick

¹³ Die Nutzungsbedingungen von Qualiservice finden sich unter folgendem Link: https://www.qualiservice.org/files/contao-theme/public/documents/downloads/Vereinbarung_Datennutzung_01_20202_barrierefrei.pdf oder auf Nachfrage bei Qualiservice unter info@qualiservice.org

Schritt 1: Einwilligungserklärung ansehen

In der informierten Einwilligung (Informed Consent) sollten Forscher/-innen festhalten, wie mit den im Forschungsprojekt erhobenen Daten umzugehen ist (vgl. Punkt 3.1). Die Informierte Einwilligung ist die Grundlage der Datenverarbeitung bei Qualiservice.

Schritt 2: Festlegen der Zielgruppe, für die anonymisiert wird

Die Frage danach, für wen anonymisiert wird, wer also potenziell Zugriff auf die anonymisierten Daten haben wird, ist bedeutsam für die Wahl des Abstraktionsniveaus der Ersetzungen. Gute Freunde/-innen, Nachbarn/-innen, Arbeitskollegen/-innen oder Forschende haben in der Regel unterschiedliches Kontextwissen über die beforschten Personen, welches zu einer Re-Identifikation herangezogen werden könnte. Sollen Interviewteile z.B. in einem Buch publiziert werden, welches jede/-r einsehen kann, muss das Abstraktionsniveau weit höher sein, als wenn – wie bei Qualiservice – ausschließlich Wissenschaftler/-innen mit berechtigtem Forschungsinteresse und nach Unterzeichnung einer Nutzungsvereinbarung auf die Daten zugreifen können.

Schritt 3: Bestimmen des Abstraktionsniveaus (zweckgebundene Anonymisierungsstufe)

Bei der Entscheidung für ein bestimmtes Abstraktionsniveau gilt es datenschutzrechtliche und forschungsethische Aspekte sowie das wissenschaftliche Nachnutzungspotenzial gegeneinander abzuwägen (vgl. Punkt 2.3). Die Entscheidung für ein Abstraktionsniveau kann auf Studienebene, aber auch auf der Ebene einzelner Datensatzelemente wie z.B. für einzelne Interviews getroffen werden. In die Entscheidung fließen die Zielgruppe der Nachnutzung, Studienspezifika wie z.B. eine besondere Sensibilität des Themas oder der Darstellung ebenso ein wie technisch/ organisatorische Maßnahmen wie z.B. der besonders gesicherte Zugang über den Gastwissenschaftlerarbeitsplatz (onsite-Nutzung). Die Entscheidung für ein bestimmtes Abstraktionsniveau muss studien- und z.T. fallspezifisch getroffen werden.¹⁴

Schritt 4: Auswahl der Materialien, die anonymisiert/ pseudonymisiert werden sollen

Neben den Forschungsdaten im engeren Sinne kann es auch notwendig sein, weitere Materialien wie Postskripte, soziodemographische Daten und sofern sie zur Kontextualisierung des Datensatzes beigelegt werden auch Projektberichte, Projektveröffentlichungen etc. in die Planung der Anonymisierung/ Pseudonymisierung mit einzubeziehen. Sie können ebenfalls personenbezogene Informationen enthalten, wie z.B. den Namen des/-r Befragten oder der Interviewer/-in im Header des Postskripts oder auch bereits veröffentlichte Informationen wie im Falle der Nennung von für die Fallauswahl relevanten Erhebungsorten in Publikationen oder Projektberichten. Diese Informationen müssen bei der Planung der Anonymisierung berücksichtigt werden.

¹⁴ Die Abwägung kann wie im Fall des von Laudel & Bielick (2019) beschriebenen Projektes auch zu dem Ergebnis führen, dass Daten nicht anonymisierbar sind, wenn ihr Nachnutzungswert erhalten bleiben soll. Liegt die Einwilligung in die Weitergabe personenbezogener Informationen vor, kann in so einem Fall geprüft werden, ob bestimmte technisch-organisatorische Vorkehrungen wie z.B. der Gastwissenschaftsarbeitsplatz vor Ort die Nachnutzung ermöglichen könnten. Evtl. kann auch eine Interessensabwägung nach § 6 Abs. 1 DSGVO angemessen sein. Das gilt es im Einzelfall zu entscheiden.

Schritt 5: Das Vorgehen bei der Anonymisierung/ Pseudonymisierung dokumentieren

Damit Sekundärnutzer/-innen nachvollziehen können, wie das Forschungsmaterial für die Nachnutzung verändert wurde, muss das Vorgehen bei der Anonymisierung/ Pseudonymisierung dokumentiert werden. Gleichzeitig unterstützt die Dokumentation ein konsistentes Vorgehen, wenn über einen längeren Zeitraum gearbeitet wird oder z.B. mehrere Personen die Daten bearbeiten. Die Dokumentation sollte Angaben zu folgenden Punkten enthalten:

- Warum werden die Daten anonymisiert oder pseudonymisiert?
- Werden die Forschungsmaterialien anonymisiert oder pseudonymisiert weitergegeben?
- Bleibt die Pseudonymisierungsliste (vgl. Schritt 6) beim Forschenden oder wird sie mit an das FDZ übergeben?
- Wie wird abstrahiert? Werden z.B. unterschiedliche Abstraktionsniveaus verwendet?
- Wird die Anonymisierung/ Pseudonymisierung vor der Datenanalyse vorgenommen oder wird schrittweise vorgegangen? D.h., werden zunächst die direkten Identifikatoren entfernt (Basisanonymisierung) und erst in einem späteren Schritt (z.B. nach Abschluss der Analyse) auch indirekte Identifikatoren abstrahiert?
- Werden die Teilnehmer/-innen in der Feldsituation darum gebeten, Namen, Orte und weitere evtl. identifizierende Informationen außen vor zu lassen?
- Wird für die Anonymisierung ein Anonymisierungstool verwendet oder wird manuell vorgegangen? Wenn ja, welches Tool wird verwendet?
- Wie werden Ersetzungen im Material kenntlich gemacht? (Verwendung finden hier z.B. Sonderzeichen wie eckige Klammern, Pipes oder farbliche Hervorhebungen. Es sollte darauf geachtet werden, dass diese nicht Teil der verwendeten Transkriptionsgrammatik sind.)
- Wird mit Zeilen- oder Absatznummern gearbeitet? (Zeilennummern des Originalmaterials werden sich durch Ersetzungen verändern, bieten aber eine genauere Zuordnung in der Ersetzungsliste, Absatznummern ändern sich nicht.)
- Werden Ersetzungen fall- oder studienspezifisch vorgenommen? D.h., wurde die Stadt Berlin nur in Interview 1 mit "Ort 1, große Großstadt in Deutschland" ersetzt (fallbezogen) oder wird Berlin in allen Interviews/ Materialien der Studie auf diese Weise maskiert (studienbezogen)?
- Welche Informationen werden ersetzt (Ersetzungskategorien wie Orte, Krankheiten etc. (vgl. Punkt 5.10)? Welche sozialwissenschaftlich relevanten Informationen sollen zu den Ersetzungen erhalten werden?
- Werden (kontrollierte) Listen (vgl. Punkt 5.7) für die Ersetzung verwendet? Wenn ja, welche Listen (bei kontrollierten Listen muss die genaue Zitation mit Versionsdatum dokumentiert werden, da diese Listen regelmäßig von den zuständigen Stellen aktualisiert werden)
- Werden Ersetzungen in bestimmten Ersetzungskategorien gezählt? (Eine Zählung hat Vorteile für den Erhalt des Nachnutzungspotenzials. Dies zeigt die beschriebene

Ersetzung von "Berlin" mit "Ort 1". Ließe man die Zählung weg, wäre nicht bestimmbar, ob im Text von einem oder von verschiedenen Orten die Rede ist. "Ort 2, große Großstadt in Deutschland" dagegen ließe sich von Ort 1 als weitere große Großstadt in Deutschland, z.B. Hamburg abgrenzen.) Es gilt zu dokumentieren, wie die Zählung kenntlich gemacht wird (z.B. Verwendung von Ziffern oder Buchstaben) und in der Verwendung konsistent zu bleiben.

- Werden Informationen gelöscht? Wenn ja, wie werden Löschungen kenntlich gemacht? (vgl. Punkt 3.5).¹⁵
- Haben mehrere Personen die Anonymisierung/ Pseudonymisierung vorgenommen? Wenn ja, wie wurde Konsistenz im gemeinsamen Vorgehen sichergestellt (gab es z.B. einen Testlauf an einem Interviewbeispiel)?

Schritt 6: Erstellen einer Ersetzungstabelle/ Pseudonymisierungsliste

Die Ersetzung personenbezogener Informationen werden in einer Ersetzungstabelle (vgl. Tabelle 1) dokumentiert. Diese sollte auch bereits im Forschungsprojekt an einem sicheren Ort getrennt von den Daten aufbewahrt werden. Die Kopfzeile einer Ersetzungsliste sollte die Interviewnummer, die Dateibezeichnung, die Zeilen- oder Absatznummer der Originalinformation, die Ersetzungskategorien, und die Ersetzungen auf unterschiedlichen Abstraktionsstufen und ggf. den Originaltext sowie Raum für Kommentare als Spaltentitel enthalten. Wenn mit pseudonymisierten Daten gearbeitet werden darf, bleibt auch die Originalinformation erhalten und wird in der Ersetzungstabelle dokumentiert. Für die Anonymisierung der Daten wird die Spalte, die die Originalinformation enthält, entfernt. Unterschiedlich stark abstrahierte Ersetzungen bleiben aber ebenso wie weitere sozialwissenschaftlich relevante Informationen erhalten und können für die flexible Anonymisierung (vgl. 3.4) herangezogen werden.

Schritt 7: Sinnverstehendes Lesen und Markieren sensibler Stellen

Im nächsten Schritt werden die zu bearbeitenden Materialien sinnverstehend gelesen und Informationen, die potenziell einen Rückschluss auf Personen zulassen, markiert. Dabei gilt es etwaige Reidentifikationsrisiken für Dritte (Interviewer, Personen, über die im Interview gesprochen wird) zu berücksichtigen.¹⁶

Beratungen von Forscher/-innen sowie die Kuration von Forschungsmaterialien im FDZ zeigen, dass Forscher/-innen aus Sorge datenschutzrechtliche Auflagen evtl. nicht zu erfüllen, insbesondere zu Beginn des Anonymisierungsprozesses sehr starke Abstraktionen wählen oder sogar Informationen löschen. Es werden dann mehr Informationen ersetzt als aufgrund datenschutzrechtlicher Vorgaben oder aufgrund forschungsethischer Erwägungen nötig ist.

¹⁵ Um das Nachnutzungspotenzial zu erhalten empfiehlt Qualiservice mit Abstraktionen auf unterschiedlichen Ebenen sowie bei der Ersetzung mit sozialwissenschaftlich relevanten Informationen zu arbeiten und Löschungen nur in ganz besonders sensiblen Kontexten vorzunehmen.

¹⁶ Auch wenn Forschungsteilnehmer/-innen in die Verarbeitung und Weitergabe ihrer personenbezogenen Daten eingewilligt haben und diese damit nicht anonymisiert werden müssen, müssen personenbezogene Informationen über Dritte anonymisiert werden.

Ein solches Vorgehen wirkt sich dann deutlich auf den Nachnutzungswert der Materialien aus. Um Überanonymisierung zu vermeiden, empfiehlt es sich deshalb, zunächst als sensibel eingeschätzte Informationen in mehreren Interviews zu markieren, bevor mit der Ersetzung begonnen wird. Das Markieren hilft dabei, ein Gefühl für die sensiblen Stellen zu bekommen und auf dieser Basis in einem weiteren Schritt die für ein mögliches Re-Identifikationsrisiko relevanten Informationen auszuwählen und ggf. im Team konkrete bedenkliche Stellen zu diskutieren. Das Markieren von ausgewählten Textbeispielen unterstützt dabei die gemeinsame Entscheidung bei Unsicherheiten bzgl. der Sensibilität der Information mit Blick auf einen Fall oder eine Studie und hilft ein konsistentes Vorgehen sicher zu stellen.

Tabelle 1: Beispiel für eine Ersetzungs-/Pseudonymisierungsliste mit unterschiedlichen Abstraktionsniveaus der Ersetzungen

Intervi- ew-Nr.	Interview- bezeichnung						
Zeilen- nr.	Ersetzungs- kategorie	Original- text	Ersetzung, Level 1	Ersetzung, Level 2	Ersetzung, Level 3	Ersetzung, Level 4	Listen- verwendung
1	Interview- ID_1.0						
5	Name	Elisabeth Gerken	Person 1	Person 1, Name der Interviewten 1	Person 1, Name der Interviewten 1, altdeutscher Vorname		
54	Beruf	Altenpfle- gerin	Beruf im Bereich Dienstleistungs- berufe und Verkäufer, ISCO08-5	Beruf im Bereich Betreuungs- berufe ISCO08-53	Beruf im Bereich Betreuungs- berufe im Gesundheit swesen ISCO08- 532	Pflege- helferin ISCO-08 5321	ISCO-08
56	Ort	Bremen	Ort 1, große Großstadt	Ort 1, große Großstadt in Norddeutsch land	Ort 1, große Großstadt in Norddeutsc hland mit Hafenanbin- dung		BBSR
62	Unternehmen	Firma Schmidt	Unternehmen B, Unternehmen im Gesundheits- und Sozialwesen - WZ Q	Unternehm- en B, Unter- nehmen im Sozialwesen WZ Q88	Unternehm- en B, Unter- nehmen zur sozialen Betreuung älterer und behinderter Menschen WZQ88.1		Wirtschafts- zweige

Schritt 8: Ersetzen sensibler Informationen: Pseudonymisieren/ Anonymisieren

Qualiservice empfiehlt ein schrittweises Vorgehen, bei dem zunächst in einer Basisanonymisierung direkte Identifikatoren wie Namen oder Anschrift von Personen ersetzt werden (formale Anonymisierung) oder die Daten pseudonymisiert werden. Die Anonymisierung weiterer Informationen wird dann im Verlauf der Analyse bzw. nach deren Abschluss vorgenommen (sobald es der Forschungszweck erlaubt). Die Frage, ob bestimmte Informationen einen Personenbezug herstellen lassen, sollte ständiger Begleiter bei der Arbeit sein – v. a., wenn bereits Informationen im Text abstrahiert wurden: So können Informationen über die Eltern eines Interviewpartners bereits ihren Personenbezug dadurch verlieren, dass der Interviewte selbst maskiert wurde. Auch pauschale Ersetzungen bestimmter Kategorien (direkte Identifikatoren wie Namen und Adressen ausgenommen) können eine Überanonymisierung begünstigen.¹⁷

Unter Anwendung des gewählten Abstraktionsniveaus (vgl. Schritt 3) werden personenbezogene Informationen so ersetzt, dass so viel wie möglich sozialwissenschaftlich relevante Information erhalten bleibt und gleichzeitig die Re-Identifikation einer Person verhindert wird. Um den Nachnutzungswert möglichst groß zu erhalten, empfiehlt es sich Ersetzungen auf unterschiedlichen Abstraktionsstufen (z.B. Ort, große Großstadt, große Großstadt im Norden Deutschlands, große Großstadt in Schleswig-Holstein etc.), vorzunehmen, so dass z.B. mit dem Anonymisierungstool QualiAnon unterschiedlich stark abstrahierte Versionen eines Textes erstellt werden können (vgl. Punkt 3.4). Orientieren sollte man sich bei Ersetzungen an dem für den Schutz der Persönlichkeitsrechte der Beteiligten notwendigen Maß. Technisch-organisatorische Maßnahmen, die zusätzlich zu Pseudonymisierung/ Anonymisierung vorgenommen werden und einem Re-Identifikationsrisiko entgegenwirken, sollten in die Abwägung einfließen.

Häufig wird die Frage nach möglichen evtl. fachspezifischen Ersetzungskategorien gestellt, die für ein bestimmtes Forschungsthema verwendet werden können. Hier kann auch der Einsatz spezifischer Listen/ Klassifikationen helfen (vgl. Punkt 3.6). Darüber hinaus können bereits in anderen Zusammenhängen von Forscher/-innen ausgearbeitete Ersetzungsschemata wiederverwendet und weiterentwickelt werden. QualiAnon erlaubt z.B. die Publikation, Weiterverwendung und Ausdifferenzierung einmal erstellter Ersetzungsschemata (vgl. Punkt 6).

Werden Ersetzungen vorgenommen, sollten diese im Interview z.B. durch Unterstreichungen, Sonderzeichen oder ähnlichem kenntlich gemacht werden, die nicht Teil der verwendeten Transkriptionsgrammatik sind. Das könnte z.B. wie folgt aussehen: [~ text text text ~]. Die Ersetzungen werden in einer Ersetzungstabelle dokumentiert (vgl. Tabelle 1). Sinnvoll kann ein schrittweises Vorgehen sein, z.B. weil Informationen noch für die Analyse benötigt werden oder weil noch nicht abschließend entschieden wurde, ob Informationen anonymisiert werden müssen oder nicht. Die Ersetzungen werden dann peu a peu vorgenommen.

¹⁷ Das pauschale Markieren bestimmter Informationen kann hingegen helfen die (evtl. auch später zu treffende) Entscheidung über die Ersetzung zu erleichtern und das Vorgehen zu systematisieren.

Häufig leidet die Lesbarkeit eines Transkriptes durch Ersetzungen. Um die Lesbarkeit zu verbessern, könnte man die Ersetzungen grammatikalisch anpassen. Steht z.B. im Originaltext „die Elbe fließt an meinem Geburtshaus entlang“ könnte man – einen Personenbezug vorausgesetzt – den Artikel „die“ in die Ersetzung einbeziehen, um die Lesbarkeit zu verbessern und erhielte z.B. „[der Fluss 1, großer mitteleuropäischer Fluss] fließt an meinem Geburtshaus entlang“. Es gilt allerdings abzuwägen, inwieweit der Originaltext zugunsten der Lesbarkeit überhaupt bearbeitet werden sollte.

Die Office-Funktion „Suchen und Ersetzen“ kann der Orientierung helfen, sollte aber vorsichtig verwendet werden, da Schreibfehler, abgebrochene Wörter, Dialektbezeichnungen für das gesuchte Wort evtl. im Transkript nicht gefunden werden und somit im Original erhalten bleiben. Von der Verwendung der Funktion „alles ersetzen“ raten wir wie z.B. auch die University of Leeds (2008) ab. Auch hier können Schreibfehler, Dialektbezeichnungen, Abkürzungen etc. dazu führen, dass nicht alle Verwendungen des gesuchten Terms gefunden werden (z.B. kann „München“ im Dialekt zu „Minga“ werden und würde nicht mit ersetzt). Darüber hinaus werden Ersetzungen evtl. in Wörtern vorgenommen, die dieselbe Buchstabenreihenfolge enthalten: Ich suche Michael und will diesen mit „Person 1, Freund des Interviewten“ ersetzen. Im Text finde ich die zugehörige Kurzform „Micha“ nicht; zusätzlich wird Michaela, die Ehefrau des Befragten bis auf das „a“ mit ersetzt und damit zu derselben Person wie der Befragte selbst.

Schritt 9: Prüfen der Anonymisierung/ Pseudonymisierung

Abschließend muss das anonymisierte/ pseudonymisierte Forschungsmaterial noch einmal geprüft werden. Die Prüfung richtet sich auf evtl. noch im Text enthaltene mögliche Re-Identifikationsrisiken für Personen. Geprüft werden aber auch die Ersetzungen: Es könnten z.B. Wortreste oder ganze Markierungen stehen geblieben sein oder es fällt auf, dass eine Ersetzung unnötig ist oder dass Zusatzinformationen fehlen. Ist die Prüfung durch die Forscher/-in selbst abgeschlossen, können die Materialien an Qualiservice weitergegeben werden. Werden anonyme Materialien weitergegeben, muss die Originalinformationen aus der Ersetzungsliste gelöscht werden. Qualiservice prüft die Anonymisierung und ergänzt diese wenn nötig.¹⁸

5 Zentrale Kategorien für die Ersetzung personenbezogener Information

Im Folgenden präsentieren wir zentrale Ersetzungskategorien im Überblick. Es empfiehlt sich, jede Ersetzung darauf hin zu prüfen, ob sie nötig ist, um das Re-Identifikationsrisiko für eine Person zu verringern. Zentral ist, ein Vorgehen zu finden, welches der jeweiligen Studie und den vorliegenden Forschungsmaterialien gerecht wird und damit gut zu Thema und Daten passt. Das bereits genannte schrittweise Vorgehen (vgl. Schritt 8) kann hier unterstützen. Von einem routinemäßigen Ersetzen der genannten Kategorien ist abzuraten. Ob eine Ersetzung

¹⁸ Wichtig zu wissen, nicht alle FDZ prüfen wie Qualiservice die Anonymisierung der übermittelten Daten. Manche FDZ überlassen dies den Datengebern/-innen. Ist man am Data Sharing interessiert, sollte man sich deshalb bzgl. des Vorgehens im gewählten FDZ und der evtl. mit einer Anonymisierung verbundenen Haftung informieren.

notwendig ist, sollte bei jeder einzelnen Ersetzung geprüft werden: Oft verändert die Ersetzung einer oder weniger Kategorie/n bereits den Informationsgehalt in einem Interview so, dass eine Re-Identifikation unwahrscheinlicher wird. Werden z.B. Lehrer/-innen einer Großstadt befragt, ist diese Gruppe viel größer als die der Justizvollzugsbeamten/-innen in derselben Stadt. Je nach im Interview vorhandenen Kontextinformationen könnten im ersten Fall evtl. berufliche Tätigkeit und der Name der Stadt, in der geforscht wurde, erhalten bleiben oder eine Abstraktion auf Regionsebene ausreichen, um den Personenbezug aus dem Text zu eliminieren.¹⁹

Zu den Kategorien, die bei der Ersetzung häufig eine Rolle spielen, zählen: Namen, Zeitangaben, Ortsangaben, Informationen zu Ausbildung und Beruf, Namen von Organisationen und Institutionen, individuelle Besonderheiten sowie Positionen und Rollen (vgl. Tabelle 2). Entlang dieser (und auch anderer) Kategorien können Ersetzungen vorgenommen und ausdifferenziert werden. Im Folgenden gehen wir auf die hier zusammengestellten Ersetzungskategorien ein und greifen dabei auf Beispiele und Erfahrungen aus der Beratungsarbeit von Qualiservice zurück. Die hier präsentierten Ersetzungskategorien bilden die Grundlage für ein Kategorienschema, welches auf der Ebene der Kategorien und themenspezifisch weiterentwickelt sowie im Qualiservice-Anonymisierungstool QualiAnon eingelesen und bearbeitet werden kann (vgl. Punkt 6).

¹⁹ Ein passendes Beispiel präsentiert auch das Finnish Social Science Data Archive (FSD) hierzu: „We were never really very religious even though my aunt was one of the first women to be ordained as priest in Finland.“ Es wird diskutiert, dass die Information, dass die Tante des Interviewten eine der ersten Pastorinnen Finnlands war, Rückschlüsse auf den Befragten ermöglichen könne. Weiß man allerdings, dass bei der ersten Ordination 94 Frauen als Pastorinnen eingesetzt wurden, relativiert sich diese Einschätzung (<https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/> (Zugriff: 15.12.2021)).

Tabelle 2: Die für die Ersetzung personenbezogener Informationen zentralen Ersetzungskategorien im Überblick (zusammengefasst aus der Beratungs- und Kurationspraxis von Qualiservice)

	Ersetzungskategorie	Mögliche Unterkategorien	Originalinformation	Mögliche Ersetzungen ohne differenzierte Abstraktionsgrade
1	Charakteristika der Person	Name, Funktion für den Befragten, Anzahl der Kinder, Geschlecht	Jutta	Name der Ehefrau
2	Zeitangaben	Datum, Zeitspanne, Monatsangaben, Altersangaben	Aug 2002 – Okt 2005	Sommer 2002 bis Herbst 2005
3	Ortsangaben	Land, Region, Bundesland, Gemeindeverband, Stadt, Stadtteil, Straßennamen, Gewässer, Gebirge Landstriche, Wohnort, Arbeitsort, Andere	Stadt Bremen	Stadt A, große Großstadt in Norddeutschland
4	Ausbildung	Bildungsabschluss, Fach, Schulart, Weiterbildung, Ausbildungsform, andere	Altenpflegeausbildung	Ausbildung zur Fachkraft für einen Assistenzberuf im Gesundheitswesen (ISCO3200)
5	Berufliche und vergleichbare Tätigkeiten	Funktionen, Beruf, Tätigkeit, ehrenamtliches Engagement, Positionen und Rollen, Andere	Rettingsanitäter	Vertreter des Rettungsdienstpersonals (ISCO08 3258)
6	Organisationen und Institutionen	Firmen, Schulen, Ausbildungsstätten, Behörden/Verwaltung, Wirtschaftszweige, Andere	Interviewdienst Horch	Unternehmen A - Unternehmen für sonstige freiberufliche, wissenschaftliche und technische Tätigkeiten (WZ M74)
7	Individuelle Besonderheiten	Krankheiten, Preise, sportliche Positionen, Unfall, Andere	Herzinfarkt	Erkrankung des Kreislaufsystems (ICD10 I00-I99)
8	Andere	Sprachen, Nationalität, Religion	italienisch	Südeuropäische Sprache, in der Landessprache, seine Erstsprache, seine Muttersprache

5.1 Ersetzung von Namen

Namen identifizieren Personen relativ eindeutig. Sie gehören zu den direkten Identifikatoren.²⁰ Haben Personen nicht in die Verwendung und Weitergabe ihrer personenbezogenen Daten eingewilligt²¹, müssen Namen (z.B. von Forschungsteilnehmer/-innen, Forscher/-innen oder von anderen sog. Dritten) anonymisiert werden. Dafür stehen unterschiedliche Optionen zur Verfügung (vgl. Tabelle 3).

Muss ein hohes Abstraktionsniveau gewählt werden, um die Person zu schützen, wählt man z.B. die Ersetzungskategorie „Name einer Person“ oder „Person“. Mehr Informationen werden allerdings schon durch die bloße Ergänzung einer Nummerierung erhalten („Name der Person 1“ oder „Person 1“). Die Nummerierung übernimmt die Funktion, die ersetzte Person 1 im Text oder bei studienbezogenem Vorgehen auch in der Studie als ein und dieselbe Person auszuweisen. Einen weiteren Mehrwert erhält man, wenn Informationen z.B. über die Funktion, die der/ die Namensträger/-in im Interview hat (z.B. Name des Interviewten) oder auch die Art der Beziehung, in der Namensträger/-innen zueinanderstehen (z.B. „Name der Ehefrau von Person 1“, „Mutter des Befragten“) erhalten werden. Weitere Informationen wie z.B. das Geschlecht des Namensträgers oder die Herkunft eines ersetzten Namens (z.B. deutscher Familienname, arabischer Familienname, Jungenname, Mädchenname) bieten weitere Möglichkeiten, um den Nachnutzungswert zu erhöhen.

Eine weitere Möglichkeit ist die Ersetzung durch andere (Vor-)Namen – sog. Pseudonyme.²² Erfahrungen mit diesem Vorgehen beschreiben z.B. Saunders, Kitzinger and Kitzinger (2015a). Die Wahl von Pseudonymen aus einem ähnlichen Kulturkreis und Sprachkreis und/ oder die Wahl von Pseudonymen mit ähnlicher Bedeutung kann einen Informationszugewinn bedeuten. Hier gilt es aber zu bedenken, dass mit der Namensgebung kulturelle und sozialstrukturelle Hinweise z.B. auf die soziale Herkunft einer Person oder auch ihr Alter verbunden sein können (Gerhards, 2005, 2010).²³ Diese könnten bei Ersetzungen mit anderen Namen verloren gehen oder im ungünstigeren Fall so verändert werden, dass „falsche Fährten“ in das Material gelegt werden. Die Wahl eines anderen Namens muss vor diesem Hintergrund gut überlegt und sollte dokumentiert werden (vgl. Punkt 4, Schritt 5). Entscheidet man sich für diese Form der Ersetzung können Namenslisten bei der Ersetzung helfen.

²⁰ Eine Übersicht zu direkten Identifikatoren präsentieren z.B. in Watteler, O., & Ebel, T. (2019).

²¹ Im Rahmen von Open Science, Zeitzeugeninterviews oder im Rahmen von partizipativer Forschung kommt es häufiger vor, dass Namen und weitere Informationen – auch auf Wunsch der Teilnehmer/-innen – erhalten bleiben sollen.

²² Der Begriff des Pseudonyms wird in diesem Kontext auch verwendet, wenn eine Dokumentation des Originalnamens nicht stattfindet, die Namensinformation also pseudonymisiert wäre.

²³ Gerhards beschreibt wie neben individuellen Präferenzen eines Paares, der soziale Status und das kulturelle Milieu die Auswahl eines Vornamens für das gemeinsame Kind und damit die Identität der Namensträger beeinflussen und auch für Distinktion zu anderen kulturellen Milieus sorgen Gerhards, J. (2005, 2010). In der Beschreibung eines Auswahlprozesses eines Paares wird dies deutlich: „The repertoire of potential first names was thus confined to those acceptable to the parents' social milieu. Names from films, television, and sports were judged to be vulgar, since they were often used by the lower classes. Old German names like Wilhelm, Uta or Otto were rejected on grounds of being too traditional and otherwise conservative. Jewish names like Sarah and Daniel would have been considered if they hadn't been used a decade earlier by the same milieu; they were shopworn and did not satisfy the parents' need to christen their child with something distinctive and individual.“ Gerhards, J. (2005).

Beispiele für offiziell herausgegebene Listen finden sich u.a. bei der Gesellschaft für Deutsche Sprache, die seit 1977 jährlich Listen der in Deutschland am häufigsten vergebenen Vornamen herausgibt oder bezogen auf deutsche Nachnamen können Informationen des Deutschen Familiennamenatlas hilfreich sein, der auch Hinweise auf die räumliche Verbreitung von Namen gibt.

Tabelle 3: Beispiele für die Ersetzung von Namen

Original	Mögliche Ersetzungen	Hinweise
Gerda	Helga, Interviewte 1 (althochdeutscher Vorname), Namensnennung der Interviewten, Person 1	
Meine Cousine Berta	Name der Cousine der Interviewten, Name der Cousine mütterlicherseits der Interviewten	
Tim Schulze	Interviewer 1 (männlich)	Auch Informationen über Interviewer/-innen sollten dokumentiert werden, da damit wichtige Hinweise den Interviewverlauf oder die Interviewführung betreffend, verbunden sein können. So macht es z.B. einen Unterschied, ob ein Werkstudent oder eine von Außen kommende Wissenschaftlerin den Personalrat interviewt; oder ob eine forschende Hebamme oder eine Soziologin einer Schwangeren Fragen zur Versorgung stellt.

Beziehen sich Aussagen auf in der Öffentlichkeit stehende Personen – z.B. der/ die Bundeskanzler/-in – und deren Aufgabe bzw. stehen die Aussagen mit dem öffentlichen Interesse in Zusammenhang, können die Namen erhalten bleiben, sofern sich die Person als Amtsträger äußert und sofern damit kein Re-Identifikationsrisiko für weitere nicht in der Öffentlichkeit stehende Dritte verbunden ist. Beziehen sich die Aussagen allerdings auf Informationen, die nichts mit dem ausgeübten Amt zu tun haben, sondern z.B. die Privatsphäre der Person betreffen, müssen diese ebenso verändert werden.

5.2 Ersetzung von Zeitangaben

Auch Zeitangaben können einen Personenbezug begründen. Zu dieser Ersetzungskategorie gehören u.a. Datumsangaben, Angaben von Zeitspannen, Angaben zum zeitlichen Verlauf sowie Beginn oder Ende von bestimmten Lebensphasen. Auch bei der Abstraktion von Zeitangaben gilt es genau zu überlegen, ob die zu ersetzende Information dazu beitragen können, eine Person zu reidentifizieren. Bei Angaben wie z.B. dem Geburtsdatum einer Person ist ein Personenbezug i.d.R. gegeben, anders kann das z.B. bei Angaben zur Dauer eines Beschäftigungsverhältnisses sein, hier entscheidet der Interviewkontext wie sensibel die Information ist – insbes. dann, wenn z.B. andere Informationen wie der Ort der Beschäftigung oder Angaben zur Firma bereits abstrahiert sind. Tabelle 4 zeigt Beispiele für die Ersetzung.

Tabelle 4: Beispiele für die Ersetzung von Zeitangaben

	Originaltext	Ersetzungsoptionen	Hinweise
Datumsangabe	01.04.2020	Im Frühjahr 2020, im April 2020, 2020, [Nennung des Datums]	Die Entsprechung für Frühjahr, Sommer, Herbst und Winter sollte dokumentiert sein, z.B. Frühjahr: 21. März bis 20. Juni
Monatsangabe	01.04.2020	Anfang April 2020, im April, Monat xy	Die Entsprechung für Anfang, Mitte, Ende eines Monats, Jahres, Jahrzehnts, Jahrhunderts etc. sollte dokumentiert werden. Für Monate könnte man z.B. 1.-10. Tag eines Monats: Anfang des Monats, 11-20. Tag = Mitte des Monats, 21-31. Tag = Ende des Monats wählen.
Jahresangabe	2002	Zu Beginn der 2000er Jahre, Anfang des 21. Jahrhunderts	
Altersangabe	Ich war 8 Jahre alt.	In meiner Kindheit, Bevor ich 10 Jahre alt war, [genaue Altersangabe]	Hier braucht es die Dokumentation der Angaben für die Verwendung der Altersintervalle. Z.B. 40-43 = Anfang 40, 44-46 = Mitte 40, 47-49 Ende 40. Von Bedeutung ist aber auch der informationelle Gehalt für die Wahl der Abstraktion. Dieser ergibt sich aus dem Interview- oder Studienkontext So kann z.B. die Altersangabe 25 leicht mit Mitte 20 ersetzt werden. Spricht jemand aber über das 65. Lebensjahr könnte über das Alter hinaus z.B. auch die Information „Rentenbeginn“ für den Nachvollzug des Erzählten wichtige Bedeutung haben und sollten sich in der Abstraktion evtl. wiederfinden. Genauere Informationen können dem Datum in Klammern hinzugefügt werden.
	Zu meinem 18. Geburtstag.	Als ich volljährig wurde, als junger Erwachsener	
Zeitspanne	Von Mitte April 2004 bis Ende Dezember 2004	Von Frühjahr 2004 bis Winter 2010; von Monat A bis Monat J (10 Monate), eine Zeit lang in 2004, mehrere Monate in 2004, mehrere Monate Anfang der 2000er Jahre, von Ostern 2004 bis Weihnachten 2004	Bei Ersetzungen der 12 Monate mit Monat A-L oder 1-12 sollte man sich bewusst darüber sein, dass die Reihenfolge der Buchstaben und Ziffern eine Reihenfolge der Monate suggerieren kann. Auch kann durch die Ersetzung mit Monat A- bis Monat B die Information über die Dauer einer Zeitspanne verloren gehen, die man vielleicht besser in Klammern (10 Monate, mehrere Monate usw.) ergänzt. Evtl. hilft die Abstraktion in Jahreszeiten.

5.3 Ersetzung von Ortsangaben

Die Vielfalt von Ortsangaben ist groß. Die Bandbreite reicht von Angaben zu Ländern, Bundesländern, Regionen, Gemeinden und Gemeindeverbänden, Städtenamen, Namen von Stadtteilen hin zu Straßennamen, Gewässerbezeichnungen, Namen von Gebirgen, Landstrichen etc. (Eine Übersicht über häufig auftretende Beispiele liefert Tabelle 5).

Ortsangaben können einen Personenbezug beinhalten. In Interviews finden sich Ortsangaben aber auch oft ohne konkreten Bezug zu einer Person. So sind die Dienstreise nach Berlin oder der Urlaub an der Ostsee per se kein personenbezogenes Datum. Oder: In einer Studie zu Lehrer/-innen in Hamburg kann der Ort Hamburg unter Umständen benannt werden, weil das mit der Information verbundene Re-Identifikationsrisiko etwa aufgrund der Zahl der zur Berufsgruppe zugehörigen Personen in dieser Großstadt gering ist. Ähnlich beschreiben es Saunders et al. (2015a): In ihrer Studie würde die Nennung des Ortes einer behandelnden Klinik ein Re-Identifikationsrisiko für die Interviewpartner/-innen darstellen, weil nur sehr wenige Kliniken die beforschte seltene Erkrankung behandeln.

Eine generelle Ersetzung aller Orte, ohne den Personenbezug zu prüfen, würde in der Regel das Nachnutzungspotenzial eines Datensatzes einschränken. Der Kontext, in den die Schilderungen der Forschungsteilnehmer/-innen eingebettet ist, ist ebenso wie der Kontext der Studie selbst entscheidend für die Ersetzung von Ortsinformationen. Dabei gilt es auch zu prüfen, ob der Ort der Erhebung bereits in Publikationen oder in Projektbeschreibungen veröffentlicht wurde und dadurch eine Anonymisierung im Interview evtl. obsolet wäre.

Ersetzungen von Ortsbezeichnungen können unterschiedlich stark abstrahiert vorgenommen werden. Z.B. könnte die Funktion, die ein Ort für den Interviewten hat, sozialwissenschaftlich relevant sein (z.B. Geburtsort, Wohnort, Arbeitsort). Daneben liefern Angaben zu Größe, Lage oder Infrastruktur eines Ortes für die Nachnutzung wichtige Informationen. Spricht eine Interviewpartnerin z.B. davon, dass sie in Bremen arbeitet und die Information ist als sensibel anzusehen, könnten trotz notwendiger Ersetzung des Ortes „Bremen“, Informationen wie „große Großstadt in Norddeutschland“ wichtige Hinweise für die Nachnutzung erhalten. Auf welcher Ebene (Gemeinde, Land, Bundesland, Region) und in welcher Breite Informationen erhalten bleiben, entscheidet sich im konkreten Fall.

Tabelle 5: Beispiele für zu anonymisierende Ortsangaben und Möglichkeiten ihrer Ersetzung

	Originaltext	Ersetzungsmöglichkeiten	Hinweise
Städte/ Stadtteile	Bremen	Ortsangabe, Ort 1, Stadt A, Stadt 1	Man kann Ortsangaben einfach mit dem Hinweis darauf ersetzen, dass es sich um eine Ortsangabe handelt. Der Informationsgehalt der Ersetzung ist so allerdings gering. Größer wird dieser bereits durch eine Zählung mit Buchstaben oder Ziffern. So kann z.B. nachvollzogen werden, ob es sich um dieselben Orte handelt.
	Bremen	Stadt 1 (große Großstadt), Stadt 1 (große Großstadt in Norddeutschland)	Hier wird sich auf die BBSR-Klassifikation bezogen. Weitere Hinweise wie die genauere Bestimmung der Lage oder Informationen zur Infrastruktur können ergänzt werden.
	Bremen	Große Großstadt A, Wohnort 1	Ein Informationsgewinn wird auch dadurch erreicht, dass die Funktion des Ortes – z.B. Wohnort, Arbeitsort, Ort der Entspannung etc. erhalten bleibt. Die Zählung kann die Unterscheidung zwischen verschiedenen Orten ermöglichen. Dies könnte z.B. relevant sein, wenn in einem Interview über berufsbedingtes Pendeln gesprochen wird. In diesem Fall könnten z.B. Entfernungen zwischen Arbeitsort 1 und Arbeitsort 2 in Klammern angegeben werden.
	Schwachhausen	Stadtteil A (oder 1) der großen Großstadt A, Stadtteil A (oder 1) der Stadt A	Auch hier können, z.B. in Klammern, Informationen zur Infrastruktur oder sozialstrukturellen und politischen Charakteristik des Stadtteils gegeben werden. Beispiele sind Angaben zur Arbeitslosigkeit, zum Migrationshintergrund, Einkommens- und Bildungsstruktur. Je nach Forschungsprojekt können aber auch Informationen wie Vorhandensein von Grünflächen, verkehrstechnische Erschließung etc. interessante Informationen sein.
	Rom	Hauptstadt im (südeuropäischen) Ausland + Nummer	Analog zu den für Städte beschriebenen Ersetzungen können Städte im Ausland abstrahiert werden. Listen wie die der European Statistics (EUROSTAT) ²⁴ könnten hierzu hilfreich sein.
	X-Stadt	Nicht identifizierbarer Ort + Buchstabe	Nichtidentifizierbare Orte könnten im Interview genauso auch wiedergegeben werden. Die Ersetzung von einer evtl. im Interview unverständlich gebliebenen Stadt mit „X-Stadt“ könnte Fragen bei der Nachnutzung aufwerfen, gerade wenn versucht wird, an anderer Stelle so viel Information zu erhalten wie möglich. Der Hinweis darauf, dass der Ort z.B. nicht identifizierbar ist oder unverständlich geblieben ist, ist für die Nachnutzung bedeutsam.

²⁴ https://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP_PUB_WELC (letzter Zugriff: 23.06.2022)

Straßennamen	Mehrmannstraße	Straße + Nummer	Funktionen einer Adressangabe für den/die Interviewten/-e können z.B. darin bestehen, dass er/sie in der genannten Straße wohnt, arbeitet, zur Schule geht etc. Eine Ziffer oder einen Buchstaben anzuhängen, kann die Information erhalten, dass es sich z.B. um verschiedene Adressen oder dieselbe Wohnadresse handelt.
	Mehrmannstraße	Wohnadresse (mit oder ohne Zählung)	
Länder	Kanada	Nordamerikanisches Land	mit oder ohne Zählung, ggf. Informationen, die das Land näher charakterisieren
Gewässer	Chiemsee	See	Der See kann z.B. in seiner Größe oder Lage weiter bestimmt werden. Evtl. ist auch eine Zählung sinnvoll, wenn mehrere Seen im Text präsentiert werden.
Region	Ostpreußen	europäische Region	
Konkrete Orte	Bremen Flughafen	Verkehrsknotenpunkt 1 in große Großstadt A	Auch hier können weitere Informationen gegeben werden, die je nach Sensibilität des Datums mehr oder weniger vom Ursprungsort – dem Flughafen – abstrahieren.
	Saftschuppen	Lager in Bahnhofsnähe	Die Ersetzung des Ortes mit seiner Funktion für die Interviewte (z.B. Arbeitsort, Treffpunkt, ...) kann - auch ohne Zählung - sozialwissenschaftlich relevante Information erhalten. Diese Form der Ersetzung lässt sich z.B. gut bei Ersetzungen von Geschäften, Diskotheken, Fitnessstudios etc. verwenden.

Erfahrungsgemäß ist es hilfreich, wenn die Ersetzung von Ortsangaben studienbezogen vorgenommen wird: D.h., dieselbe Ortsbezeichnung wird in allen Interviews einer Studie auf immer dieselbe Weise ersetzt. Somit bleibt der Datensatz auch eher als Ganzer verbunden und die Interviews untereinander aufeinander beziehbar. In einer Studie zu Wendeerfahrungen, die in den 1990er Jahren in zwei Orten durchgeführt wurde, könnten so z.B. die Erfahrungen von Menschen eines Ortes 1 (kleine Landgemeinde, starke religiöse Prägung, hohe Arbeitslosigkeit) mit den Erfahrungen der Menschen des Ortes 2 (mittlere Kleinstadt, Universitätsstadt, gute Infrastruktur) verglichen werden, obwohl der Ort wegen eines mit seiner Nennung möglichen Re-Identifikationsrisikos für die Interviewten abstrahiert wurde.

Werden Ziffern für die nähere Bestimmung verwendet, kann eine unbegrenzte Anzahl von Orten eindeutig zugeordnet werden. Die Verwendung von Buchstaben (A bis Z) führt schnell zu unübersichtlichen und oft fehleranfälligen Buchstabenreihen (z.B. muss man nach der Verwendung der Ersetzung „Stadt Z“ mit „Stadt AA“ o.ä. beginnen). Die Verwendung von Buchstaben beugt wiederum der Annahme einer möglichen Priorisierung einzelner bezifferter Orte vor, die im Interview nicht intendiert war – z. B. wenn verschiedene Arbeits- oder Wohnorte einer Person im Text präsentiert werden und mit der Bezifferung der Eindruck vermittelt würde, dass Wohnort 1 der priorisierte Wohnort sei.

Für die teilstandardisierte Ersetzung mit Listen (vgl. Punkt 3.6) von Orten in Deutschland kann die Übersicht über Stadt- und Gemeindetypen des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR) verwendet werden. Diese Liste beinhaltet u.a. den Gemeindennamen, das Bundesland, den Namen des Gemeindeverbandes und seine Nummer, Angaben zu Fläche und Einwohnerzahl sowie differenzierte und weniger differenzierte Angaben zu den Gemeindetypen²⁵. Bei Anwendung der Liste kann der Gemeindename in der Liste gesucht werden und je nach gewähltem Vorgehen durch zugehörige Informationen auf einer höheren Abstraktionsebene wie dem Bundesland oder dem differenzierten Gemeindetyp abstrahiert werden.

Die Gemeindetypen, die das BBSR verwendet, sind in drei Haupttypen kategorisiert, die wiederum qualitativ unterschieden werden können. Anhand der Legende können Informationen über Regionen vermittelt werden. Die Gemeindetypen geben Auskunft über Fläche und Einwohnerzahl und auch darüber wie viele Gemeinden und Gemeindeverbände der jeweiligen Kategorie zugeordnet werden können.

25

<https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/gemeinden/StadtGemeindetyp/StadtGemeindetyp.html>, Zugriff: 01.08.2023

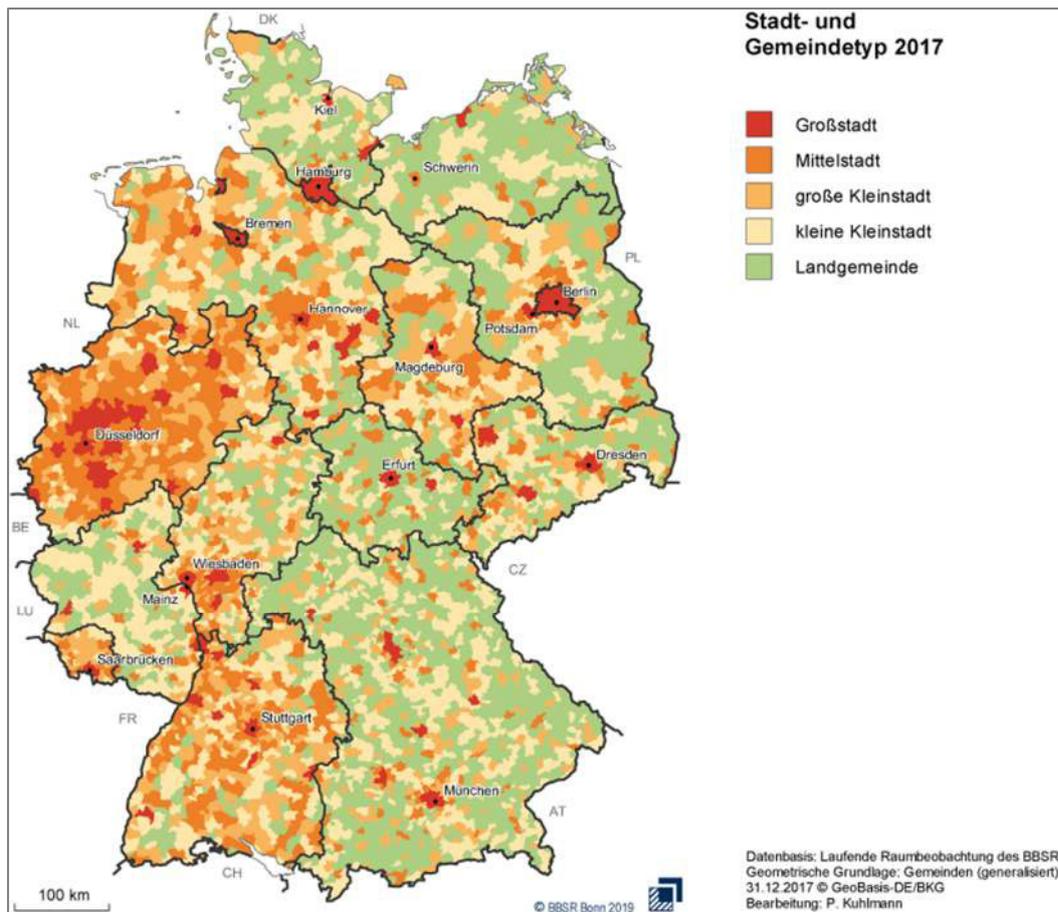


Abbildung 2: Stadt- und Gemeindetypen des BBSR, Quelle: <https://www.bbsr.bund.de/BBSR/DE/forschung/raumbewertung/Raumabgrenzungen/deutschland/gemeinden/StadtGemeindetyp/StadtGemeindetyp.html>, Zugriff: 01.08.2023

Über die Gemeindekennziffer können Informationen zu den jeweiligen Orten, die in weiteren Listen des BBSR aufgelistet sind, kombiniert werden. Hier finden sich z.B. ortsbezogene Informationen zur Infrastruktur des Ortes wie Besiedelung und Lage, kreisbezogene Informationen auf Regionsebene (z.B. Arbeitslosigkeitsregionen, Wohnungsmarktregionen etc.) oder zu seiner (aggregierten) Einwohner- und Arbeitslosenzahl. Eine Ersetzung könnte wie folgt aussehen: „Ort 1, große Großstadt in Mitteldeutschland, niedrige Arbeitslosigkeit“

Weitere Verlinkungen zu Informationen auf europäischer Ebene sind möglich über die NUTS-Regionen²⁶ und die siedlungsstrukturellen Regionstypen Europas. Auch auf der Ebene der NUTS-Regionen kann man sich mittels der Gemeindekennziffer bewegen. Weitere Informationen zur Beschreibung konkreter Orte und Regionen auf europäischer Ebene sind z.B. über die Statistiken der Europäischen Kommission²⁷ erhältlich. Darüber hinaus stellen die United Nations Organisation (UNO oder IFO) eine Liste zur Verfügung, die auch

²⁶ Nomenclature of territorial units for statistics

²⁷ EUROSTAT: <https://ec.europa.eu/eurostat/de/web/regions-and-cities>

Länderbezeichnungen und Ozeanien beinhaltet.^{28,29} Entscheidet man sich für die Verwendung von Listen, können diese in QualiAnon eingelezen und so für die Bildung von Ersetzungen verwendet werden (vgl. Punkt 6).

5.4 Ersetzung von Informationen zu Ausbildung und Beruf

Angaben zu Berufen, beruflichen Tätigkeiten, Ausbildungen etc. können personenbezogene Daten enthalten. Auch bei dieser Kategorie ist es für den Erhalt des Nachnutzungspotenzials des Materials wichtig, Ersetzungen gut abzuwägen, da eine Abstraktion immer mit einem Informationsverlust verbunden ist und das Nachnutzungspotenzial des Materials verringert. Bezogen auf Ausbildung und Beruf diskutieren z.B. Thomson et al. (2005) welche Schwierigkeiten in einem interdisziplinären Forschungsprojekt mit der Ersetzung von „nurse“ und „physician“ durch „health professional“ verbunden waren.³⁰ In manchen Fällen bietet sich evtl. auch an, die Berufsbezeichnungen unberührt zu lassen, weil z.B. die Ersetzung anderer Informationen genügt, um eine Re-Identifikation von Personen zu verhindern ebenso wie eine ausreichend große Anzahl von Vertreter/-innen einer Berufsgruppe (siehe das Beispiel der Priesterin in Fußnote 19) bzw. eben gerade nicht (siehe das Beispiel der Justizvollzugsbeamten auf Seite 21). Eine Abwägung sollte deshalb fallbezogen vorgenommen werden.

Berufe können auf unterschiedliche Weise abstrahiert werden. Beispiele für Ersetzungen finden sich in Tabelle 6. Um keine „falschen Fährten“ zu legen, empfiehlt Qualiservice Beruf und berufliche Tätigkeiten unterschiedlich stark zu abstrahieren und nicht mit anderen Berufen zu ersetzen. Ersetzte man z.B. den Physiotherapeuten mit dem Masseur würde man zwei Berufsbilder vermischen, die sich mindestens bzgl. der Ausbildung und deren Zugangsvoraussetzungen unterscheiden. Besser wäre Physiotherapeut und Masseur zu abstrahieren und mit einer Sammelkategorie „personenbezogener Dienstleistungsberuf im Gesundheitswesen“ zu ersetzen, die durch weitere Informationen wie „Fokus auf den Bewegungsapparat“ für den Physiotherapeuten ergänzt werden könnte. Es gehen zwar auch dann Informationen verloren (vgl. Thomson et al., 2005), aber es werden keine Informationen in den Text eingebaut, die im Original nicht enthalten sind.

Bei der Suche nach adäquaten Ersetzungen mit unterschiedlich hohem Abstraktionsniveau können hierarchisch aufgebaute standardisierte Listen helfen. Für Berufe eignet sich die

²⁸ Informationen und Tabellen zu lokalen Verwaltungseinheiten, den sog. NUTs finden sich hier:

<https://ec.europa.eu/eurostat/de/web/regions-and-cities> und auf folgender Website

http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NUTS_2013L&StrLanguageCode=EN&StrLayoutCode=HIERARCHIC (Zugriff 12.05.2020).

²⁹ https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/_node.html

³⁰ „For example, in interdisciplinary discussions between researchers in business and nursing, the business researchers proposed replacing the words "nurse" and "physician" with the broader category of "health care professional" for the sake of preserving the individual professionals' anonymity. From an organizational analysis perspective, concerned with the professions as a group, this was a possibly acceptable solution. Our nursing colleague objected, however, saying that a participant's identity as a nurse or physician was crucial to data analysis“ Thomson et al. (2005).

International Classification of Occupations (ISCO)³¹ der Internationalen Arbeitsorganisation (ILO). Diese Klassifikation ist etabliertes Instrumentarium, um Angaben zu Berufen und beruflichen Tätigkeiten in quantitativen Datensätzen zu anonymisieren (Ebel, o. J.; Züll, 2015). Die ISCO unterscheidet Berufshauptgruppen, Berufsgruppen, Berufsuntergruppen und Berufsgattungen in einem hierarchisch gegliederten Klassifikationssystem, welches unterschiedlich stark abstrahierte Informationen zu den jeweiligen Berufen und beruflichen Tätigkeiten enthält.³² Die Einordnung der Qualifikationsstufen folgt der International Standard Classification of Education (ISCED)³³ – mit Grundbildung, Sekundärbildung/ Grundberufsausbildung/ berufliche Erstausbildung, postsekundäre Bildung und Tertiärbildung sowie der Hochschulausbildung. Dies ermöglicht es vier Qualifikationsstufen (skill level) (Hilfsarbeitskräfte, Fachkräfte, gehobene Fachkräfte und akademische Berufe) zu klassifizieren.

Die ISCO-Klassifikation ist hierarchisch aufgebaut. Dies ermöglicht, Informationen unterschiedlich stark zu abstrahieren, ohne falsche Fährten zu legen. Je höher der gewählte Abstraktionsgrad, desto größer ist die Zahl derjenigen Berufe, die in die jeweilige Kategorie fallen, was das Re-Identifikationsrisiko verringert. So kann der Beruf der Altenpflegerin je nach Abstraktionsstufe zum „Dienstleistungsberuf (ISCO 08 Code 5)“, „Betreuungsberuf (ISCO08 Code 53) oder z.B. zu einem „Betreuungsberuf im Gesundheitswesen (ISCO-08-Code 532) werden. Zusätzliche Informationen wie z.B. die Zielgruppe, die betreut wird oder der Betreuungssektor (ambulant vs. stationär) können als weitere wichtige Hinweise für die Nachnutzung ergänzt werden. Im Beispiel könnte die Ersetzung folgendermaßen aussehen:

„Frau A.: (1) Also, ich, (1) beruflich hatte ich eigentlich anders angefangen. Ich war [[Nennung des Berufs, Gruppe: Hauswarte und Hauswirtschaftsleiter ISCO: 515 000]], hab dann halt geheiratet, zwei Kinder, aufgehört zu arbeiten, war ungefähr 10 Jahre zu Hause und hab dann über nen [[Ausbildungskurs in der Gesundheitsversorgung]] in [[einer Einrichtung der Gesundheitsversorgung für ältere Menschen³⁴]] ein Praktikum gemacht und bin dann da gleich gefragt worden, ob ich nicht mal vertretungsweise als [[Nennung des Berufs, Betreuungsberuf im Gesundheitswesen (ISCO 532 000)]] (2) einspringen könnte. Hab ich dann auch gemacht und bin dann da einfach hingengeblieben. Das, aus dem Mal einspringen wurden dann [[mehrere, >5]] Jahre Arbeit und dann hab ich gedacht, na gut, wenn ich jetzt diese Arbeit schon mach, dann mach ich auch die Ausbildung dazu. und hab dann die [[Ausbildung zur Fachkraft für einen Assistenzberuf im Gesundheitswesen (ISCO 3200)]] gemacht ...“³⁵

Auch Angaben zur Ausbildung können personenbezogen sein. Die Komplexität und Vielgestaltigkeit und damit auch fehlende Vergleichbarkeit der Ausbildungs- und

³¹https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=CL_ISCO08&StrLanguageCode=DE&IntPcKey=&StrLayoutCode= (Zugriff am 01.12.2021)

³² Neben der ISCO-08 findet in den Sozialwissenschaften auch die Klassifikation der Berufe (KldB) Anwendung. In der KldB 2010 werden Berufsbezeichnungen erfasst. Sie hält eine komplette Liste von Berufsbezeichnungen in bestimmten Branchen vor. Sollen nicht Berufe, sondern Tätigkeiten im Mittelpunkt der Information stehen, bietet sich evtl. eher die Nutzung der ISCO-08 an.

³³ Vgl. <https://www.bfs.admin.ch/asset/de/do-d-15.01-iscsed-01>, Zugriff: 31.07.2023

³⁴ Könnte z.B. weiter ergänzt werden mit sozialwissenschaftlich relevanter Information wie *Einrichtung im ländlichen Raum, großer Arbeitgeber für Frauen, mit weniger als 100 Personen in der Versorgung, ...*

³⁵ Die hier vorgenommenen Ersetzungen sind in doppelten Klammern eingefügt.

Bildungssysteme stellt eine besondere Herausforderung dar. Reicht ein hoher Abstraktionsgrad aus, kann die ISCED der UNESCO³⁶ unterstützen. Die ISCED zielt auf die Vergleichbarkeit der Bildungsinstitutionen auf europäischer Ebene ab. Sie ist in den „skill levels“ der ISCO-08 repräsentiert. Die Klassifikation beinhaltet zehn Qualifikationsstufen, welche von frühkindlicher Bildung im Kindergarten bis zur abgeschlossenen Promotion an Universitäten reichen. Ein Blick auf das komplexe und vielfältig verzweigte deutsche Bildungssystem macht aber deutlich, dass hier der Verwendung der ISCED schnell Grenzen gesetzt sind, wenn Informationen zu Binnendifferenzierung, Varianz und Spezifika des deutschen Bildungssystems im Datensatz erhalten werden sollen. Ob ein Mehrwert mit der Listenverwendung verbunden ist, gilt es studien- und v.a. themenbezogen zu entscheiden.

Tabelle 6: Beispiele für die Ersetzung von Angaben zu Ausbildung und Beruf

Original	Mögliche Ersetzungen	Anmerkung
Ich studiere Sozialpolitik.	Ich studiere [[Angabe des Fachs]]. Ich studiere [[ein sozialwissenschaftliches Fach]]	
An dieser Schule habe ich Abitur gemacht.	An dieser Schule habe ich [[einen Schulabschluss]] gemacht.	
Ich bin Altenpflegerin.	Ich bin [[Nennung des Berufs]]. Ich bin [[Pflegerin/ Pflegekraft]] Ich bin [[Nennung des Berufs, Dienstleistungsberufe und Verkäufer, ISCO-08: Code 05:]] Ich bin [[Nennung des Berufs, Betreuungsberufe im Gesundheitswesen, ISCO08 Code 532:]]. Ich bin [[Nennung des Berufs, Betreuungsberufe im Gesundheitswesen, Haus- und Familienpfleger, ISCO08 Code 5322:]]	Freie Ersetzungen oder auch weitere Bestimmungen des Tätigkeitsbereichs (außerklinisch, stationär, ambulant) können herangezogen werden. Um die Lesbarkeit zu verbessern, kann die Dokumentation des ISCO-Codes auch auf die Anonymisierungsdokumentation begrenzt werden.
Mein Spezialgebiet ist Statistik.	Mein Spezialgebiet ist [[Angabe des Fachgebiets]]. Mein Spezialgebiet ist [[Angabe des mathematischen Fachgebiets]].	
Dann habe ich eine Praxisklasse absolviert.	Dann habe ich [[ein Schuljahr in einem Modellprojekt]] absolviert. Dann habe ich [[ein Schuljahr in einem Modellprojekt in Kooperation mit Betrieben absolviert]] Dann habe ich [[ein Schuljahr in einem Modellprojekt für SchülerInnen mit Hauptschulabschluss und Lern- und Leistungsrückständen]] absolviert.	Das Modell der Praxisklasse wird in einem bestimmten regionalen Gebiet durchgeführt. Ähnliche Infrastrukturprojekte gibt es regional und überregional. Der konkrete Name könnte aber je nach Kontext eine Re-Identifikation ermöglichen, da solche Modelle auf bestimmte unterschiedlich große Regionen begrenzt sind.

³⁶ <http://uis.unesco.org/en/topic/international-standard-classification-education-isced> (Zugriff: 01.12.2021)

5.5 Ersetzung von Organisationen und Institutionen

Auch Informationen zu Organisationen und Institutionen können personenbezogen sein. Zu dieser Gruppe gehören Informationen über Firmen, in denen Menschen arbeiten, Theater und Museen, in die sie gehen, Projekte, für die sie sich engagieren ebenso wie Schulen, Kliniken und Behörden, sofern sie einen Personenbezug herstellen lassen und im Interviewkontext von Bedeutung sind. Eine Übersicht über Beispiele für diesen Themenbereich bietet Tabelle 7. Um ein Überanonymisieren zu vermeiden und das Nachnutzungspotenzial zu erhalten, gilt es auch hier von Fall zu Fall zu entscheiden, ob eine Ersetzung vorgenommen und wenn ja, wie stark abstrahiert werden muss bzw. welche Zusatzinformationen wie Größe, Ausrichtung, besondere Charakteristika für die Sekundärnutzung hilfreich sein können. Um Unternehmen zu kennzeichnen kann z.B. die Größe des Unternehmens, die akkumulierte Mitarbeiterzahl, die nationale oder internationale Ausrichtung von Bedeutung sein (z.B., um ein McDonalds-Restaurant vom kleinen Schnellimbiss um die Ecke abzugrenzen).

Tabelle 7: Beispiele und Ersetzungsvorschläge für Angaben zu Institutionen und Organisationen

Beispiel	Mögliche Ersetzungen	Kommentar
Firmen		
Karstadt	Warenhauskette, Warenhauskette 1, Warenhauskette A	
Bäcker Neumann	Firma 1 (kleines Familienunternehmen), Firma 1 Lebensmittelbranche (kleines Familienunternehmen), ein Lebensmittelgeschäft	Die Regeln für Nummerierung sollten dokumentiert sein, ebenso wie z.B. die Abgrenzung von kleinen, mittleren und großen Unternehmen
Wir sind ca. 25 Mann	Wir sind ein mittelgroßer Betrieb mit 20-30 Mitarbeitern	Betriebsgrößen sollten festgelegt und dokumentiert werden, ebenso wie der Umgang mit gegenderten Formen. Ist z.B. keine Frau beschäftigt, sollte auch nicht von Mitarbeiter*innen gesprochen werden).
Burger King	Systemgastronomie – mehr als 1000 Beschäftigte	Informationen wie z.B. „international agierend“ könnten ergänzt werden.
Zwei-Mann-Betrieb	sehr kleiner Betrieb	
drei Lehrlinge	3-5 Lehrlinge, unter 5 Lehrlinge	Clustern kann eine Option sein.
Schulen/ Ausbildungsstätten		
Volkshochschule	Weiterbildungseinrichtung + Ziffer	
Müller-Schule	Gymnasium + Ziffer, Name der Schule (Gymnasium)	
Kolping	Konfessioneller Sozialverband	
Behörden/Verwaltung		
Handelskammer	Öffentlich-rechtliche Personalkörperschaft im Handel	
Beim Finanzamt	Bei einer Landesbehörde, Landesamt	
Andere		
Radio Bremen	Rundfunkanstalt	
Überseemuseum	Museum	

5.6 Informationen zu Wirtschaftszweigen zur Bestimmung von Berufen und Organisationen

Informationen zu Wirtschaftszweigen sind seltener Teil einer Ersetzung. Viel häufiger ergänzen sie Informationen zu Berufen, beruflichen Tätigkeiten und Organisationen. Um den Wirtschaftszweig in die Ersetzung einfließen zu lassen, kann man einerseits über die im Text benannte Firma recherchieren und auf dieser Basis Informationen bereitstellen. Hilfreich können aber auch hier standardisierte Listen wie Handelsregister oder die Klassifikation der Wirtschaftszweige (WZ) sein. Diese Klassifikation wird regelmäßig vom Statistischen Bundesamt herausgegeben und ist kostenfrei zugänglich.^{37,38} Sie gliedert sich in 21 Unterabschnitte (A-U), die sich wiederum in 88 Abteilungen, 272 Gruppen, 615 Klassen und 839 Unterklassen aufgliedern. Beispiele der Anwendung finden sich in Tabelle 8.

Tabelle 8: Beispiele für die Anwendung der WZ08

Original	Level 1	Level 2	Level 3	Level 4
Siemens	Verarbeitendes Gewerbe (WZ 2008 Abs.C)	Herstellung von Datenverarbeitungsgeräten, elektronischen und optischen Erzeugnissen (WZ08,26)	Herstellung von optischen und fotografischen Instrumenten und Geräten (WZ08,26.70.0)	
Charité	Einrichtung im Gesundheits- und Sozialwesen (WZ08, Abs.Q)	Einrichtung im Gesundheitswesen (WZ08,86)	Krankenhaus (WZ08, 88.1)	Hochschulklinik (WZ08, 86.10.01)

Berichtet ein Interviewpartner im Interview davon, dass er bei der Charité angestellt war, kann also in der anonymisierten Version unterschiedlich stark abstrahiert werden, dass er in einer Institution/ einem Unternehmen etc. im Gesundheits- und Sozialwesen; im Gesundheitswesen, in einem Krankenhaus oder einer Hochschulklinik beschäftigt war. Mit zunehmendem Abstraktionsgrad erhöht sich damit der mitgelieferte Informationsgehalt. Die Ersetzung der Firma Siemens unter Anwendung der WZ2008 könnte im Beispiel wie folgt aussehen:

Frau Bach: „... Wir sind hierhergezogen, weil eben [[Unternehmen B – WZ08,26: Herstellung von Datenverarbeitungsgeräten, elektronischen und optischen Erzeugnissen]] natürlich eine attraktive Arbeitsstelle war. Und das sind dann so 200km Entfernung...“

³⁷<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html?nn=205976> (Zugriff: 01.12.2021)

³⁸ „Die Gliederung der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), wurde unter intensiver Beteiligung von Datennutzern und Datenproduzenten in Verwaltung, Wirtschaft, Forschung und Gesellschaft geschaffen. Sie berücksichtigt die Vorgaben der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft (Nomenclature statistique des activités économiques dans la Communauté européenne (NACE) Revision 2), die mit der Verordnung (Europäische Gemeinschaft (EG)) Nummer 1893/2006 des Europäischen Parlaments und des Rates vom 20. Dezember 2006 (Amtsblatt Europäische Gemeinschaft Nummer Liste 393 Seite 1) veröffentlicht wurde. Die Zustimmung der Europäischen Kommission gemäß Artikel 4, Absatz 3, der oben genannte Verordnung liegt vor.“
(<https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html>) (Zugriff: 01.12.2021).

Auf europäischer und internationaler Ebene gibt es weitere Klassifikationen, die hilfreiche Informationen beinhalten können. Dazu gehört z.B. die ISIC (International Standard Industrial Classification of all Economic Activities)³⁹ der Vereinten Nationen, die Wirtschaftsbereiche und Industriezweige international aufgliedert.

5.7 Ersetzung von Gesundheitsinformationen

Gesundheitsinformationen stellen einen Teilbereich der „individuellen Besonderheiten“ dar (vgl. Tabelle 2), die daneben z.B. auch Informationen über Preise, sportliche Positionen, Unfälle etc. zusammenfasst. In die Verarbeitung von Gesundheitsinformationen muss explizit eingewilligt werden. Sie sind durch Art. 9 DSGVO besonders geschützt.

Sollen Gesundheitsinformationen abstrahiert werden, kann je nach Forschungsthema und Abstraktionsniveau z.B. von einer „Erkrankung“, einer „langjährigen Erkrankung“, einer „Diagnose“ der „ersten oder zweiten Erkrankung“ oder z.B. von einer „Krebserkrankung“ usw. gesprochen werden. Hilfreich bei der Suche nach standardisierten Ersetzungen kann die International Classification of Diseases and Related Health Problems (ICD)⁴⁰ sein. Die ICD wird weltweit eingesetzt, um Diagnosen zu klassifizieren. Sie wird von der Weltgesundheitsorganisation (WHO) herausgegeben und regelmäßig aktualisiert. Die deutschsprachige Version wird vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM)⁴¹ jährlich überarbeitet und angepasst.

Die ICD ist hierarchisch gegliedert und beinhaltet 22 Hauptgruppen von Krankheiten (I-XXII), auf der höchsten Abstraktionsebene. Darunter finden sich z.B. Endokrine, Ernährungs- und Stoffwechselkrankheiten (Gruppe IV), Krankheiten des Kreislaufsystems (Gruppe IX) oder auch Krankheiten der Haut und der Unterhaut (Gruppe XII). Diese Hauptgruppen verzweigen sich in immer differenziertere Untergruppen. Das Hierarchisierungsniveau lässt sich aus der jeweils zugehörigen ICD-Codenummer ablesen. Der ICD-Code lässt sich bis auf fünf Ebenen auffächern, die unterschiedlich stark spezifizierte Informationen zu Krankheiten bereitstellen. Ein Herzinfarkt eines Interviewpartners könnte beispielsweise als „Erkrankung“, als „Krankheit des Kreislaufsystems (Gruppe IX, ICD10)“ oder als „ischämische Herzkrankheit (I20 ICD10)“ ersetzt werden. Von Abstraktionsebene zu Abstraktionsebene erhöht sich die Zahl der Menschen, die am genannten Gesundheitsproblem leiden, was den konkreten Personenbezug verringert.

³⁹ <https://ilostat ilo org/resources/concepts-and-definitions/classification-economic-activities/> (Zugriff: 01.12.2021)

⁴⁰ <https://www who int/classifications/classification-of-diseases> (Zugriff: 01.12.2022)

⁴¹ Ursprünglich oblag die Bearbeitung der deutschen Version der ICD dem Deutschen Institut für Medizinische Dokumentation und Information (DIMDI). Dieses wurde am 26. Mai 2020 mit dem Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) unter dem Dach des BfArM zu einer Behörde zusammengeführt.

Kapitel IX
Krankheiten des Kreislaufsystems
(100-199)

Dieses Kapitel gliedert sich in folgende Gruppen:

<u>100-102</u>	Akutes rheumatisches Fieber
<u>105-109</u>	Chronische rheumatische Herzkrankheiten
<u>110-115</u>	Hypertonie [Hochdruckkrankheit]
<u>120-125</u>	Ischämische Herzkrankheiten
<u>126-128</u>	Pulmonale Herzkrankheit und Krankheiten des Lungenkreislaufes
<u>130-152</u>	Sonstige Formen der Herzkrankheit
<u>160-169</u>	Zerebrovaskuläre Krankheiten
<u>170-179</u>	Krankheiten der Arterien, Arteriolen und Kapillaren
<u>180-189</u>	Krankheiten der Venen, der Lymphgefäße und der Lymphknoten, anderenorts nicht klassifiziert
<u>195-199</u>	Sonstige und nicht näher bezeichnete Krankheiten des Kreislaufsystems

Abbildung 3: ICD10, Aufgliederung der Herzkreislauferkrankungen Gruppe IX in Unterkategorien auf der zweiten Abstraktionsebene⁴²

Unter Beachtung der Einwilligungserklärung und bereits vorgenommener Ersetzungen gilt es genau abzuwägen, was ersetzt werden muss, um einen Re-Identifikation zu verhindern. Der Herzinfarkt könnte je nach Studienkontext als Einzeldatum auch unspezifisch genug sein und nicht notwendigerweise abstrahiert werden müssen. Beispiele für dieses Abwägen des Abstraktionsgrades von Ersetzungen, beschreiben z.B. Saunders et al. (2015a) in ihren Ausführungen zur Anonymisierung der Forschungsmaterialien einer Studie mit Angehörigen von chronisch erkrankten Menschen, die in der Wahrnehmung ihrer selbst und ihrer Umwelt stark eingeschränkt waren:

„Over and above name, place, occupation and religion there are often features in any dataset, which may provide unique (or closely narrowing) identifying information. We have given examples of a range of such features throughout the discussion above – including idiosyncratic details of a case. Depending on the substantive focus of the research there may also be a recurring issue that needs special attention. In our case this was the nature of the original injury. Such information was often interwoven throughout accounts and also has prognostic significance. Brain damage resulting from ‘non-traumatic’ injury (e.g. from oxygen deprivation following cardiac arrest) is more predictably devastating than resulting from traumatic brain injury (e.g. from a blow to the head). Distinguishing between these types of injury is therefore analytically important. However, beyond this we commonly used generic phrases such as ‘road traffic accident’ or ‘sporting injury’, which allowed us to describe a range of events while not revealing more specific information.“ (Saunders et al. 2015a, 627)

Neben der ICD bietet die WHO weitere Klassifikationen an, die für die Abstraktion personenbezogener Gesundheitsinformationen in spezifischen Forschungsprojekten interessant sein können. Dazu gehören z.B. die International Classification of Functioning, Disability and Health (ICF)⁴³ (dts: Internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit) oder die International Classification of Health Interventions

⁴² zugänglich unter <https://www.praxis-wiesbaden.de/icd10-gm-diagnosen/#IX> (Zugriff am 23.06.2022)

⁴³ <https://www.who.int/classifications/icf/en/> (Zugriff: 01.12.2021)

(ICHI).⁴⁴ Die ICD kann in das Qualiservice Anonymisierungstool eingelesen werden. Werden Listen für Ersetzungen verwendet, gilt es die listenspezifischen Nutzungsbedingungen zu befolgen und die jeweilige Version der Klassifikation zu dokumentieren und zu zitieren (vgl. Punkt 4, Schritt 5).

5.8 Ersetzung von Positionen und Rollen

Personen können auch über Angaben zu Positionen und Rollen in familiären oder in Arbeitszusammenhängen, bzgl. politischer Gremien oder ehrenamtlichen Engagements oder auch durch Platzierungen in sportlichen Wettbewerben identifiziert werden. Eine Auswahl an Beispielen findet sich in Tabelle 9.

Wieder muss geprüft werden, ob im Text erhaltene Informationen einen Personenbezug herstellen lassen oder ob z.B. schon die Ersetzung anderer Informationen im Text eine Re-Identifikation nahezu ausschließt, so dass evtl. die Originalinformation erhalten bleiben kann.

⁴⁴ <https://www.who.int/classifications/ichi/en/> (Zugriff: 01.12.2021)

Tabelle 9: Beispiele für die Ersetzung von Informationen zu Positionen und Rollen

Original	Mögliche Ersetzungen	Anmerkung
Er war erster Vorsitzender im hiesigen Schützenverein.	Er [[hatte eine hohe Position]] im hiesigen Schützenverein. Er war erster Vorsitzender im hiesigen [[Verein]].	Sind z.B. Erhebungsorte bekannt, könnte auch der Verweis auf den „hiesigen“ Schützenverein zu ersetzen sein. Ersetzt man Spezifika des Vereins kann evtl. die Funktion des Vorsitzenden erhalten bleiben (siehe zweites Beispiel).
Mein Vater war Schmiedemeister.	Mein Vater [[hatte Meisterfunktion]]. Mein Vater war [[Handwerksmeister]].	
	Mein Vater war [[Handwerksmeister im verarbeitenden Gewerbe (WZ2008 Abs. C)]] Mein Vater war [[Handwerksmeister in der Herstellung von Metallerzeugnissen (WZ 2008 Abs. C, 25)]]	Informationen zu Wirtschaftszweigen können verwendet werden, um unterschiedlich stark abstrahierte sozialwissenschaftlich relevante Information zu erhalten (vgl. Punkt 5.6).
In diesem Jahr wurde ich Referentin der Senatorin für Bildung.	In diesem Jahr wurde ich Referentin [[in der öffentlichen Verwaltung des Landes]]. In diesem Jahr wurde ich Referentin der Senatorin [[xx (nähere Bestimmung des Senatsressorts)]].	Auch Positionen anderer können zur Re-Identifikation einer Person beitragen. Im Beispiel wird die „Senatorin für Bildung“ ersetzt und die Position der Referentin erhalten. Diese Entscheidung kann je nach Kontext der Studie unterschiedlich ausfallen und auch dazu führen, dass z.B. die Referendarstelle abstrahiert wird. Im zweiten Beispiel werden Informationen weggelassen. Die Löschung der beiden Wörter ist mit jeweils einem X für ein Wort markiert.
Er aß privat regelmäßig mit Helmut Kohl zu Mittag.	Er aß privat regelmäßig mit [[einem deutschen hochrangigen Politiker]] zu Mittag. Er aß privat regelmäßig mit [[einem deutschen Bundeskanzler]] zu Mittag	Die Ersetzung von Namen von in der Öffentlichkeit stehenden Personen wie der Person von Helmut Kohl ist ein besonderes Thema (siehe Punkt 5.1). Im Beispiel wird angenommen, dass die Nennung Helmut Kohls mit einem Re-Identifikationsrisiko für den oder die Forschungsteilnehmer/in einhergeht.

5.9 Weitere Informationen, die einen Personenbezug enthalten können

Je nach Kontext des Textes und der Forschungsfragen können neben den genannten Ersetzungskategorien auch Informationen aus anderen Bereichen wie z.B. Angaben zu Sprachen, Nationalitäten, Konfessionen oder Sportarten ggf. die Re-Identifikation einer Person ermöglichen. Auch hier gilt es zunächst zu prüfen, ob und wenn ja, wie viel der Information ersetzt werden muss, um ein Über- oder Unteranonymisieren zu vermeiden. Insbesondere gilt es Informationen zur Religionszugehörigkeit oder zur sexuellen Orientierung, die wie auch Gesundheitsinformationen von der DSGVO besonders geschützt sind, zu prüfen. Die Abwägung von notwendigen Schutzmaßnahmen und wissenschaftlichem Erkenntnisinteresse

ist bei der Abstraktion von Informationen ständiger Begleiter gut ablesbar in den Beschreibungen von Saunders et al. (2015a, 624):

„We compromised through cutting references which would narrow down identification of their religion/cultural background too tightly – this involved losing some very interesting data but seemed a reasonable compromise, albeit one which made the quotations much more anodyne than they had been in the original form. The question we ask ourselves here is whether we have veered too far, and ‘white-washed’ the data, ‘forfeiting much of the richness yielded by the study’ (Parry and Mauthner, 2004: 144) and draining it of meaning?“

Schwierigkeiten treten auf, wenn Informationen zur Religion oder zu religiösen Handlungen für die Forschungsfrage wichtige Informationen enthalten. Saunders et al. (2015a, 623) beschreiben z.B. wie in einer religiösen Handlung deutlich wird, dass die Bewusstseinsleistung eines am Hirn verletzten Patienten größer ist als von Ärzten eingeschätzt:

„...we often replaced religion/culture with similar but unrelated items, or with generalised descriptions (e.g. Jaspal’s Hinduism became ‘Jaspal’s religious faith’). This strategy became problematic when religion, culture or ethnic identity was not simply ‘background information’ but crucial ‘context for deeper and fuller understanding’ (Clark, 2006: 6). After lengthy discussion we decided to retain certain features in some cases. For example George’s brother, David, had been injured six years prior to the interview, and at the time was diagnosed by medical professionals as having ‘low awareness’. George and his wife, Linda, (interviewed together) were very optimistic about David’s situation. Their religion underpinned their shared views about David’s future, along with a strong sense of family togetherness that they associated with their cultural heritage (from outside the UK). They consistently represented these factors as underpinning their belief in David’s ongoing recovery...“

Zum Teil gibt es auch für diese Kategorie listenartige Übersichten, die bei der Suche nach passenden Ersetzungen helfen können. Diese müssen aber im Einzelfall auf ihre Eignung (fachliche Qualität, Zugänglichkeit und Aktualisierung) geprüft werden. Hilfreiche Informationen zu Sprachen könnte z.B. die Website „Ethnologue language name index“⁴⁵ bieten. Daneben bieten offizielle Seiten wie das Statistische Bundesamt, die UNO z.B. zu Ländern, deren Bevölkerung und ihren regionalen Profilen⁴⁶ oder die WHO⁴⁷ oder EUROSTAT⁴⁸ Informationen, die die Suche nach passenden Abstraktionen im Sinne sozialwissenschaftlich relevanter Informationen unterstützen könnten.

5.10 Kumulation von personenbezogenen Informationen verschiedener Kategorien

Ziel der hier beschriebenen Vorgehensweisen ist es, datenschutzrechtlichen Aspekten nachzukommen und dabei das Nachnutzungspotenzial eines Datensatzes zu erhalten. Wir haben einzelne Ersetzungskategorien beschrieben, die häufig personenbezogene Informationen enthalten und empfehlen ein Vorgehen, welches alle potenziell zu ersetzenden

⁴⁵ <https://www.ethnologue.com/data-consulting> (Zugriff: 01.12.2021)

⁴⁶ <http://data.un.org/> (Zugriff: 01.12.2021)

⁴⁷ <https://www.who.int/data/collectionsweitere> (Zugriff: 01.12.2021)

⁴⁸ <https://ec.europa.eu/eurostat/web/main/data/databaseInformationen> (Zugriff: 01.12.2021)

Informationen im Einzelnen auf ihren Personenbezug hin prüft. Bei der Entscheidung, ob ein Personenbezug enthalten ist oder nicht, sollte berücksichtigt werden, dass vorgenommene Ersetzungen einer Kategorie den Personenbezug von Informationen anderer Ersetzungskategorien beeinflussen. Sind z.B. neben direkten Identifikatoren wie Namen und Adressen auch Orts- und Zeitangaben im Text abstrahiert, kann dies ggf. für eine Anonymisierung eines Textes bereits ausreichen. Umgekehrt, kann die Nennung des Erhebungsortes in Publikationen und Kontextmaterialien dazu führen, dass im Interviewtext viel umfassender eingegriffen werden muss, um dieses Ziel zu erreichen.

Daneben beeinflusst auch das Studiendesign die Entscheidung für ein passendes Anonymisierungskonzept. Bei der Archivierung von Panelstudien, die über mehrere Jahre hinweg dieselben Personen beforschen, soll z.B. eine konsistente Anonymisierung über alle Wellen hinweg sichergestellt werden. Hier kann es ggf. helfen – anders als beschrieben – bestimmte Informationen wie z.B. Ortsbezeichnungen konsistent über die Studie hinweg von Beginn auch ohne Prüfung des jeweiligen Personenbezugs standardisiert in derselben Weise zu ersetzen, um eine konsistente Handhabung der Ortsinformation zu erleichtern. D.h., ohne zu prüfen, ob z.B. die Nennung des Ortes Kappeln an der Schlei an den verschiedenen Stellen der Materialien einen Personenbezug beinhaltet (z.B. Geburtsort oder eben nur Urlaubsort ist) würden alle Nennungen von Kappeln ersetzt, um auf pragmatische Weise sicher zu stellen, dass Ort 1, kleine Mittelstadt in Wassernähe mit viel Landwirtschaft aus Welle 1 dem in Welle 5 entspricht. Man würde hier auf den möglichen Erhalt von Originalinformation zugunsten eines konsistenten und über mehrere Wellen stabil zu haltenden Anonymisierungskonzeptes verzichten. Diese Vorgehensweise hätte den Vorteil, dass evtl. aufwendige Nachbearbeitungen von Interviews aus vorangegangenen Wellen reduziert werden könnten, die durch Bezüge zwischen Interviews derselben Personen und dem damit evtl. verbundenen Reidentifikationsrisiko notwendig werden könnten.

6 QualiAnon - Das Qualiservice Anonymisierungstool für textbasierte Forschungsmaterialien

Personenbezogene Informationen in textgebundenen Forschungsmaterialien zu abstrahieren, ist ein ressourcenaufwändiger Prozess. Das gilt insbesondere dann, wenn die Materialien für die wissenschaftliche Nachnutzung aufbereitet werden. Schon vor Jahren begann Qualiservice deshalb mit der Entwicklung und Anwendung eines Anonymisierungstools, mit dessen Hilfe Ersetzungen fall- oder studienbezogen vorgenommen und bearbeitet werden können (vgl. Kretzer, 2013) und welches sicher stellt, dass Identifikationsdaten getrennt von anderen Informationen wie z.B. von anonymisierten Interviewtranskripten gespeichert werden können.

QualiAnon ist so konzipiert, dass Anwender/-innen jederzeit die Kontrolle über die von der Software ausgeführten Prozesse und bearbeiteten Informationen haben. Das Tool läuft ohne Anbindung an das Internet. Auf automatisch ablaufende Prozesse wurde bewusst verzichtet. D.h., QualiAnon unterstützt den Ersetzungsprozess, es nimmt aber keine Ersetzungen

automatisch vor. Das kritische sinnverstehende Lesen und Einordnen von Informationen als sensibel muss von den Anwendern/-innen vorgenommen werden.

Bei der Entwicklung spielten u.a. die folgenden Überlegungen eine besondere Rolle: Das Anonymisierungstool sollte sensible Informationen markieren und – unter Verwendung verschiedener Abstraktionsniveaus - ersetzen können. Die Art der Ersetzung sollte dokumentierbar und bearbeitbar sein und sofern gewollt auch die Originalinformation erhalten werden (Pseudonymisierung).

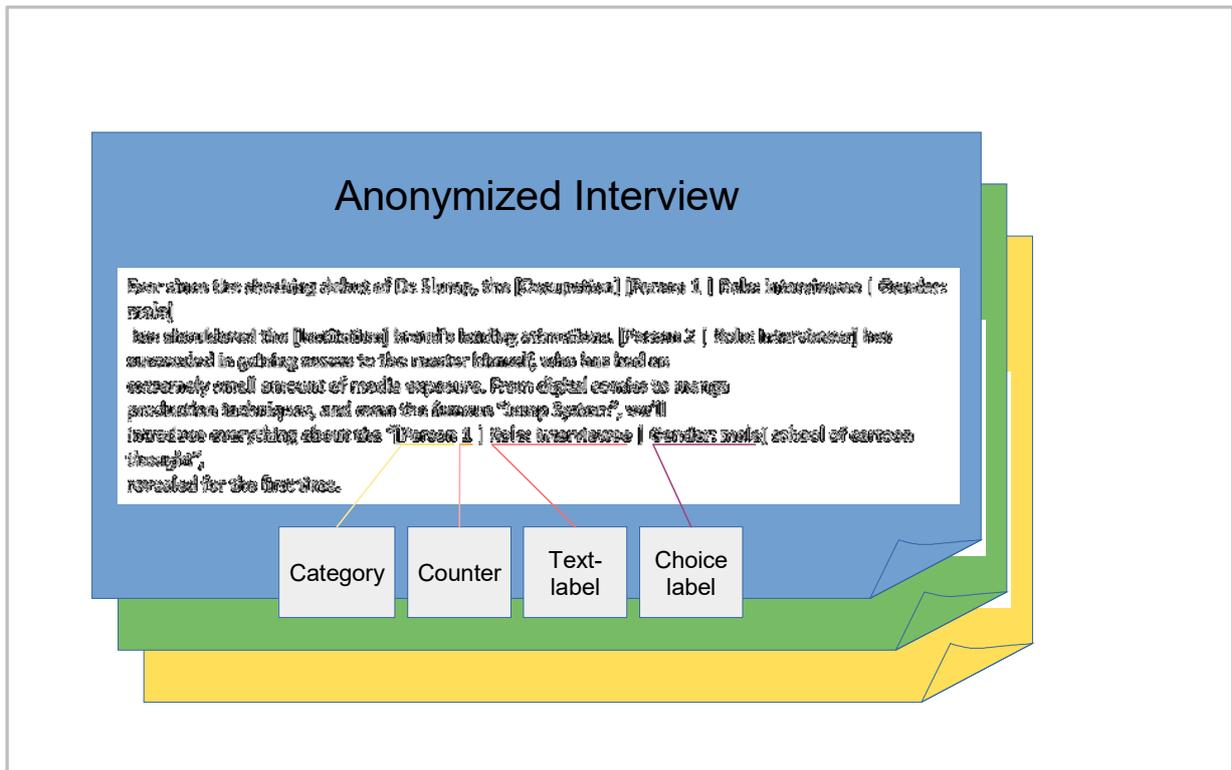


Abbildung 4: Anonymisierungsbeispiel für die Anwendung von QualiAnon (Darstellung: Tom Nicolai)

Abbildung 4 zeigt ein Beispiel für die Anonymisierung eines Namens in einem Interviewausschnitt mit QualiAnon. Für jede Entität, die QualiAnon ersetzt, kann aus einem bearbeitbaren System von Ersetzungskategorien, eine Kategorie gewählt werden (z.B. Person, Beruf, Ort, etc.). In eckigen Klammern findet sich die Information, die im Beispiel zur Ersetzung des Namens verwendet wurde: „[Person 1 | Role: Interviewee | Gender: male]“.

Die verwandten Ersetzungskategorien können für alle Interviews einer Studie durchnummeriert werden, so dass Person 1 im gesamten Interview oder auch in weiteren Interviews als dieselbe Person erkennbar ist. Darüber hinaus können je nach Sensibilität der Information und angestrebter Sicherheitskategorie unterschiedliche Abstraktionsebenen der Ersetzung gewählt werden, was die Umsetzung des Konzeptes der „flexiblen Anonymisierung“ (vgl. Punkt 3.4) ermöglicht. Im Beispiel sind es die Information, dass es sich bei Person 1 um den Interviewten handelt und die Information, dass der Interviewte männlichen Geschlechts

ist. Die bei der Ersetzung entstehenden Kategoriensysteme können exportiert und mit anderen Forschern/-innen geteilt werden (vgl. das zum Tool gehörige Manual (Nicolai, 2021)). Auf diese Weise können themen- oder bereichsspezifische Ersetzungsschemata entstehen, die Anwender/-innen, die zu ähnlichen Themen arbeiten, qualifizierte Anhaltspunkte oder auch ausgearbeitete Anonymisierungskonzepte für das eigene Projekt liefern.

Damit Forscher/-innen auch außerhalb der besonders geschützten Kurationsumgebung eines FDZ QualiAnon nutzen können, waren verschiedene Weiterentwicklungen nötig. Um einen datenschutzrechtlich konformen Umgang mit den Forschungsdaten sicher zu stellen, sollten Nutzer z.B. Identifikationsdaten getrennt von anderen Informationen abgelegt können.

Das Tool arbeitet mit einem System von Universally Unique Identifiers (UUIDs) (vgl. Abbildung 5). Das anonymisierte Interview („blackened interview“) enthält die Marker-ID, die bei der Bearbeitung im Originaltext gesetzt wird. Die Replacement-ID bezieht sich auf die jeweils gewählte Übersetzung. In einem weiteren Dokument werden Marker-IDs und Replacement-IDs einer Studie sowie die Originalinformation gespeichert.⁴⁹

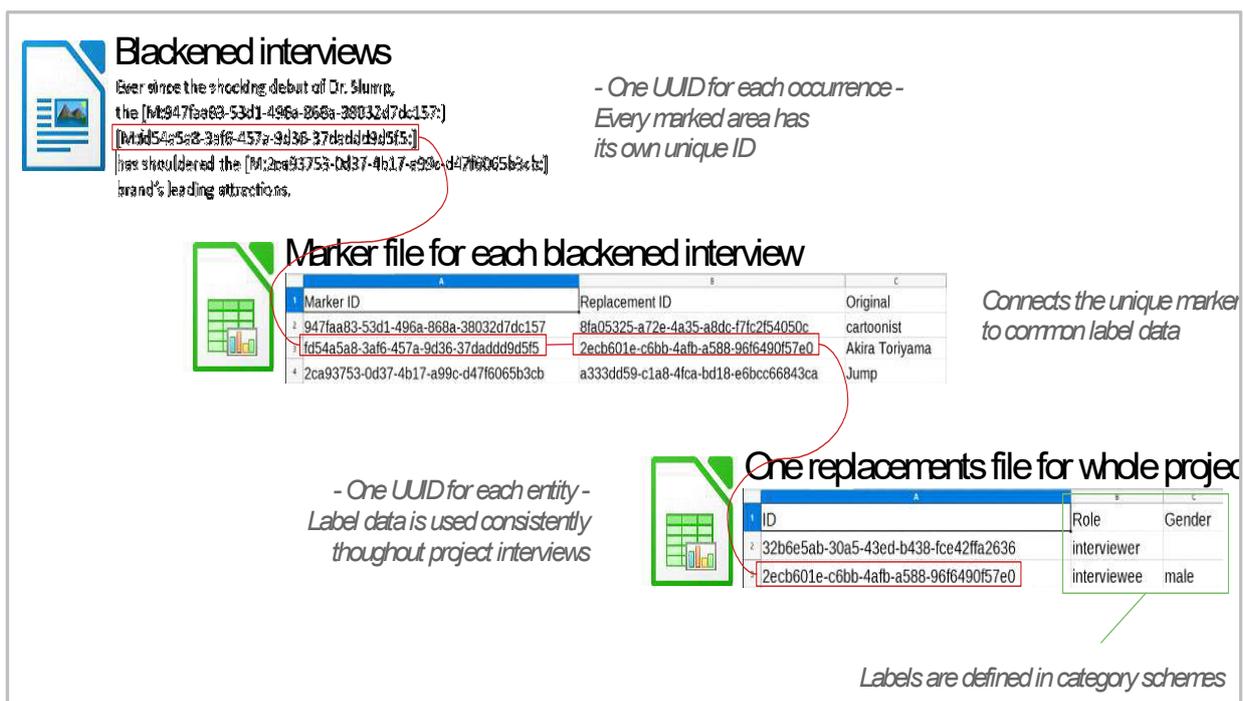


Abbildung 5: Das Datenmodell von QualiAnon - Verbindung von UUIDs (Darstellung Tom Nicolai)

⁴⁹ Sollen Daten anonymisiert werden, muss die Originalinformation aus diesem Dokument gelöscht werden.

Das von Nicolai et al. (2021) entwickelte Tool QualiAnon wurde im Mai 2021 im virtuellen Launch vorgestellt und wird kontinuierlich weiterentwickelt. Abbildung 6 zeigt wie die einzelnen Bausteine des Programms während der Bearbeitung eines Projektes organisiert werden können.

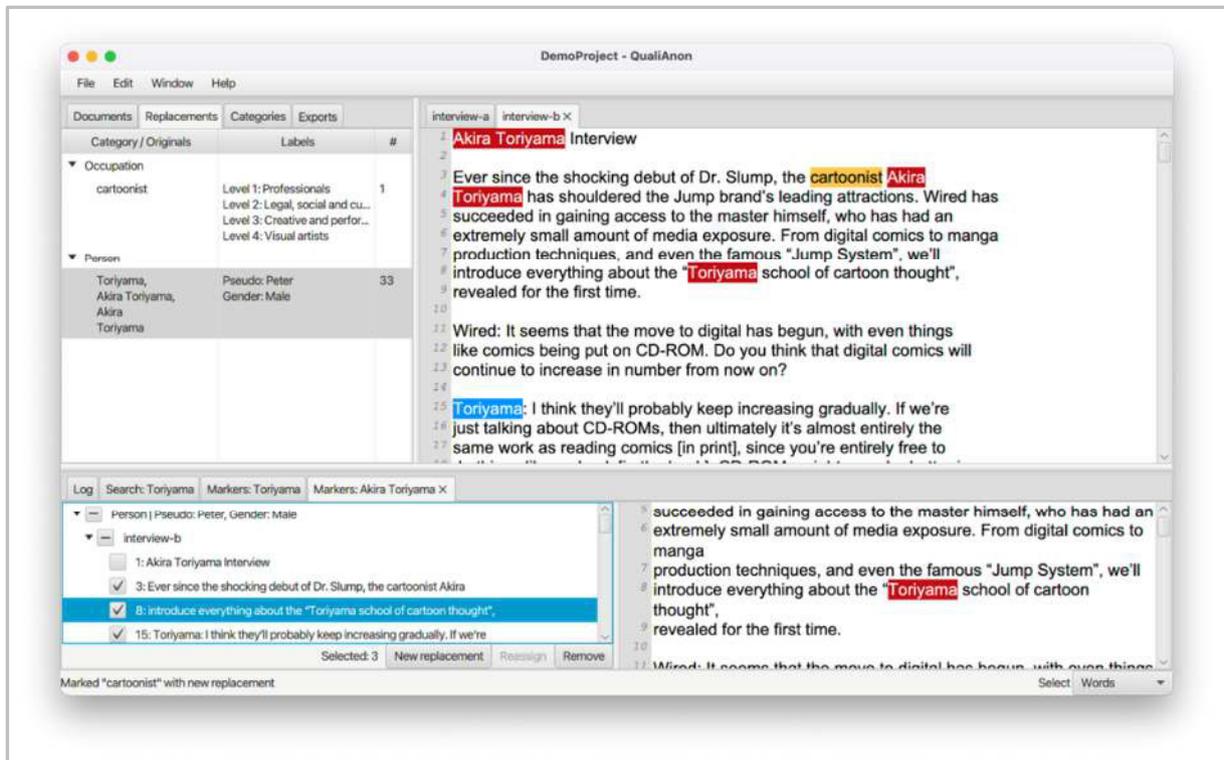


Abbildung 6: Bildschirmfoto Programmanwendung (Tom Nicolai)

QualiAnon ist als Open Source-Tool entwickelt. Der Code wird über ein Github-Repository⁵⁰ veröffentlicht. Bis zur Veröffentlichung des Codes erhält man kostenfrei Zugang über Qualiservice unter der Adresse qualianon@uni-bremen.de.⁵¹ Die einzelnen Anwendungsschritte können im detaillierten Anwendungsmanual (Nicolai, 2023) nachvollzogen werden.

⁵⁰ <https://github.com/> (Zugriff: 01.12.2021)

⁵¹ Sollten Sie Ihre Daten für die Sekundärnutzung bei Qualiservice aufbereiten wollen, kontaktieren Sie uns bitte immer vorab.

7 Zusammenfassung und Ausblick

Die Anonymisierung von personenbezogenen Daten ist eine Strategie um datenschutzrechtliche und forschungsethische Anforderungen und damit auch Anforderungen an die gute wissenschaftliche Praxis zu erfüllen (DFG, 2022). Anders als in der quantitativen Sozialforschung fehlen in der qualitativen Sozialforschung etablierte Anonymisierungsverfahren, was auch mit der Beschaffenheit der Forschungsmaterialien, ihrer besonderen Informations- und Kontextdichte sowie dem hohen Grad des Aufeinanderverwiesenseins der Informationen liegt. Eine weitere Herausforderung für die Anonymisierung von qualitativen Textdaten liegt darin, dass die Nähe zum Ursprungstext soweit wie möglich erhalten werden muss, um qualitative Forschungsmaterialien wissenschaftlich nachnutzen zu können. Schritt für Schritt, von Fall zu Fall und oft auch bezogen auf die Nennung ein und derselben Information im Text muss entschieden werden, ob Informationen in der jeweiligen Verwendung im Text ein Re-Identifikationsrisiko bergen und ersetzt werden müssen und welche Informationen im Originalwortlaut erhalten werden können. Gleichzeitig gilt es nach Möglichkeiten zu suchen, die Lesbarkeit des Textes zu erhalten.

Diese besonderen Herausforderungen sprechen gegen die Realisierung eines standardisierten Anonymisierungsverfahrens in der qualitativen Sozialforschung. Gleichzeitig ist es sinnvoll aus Gründen begrenzter Ressourcen und mit Blick auf den Erhalt des Nachnutzungspotenzials auch in der qualitativen Sozialforschung (themen- und disziplinspezifische) Best Practices zu etablieren, die es ermöglichen, Strategien weiter zu entwickeln und die Vor- und Nachteile bestimmter Vorgehensweisen festzuhalten. Die Verwendung von autorisierten, d.h. regelmäßig aktualisierten und von öffentlichen Institutionen, erstellten Listen bei der Ersetzung von z.B. Orten oder Krankheiten kann hier ebenso wie die Weiterentwicklung und das Teilen themenspezifischer Schemata von Ersetzungskategorien helfen, themenspezifische Teilstandardisierungen zu entwickeln. Allerdings muss im Einzelfall abgewogen werden, ob die Verwendung standardisierter Listen einen Mehrwert für die Erhaltung sozialwissenschaftlicher Information bringt oder ob nicht ganz andere Informationen (die nicht standardisiert erfasst sind) für die Nachnutzung bedeutsam sind.

Um die Anonymisierung qualitativer Forschungsdaten zu planen und den notwendigen Aufwand abzuschätzen, empfiehlt es sich zunächst ein Anonymisierungskonzept zu erstellen und zu dokumentieren. Der Zweck, für den anonymisiert wird, ist dabei für die Schätzung des notwendigerweise zu betreibenden Aufwandes zentral – ein bebildertes Zitat in einem wissenschaftlichen Fachjournal ist weit weniger aufwendig zu anonymisieren als komplette Interviewtranskripte, die für die Nachnutzung aufbereitet werden. Darüber hinaus stellt sich auch die Frage, für wen anonymisiert werden soll: Sollen die anonymisierten Forschungsmaterialien öffentlich zugänglich sein, sodass im Zweifel auch Freunde/-innen oder Kollegen/-innen von Befragten die Materialien einsehen können, oder wird der Zugang wie bei Qualiservice auf Wissenschaftler/-innen, die eine Nutzungsvereinbarung unterschreiben, reglementiert sein? Auch Fragen der Über- und Unteranonymisierung stellen sich bezogen auf den jeweiligen Verwendungszweck unterschiedlich und müssen vor dem Hintergrund der

geplanten Art der Nachnutzung beurteilt werden. Die ausführliche Dokumentation des gewählten Anonymisierungsprozederes hilft, ein konsistentes Vorgehen im Forschungsteam sicherzustellen und ermöglicht Sekundärnutzern/-innen den Nachvollzug der am Text vorgenommenen Veränderungen.

Abschließend gilt es festzuhalten, dass absolute Anonymität in einer digitalisierten Welt wie unserer bei gleichzeitigem Erhalt des wissenschaftlichen Nachnutzungswertes für qualitative Forschungsmaterialien nicht realisierbar ist. Was aber realisierbar ist, ist durch die Möglichkeiten, die die faktische Anonymisierung und auch die Pseudonymisierung bieten in Verbindung mit weiteren organisatorisch-technischen Maßnahmen wie Verschlüsselung, Passwortschutz, Zugangsbeschränkungen, getrennte Speicherung von Identifikatoren und Forschungsmaterialien, onsite-Nutzung etc. das Risiko für eine mögliche Re-Identifikation von Personen soweit zu minimieren, dass die häufig unter hohem Ressourcenaufwand erhobenen und in der Regel nicht in ihrer kompletten Tiefe ausgewerteten qualitativen Daten weiter beforscht werden können, sie bei der Vorbereitung von Anträgen verwendet und in der akademischen Lehre genutzt werden können. Für diesen nachhaltigen Ansatz arbeitet Qualiservice zusammen mit Forscher/-innen an innovativen Lösungen, die praktikabel sind und genügend Raum für die studien- und fallspezifische Flexibilität lassen. Wir hoffen mit unserer Handreichung Forschende bei der Entwicklung eines passenden Vorgehens unterstützen zu können, stehen für Rückfragen gern zur Verfügung und freuen uns über Anregungen, Erfahrungsberichte und Feedback.

8 Quellenverzeichnis

- Bishop, L. (2005). Protecting Respondents and Enabling Data Sharing: Reply to Parry and Mauthner. *Sociology*, 39(2), 333-336.
- Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and Informed Consent: Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives. *FQS*, 1(3), Art. 7.
- Deutsche Forschungsgemeinschaft (DFG). (2019). *Handreichung des Fachkollegiums 106 Sozial- und Kulturanthropologie, Außereuropäische Kulturen, Judaistik und Religionswissenschaft zum Umgang mit Forschungsdaten*.
https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/handreichung_fachkollegium_106_forschungsdaten.pdf
- Deutsche Forschungsgemeinschaft (DFG). (2022). Leitlinien zur Sicherung guter wissenschaftlicher Praxis.
- Deutsche Gesellschaft für Erziehungswissenschaft (DGfE). (2005). *Anonymisierung von Daten in der qualitativen Forschung: Probleme und Empfehlungen*
https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2005.11_Anonymisierung_von_Daten.pdf
- Deutsche Gesellschaft für Erziehungswissenschaft (DGfE). (2016). Ethikkodex der Deutschen Gesellschaft für Erziehungswissenschaft. *Erziehungswissenschaft*, 53 (27).
- Deutsche Gesellschaft für Soziologie (DGS). (2017). *Ethikkodex der deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS)*.
<https://www.soziologie.de/die-dgs/ethik/ethik-kodex/> (Zugriff 25.10.2019)
- Deutsche Gesellschaft für Volkskunde (DGV) (Ed.). (2018). *Positionspapier zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten*. https://www.d-g-v.de/wp-content/uploads/2019/04/dgv-Positionspapier_FDM.pdf
- Ebel, T. (o. J.). *Empfehlungen zur Anonymisierung quantitativer Daten*
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjy3_D4pKnqAhVQwKQKHailAMkQFjABegQIBRAB&url=https%3A%2F%2Fwww.gesis.org%2Ffileadmin%2Fupload%2Finstitut%2Fwiss_arbeitsbereiche%2Fdatenarchiv_analyse%2FAnonymisierung_quantitativer_Daten-0150512.pdf&usq=AOvVaw1iDH8hMUXt4614vUc851CS (Zugriff: 01.08.2023)
- Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung) Version des ABl. L 119, 04.05.2016; ber. ABl. L 127, 23.05.2018, (2018). <https://dsgvo-gesetz.de/> (Zugriff: 01.08.2023)
- Fielding, N. (2004). Getting the most from archived qualitative data: epistemological, practical and professional obstacles. *Int J Social Research Methodology*, 7(1), 97-104.
- Finnish Social Science Data Archive. (2021). *Anonymization and Personal Data*.
<https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/>
- Gerhards, J. (2005). *The Name Game*. Transaction Publishers.
- Gerhards, J. (2010). *Die Moderne und ihre Vornamen*. VS Verlag für Sozialwissenschaften.
https://doi.org/https://doi.org/10.1007/978-3-531-92353-6_9
- Heuer, J.-O., Hollstein, B., & Mozygemba, K. (2021). Data Sharing mit Qualiservice. *Soziologie*, 50 (4), 459-471.
- Heuer, J.-O., Kretzer, S., Mozygemba, K., Huber, E., & Hollstein, B. (2020). *Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung. Eine Handreichung für Forschende zur Erstellung eines Studienreports für das Forschungsdatenzentrum Qualiservice*.
<https://doi.org/http://dx.doi.org/10.26092/elib/166>
- Hopf, C. (2004). Forschungsethik und qualitative Forschung. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *Qualitative Forschung. Ein Handbuch* (pp. 589-600). Rowohlt.

- Huber, E., & Imeri, S. (2021). Informed consent in ethnographic research: A common practice facing new challenges. In *Qualiservice Working Papers*. Bremen: Forschungsdatenzentrum Qualiservice. <https://doi.org/10.26092/elib/1070>
- Kelly, A. (2009). In defence of anonymity: rejoining the criticism. *British Educational Research Journal*, 35(3), 431-445.
- Kretzer, S. (2013). Arbeitspapier zur Konzeptentwicklung der Anonymisierungs-/Pseudonymisierung in Qualiservice. https://wiso.genios.de/document/SSOA__47605
- Kretzer, S., Mozygamba, K., Heuer, J.-O., & Huber, E. (2020). Erläuterungen zur Verwendung der von Qualiservice bereitgestellten Vorlagen für die informierte Einwilligung. *Qualiservice Working Papers*, QS-WP-2-2020. <https://doi.org/http://dx.doi.org/10.26092/elib/192>
- Laudel, G., & Bielick, J. (2019). Forschungspraktische Probleme bei der Archivierung von leitfadengestützten Interviews. *Forum Qualitative Sozialforschung*, 20 (2), 1-32. <https://doi.org/https://doi.org/10.17169/fqs-20.2.3077>
- Medjedovic, I. (2011). Secondary Analysis of Qualitative Interview Data: Objections and Experiences. Results of a German Feasibility Study. *FQS*, 12(3), Art 10. <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103104>
- Medjedovic, I. (2014). *Qualitative Sekundäranalyse. Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung*. Springer VS.
- Meyermann, A., & Porzelt, M. (2014). *Hinweise zur Anonymisierung qualitativer Daten* (Vol. 1). Deutsches Institut für Internationale Pädagogische Forschung. <http://www.forschungsdatenbildung.de/files/fdb-informiert-nr-1.pdf>
- Mozygamba, K., & Kretzer, S. (2022). Datenvielfalt im Data-Sharing – eine kooperative Aufgabe von Forschenden und Forschungsdatenzentrum. In C. Lohmeier & T. Wiedemann (Eds.), *Datenvielfalt in kommunikationswissenschaftlichen Forschungskontexten*. Springer VS.
- Nicolai, Tom & Mozygamba, Kati (2023): QualiAnon User Manual, v1.3 Qualiservice Technical Report 1-2023, Bremen, doi: <https://doi.org/10.26092/elib/2576>
- Nicolai, T., Mozygamba, K., Kretzer, S., & Hollstein, B. (2021). *QualiAnon - Qualiservice Tool for Anonymizing Text Data*. In University of Bremen.
- Opitz, D., & Mauer, R. (2005). Erfahrungen mit der Sekundärnutzung von qualitativem Datenmaterial – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten. *FQS*, 6, Art. 13.
- Rat für Sozial- und Wirtschaftsdaten (RatSWD). (2017a). *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften*. Bundesministerium für Bildung und Forschung (BMBF), https://www.ratswd.de/dl/RatSWD_Output9_Forschungsethik.pdf
- Rat für Sozial- und Wirtschaftsdaten (RatSWD). (2017b). *Handreichung Datenschutz*. Bundesministerium für Bildung und Forschung. https://www.ratswd.de/dl/RatSWD_Output5_HandreichungDatenschutz.pdf
- Rat für Sozial- und Wirtschaftsdaten (RatSWD). (2020). *Handreichung Datenschutz* (2. vollständig überarbeitete Auflage. ed., Vol. Output 8). Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/https://doi.org/10.17620/02671.50>
- Richter, C., Kwelik, N., Müller, M., & Severing, L. (2021). Qualitative Daten anonymisieren und für Sekundäranalysen aufbereiten: Das Bochumer Anonymisierungsmodell (BAM). In C. Richter & K. Mojescik (Eds.), *Qualitative Sekundäranalysen. Daten der Sozialforschung aufbereiten und nachnutzen*. Springer VS.
- Saunders, B., Kitzinger, J., & Kitzinger, C. (2015a). Anonymising interview data: challenges and compromise in practice. *Qualitative Research*, 15(5), 616-632. <https://doi.org/10.1177/1468794114550439>
- Saunders, B., Kitzinger, J., & Kitzinger, C. (2015b). Participant Anonymity in the Internet Age: From Theory to Practice. *Qualitative Research in Psychology*, 12(2), 125-137. <https://doi.org/10.1080/14780887.2014.948697>

- Stein, A. (2010). Sex, Truths, and Audiotape: Anonymity and the Ethics of Exposure in Public Ethnography. *Journal of Contemporary Ethnography*, 39(5), 554-568.
- Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T., & Estabrooks, C. A. (2005). Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *FQS*, 6(1), Art. 29.
- UK Data Service. (2021). <https://ukdataservice.ac.uk/events/webinar-how-to-anonymise-qualitative-and-quantitative-data/> (Zugriff: 01.08.2023)
- University of Leeds. (2008). *Timescapes Anonymisation Guidelines. Version in use 18 Aug 08 lb*. www.timescapes.leeds.ac.uk/assets/files/timescapes/Timescapes-Anonymisation-Guidelines-18Aug08-in-use.doc (Zugriff: 30.06.2020)
- von Unger, H. (2014). Forschungsethik in der qualitativen Forschung: Grundsätze, Debatten und offene Fragen. In H. von Unger, P. Narimani, & R. M'Bayo (Eds.), *Forschungsethik in der qualitativen Forschung: Reflexivität, Perspektiven, Positionen*. (pp. 15-39). VS Verlag.
- von Unger, H. (2018). Forschungsethik, digitale und biographische Interviews. In H. Lutz, M. Schiebel, & E. Tuiider (Eds.), *Handbuch Biographieforschung* (pp. 681-693). Springer VS.
- von Unger, H. (2020). Archivierung und Nachnutzung qualitativer Daten aus forschungsethischer Perspektive. In RatSWD (Ed.), *Handreichung Datenschutz*. <https://doi.org/10.17620/02671.50>
- Watteler, O., & Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In U. Jensen, S. Netscher, & K. Weller (Eds.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (pp. 57-80). Barbara Budrich. <https://doi.org/https://doi.org/10.3224/84742233.05>
- Züll, C. (2015). *Berufscodierung* (SDM - Suvery Guidelines). gesis Leibniz-Institut für Sozialwissenschaften. Doi: 10.15465/sdm-sg_019

9 Anhang: Beispiele für die Ersetzung von personenbezogenen qualitativen Textdaten

Textbeispiel 1

Originaltext:

P08: Naja, und jedenfalls dann hab' ich dann halt so Wettkämpfe mitgemacht. Das ging los, ich war, also, das fing mit der Vereinsmeisterschaft eben an, dann wurde ich Vereinsmeisterin, wurd' ich Kreismeisterin, Bezirksmeisterin, Landesmeisterin, ja, und denn, das ging also immer bis zur Landesmeisterin, vier Jahre so, und dann im 5. Jahr, also mein Traum war's immer, bis zur deutschen Meisterschaft zu kommen, ja und dann hab' ich das 1970 dann geschafft, ne? Also, erst die Endausscheidung in, in Rothenburg, da, in der Sven-Hannawald-Halle. Das war die Landesmeisterschaft, also ich hab' denn Luftgewehr und Kleinkaliber geschossen, ne? (räuspert sich).

IV1: Hm, hm.

P06: Ja, und dann kam ich denn nach Berlin (undeutlich) da, im Olympia-Stadion, also, die Schieß (undeutlich), und das war immer mein Wunsch gewesen, und dann war das.

Markierung evtl. sensibler Stellen in Textbeispiel 1

P08: Naja, und jedenfalls dann hab' ich dann halt so Wettkämpfe mitgemacht. Das ging los, ich war, also, das fing mit der Vereinsmeisterschaft eben an, dann wurde ich Vereinsmeisterin, wurd' ich Kreismeisterin, Bezirksmeisterin, Landesmeisterin, ja, und denn, das ging also immer bis zur Landesmeisterin, vier Jahre so, und dann im 5. Jahr, also mein Traum war's immer, bis zur deutschen Meisterschaft zu kommen, ja und dann hab' ich das 1970 dann geschafft, ne? Also, erst die Endausscheidung in, in Rothenburg, da, in der Sven-Hannawald-Halle. Das war die Landesmeisterschaft, also ich hab' denn Luftgewehr und Kleinkaliber geschossen, ne? (räuspert sich).

IV1: Hm, hm.

P06: Ja, und dann kam ich denn nach Berlin (undeutlich) da, im Olympia-Stadion, also, die Schieß (undeutlich), und das war immer mein Wunsch gewesen, und dann war das.

Anonymisierung Testbeispiel 1:

P08: Naja, und jedenfalls dann hab' ich dann halt so Wettkämpfe mitgemacht. Das ging los, ich war, also, das fing mit der Vereinsmeisterschaft eben an, dann wurde ich Vereinsmeisterin, wurd' ich Kreismeisterin, Bezirksmeisterin, Landesmeisterin, ja, und denn, das ging also immer bis zur Landesmeisterin, vier Jahre so, und dann im 5. Jahr, also mein Traum war's immer, bis zur deutschen Meisterschaft zu kommen, ja und dann hab' ich das Jahresangabe dann geschafft, ne? Also, erst die Endausscheidung in, in Ort A, Landgemeinde, da, in der Austragungsort für Wettkämpfe 1. Das war die Landesmeisterschaft, also ich hab' denn in zwei Disziplinen geschossen, ne? (räuspert sich).

IV1: Hm, hm.

P06: Ja, und dann kam ich denn nach Ort B, große Großstadt (undeutlich) da, im Austragungsort für Wettkämpfe 2, also, die Schieß (undeutlich), und das war immer mein Wunsch gewesen, und dann war das.

Textbeispiel 2

Originaltext:

A: (1) Also, ich, (1) beruflich hatte ich eigentlich anders angefangen. Ich hab dann halt geheiratet, zwei Kinder, aufgehört zu arbeiten, und das war ganz schlimm, Schwiegervater hatte am ersten April siebenachtzig (ei)nen Schlaganfall. Ich war ungefähr 10 Jahre zu Hause und hab dann über nen Krankenpflegekurs in einem Altersheim ein Praktikum gemacht und bin dann da gleich gefragt worden, ob ich nicht mal vertretungsweise als Pflegehelferin (2) einspringen könnte. Hab ich dann auch gemacht und bin dann da einfach hingengeblieben. Aus dem Mal einspringen wurden dann sieben Jahre Arbeit und dann hab ich gedacht, na gut, wenn ich jetzt diese Arbeit schon mach, dann mach ich auch die Ausbildung dazu. und hab dann die Ausbildung zur Pflegehelferin gemacht, hab dann gleich als Leitung im Pflegebereich gearbeitet. Aber ich hatte natürlich sieben Jahre Vorlauf schon Arbeit in dem Bereich, ne.

Markierung evtl. sensibler Stellen in Textbeispiel 3

A: (1) Also, ich, (1) beruflich hatte ich eigentlich anders angefangen. Ich hab dann halt geheiratet, zwei Kinder, aufgehört zu arbeiten, und das war ganz schlimm, Schwiegervater hatte am ersten April siebenachtzig (ei)nen Schlaganfall. Ich war ungefähr 10 Jahre zu Hause und hab dann über nen Krankenpflegekurs in einem Altersheim ein Praktikum gemacht und bin dann da gleich gefragt worden, ob ich nicht mal vertretungsweise als Pflegehelferin (2) einspringen könnte. Hab ich dann auch gemacht und bin dann da einfach hingengeblieben. Aus dem Mal einspringen wurden dann sieben Jahre Arbeit und dann hab ich gedacht, na gut, wenn ich jetzt diese Arbeit schon mach, dann mach ich auch die Ausbildung dazu. und hab dann die Ausbildung zur Pflegehelferin gemacht, hab dann gleich als Leitung im Pflegebereich gearbeitet. Aber ich hatte natürlich sieben Jahre Vorlauf schon Arbeit in dem Bereich, ne.

Anonymisierung Textbeispiel 3:

A: (1) Also, ich, (1) beruflich hatte ich eigentlich anders angefangen. Ich hab dann halt geheiratet, zwei Kinder, aufgehört zu arbeiten, und das war ganz schlimm, Schwiegervater hatte am Ende der 1980er eine zerebrovaskuläre Erkrankung-ICD10 I60-I69. Ich war ungefähr 10 Jahre zu Hause und hab dann über nen Ausbildungskurs in der Gesundheitsversorgung ein Praktikum gemacht und bin dann da gleich gefragt worden, ob ich nicht mal vertretungsweise als Betreuungsberuf im Gesundheitswesen (ISCO 532) (2) einspringen könnte. Hab ich dann auch gemacht und bin dann da einfach hingengeblieben. Aus dem Mal einspringen wurden dann sieben Jahre Arbeit und dann hab ich gedacht, na gut, wenn ich jetzt diese Arbeit schon mach, dann mach ich auch die Ausbildung dazu. und hab dann die Ausbildung zur Fachkraft für einen Betreuungsberuf im Gesundheitswesen (ISCO-08 532) gemacht, hab dann gleich als leitende Funktion im Betreuungsberuf im Gesundheitswesen (ISCO 532-08) gearbeitet. Aber ich hatte natürlich sieben Jahre Vorlauf schon Arbeit in dem Bereich, ne.

Textbeispiel 3

Originaltext Textbeispiel 2

P001: Nein, mein Mann war bei (ei)ner anderen Firma. Ich weiß nicht, kennen Sie die Hamburger Geschäfte? Mein Mann war bei Flimmermüller und ich bin bei Hagenschmidt gewesen.

IV1: Flimmermüller, das ist mir ein Begriff.

P001: Und da hab ich noch zehn Jahre immer +/.

IV1: Aber das heißt, das war dann immer Saisonarbeit. Also immer, wenn Dom +/.

P001: Nur noch Dom dann. Weil ich ja als Verkäuferin in Finkenwerder gearbeitet hab, aber bis Mittag. Und dann fing ich nachmittags um drei auf (de)m Markt an, bis abends, noch zusätzlich.

IV1: Ach, die Verkäuferinnenstelle, die hatten Sie hier in Finkenwerder? Also, das ist noch (ei)ne andere.

P001: (Ei)Ne andere Stelle wieder. Da in der Obst- und Gemüse-(Abteilung), da war ich zwölf Jahre als Verkäuferin.

Markierung evtl. sensibler Stellen in Textbeispiel 2

P001: Nein, mein Mann war bei (ei)ner anderen Firma. Ich weiß nicht, kennen Sie die **Hamburger** Geschäfte? Mein Mann war bei **Flimmermüller** und ich bin bei **Hagenschmidt** gewesen.

IV1: **Flimmermüller**, das ist mir ein Begriff.

P001: Und da hab ich noch zehn Jahre immer +/.

IV1: Aber das heißt, das war dann immer Saisonarbeit. Also immer, wenn **Dom** +/.

P001: Nur noch **Dom** dann. Weil ich ja als Verkäuferin in **Finkenwerder** gearbeitet hab, aber bis Mittag. Und dann fing ich nachmittags um drei auf (de)m Markt an, bis abends, noch zusätzlich.

IV1: Ach, die Verkäuferinnenstelle, die hatten Sie hier in **Finkenwerder**? Also, das ist noch (ei)ne andere.

P001: (Ei)Ne andere Stelle wieder. Da in der Obst- und Gemüse-(Abteilung), da war ich zwölf Jahre als Verkäuferin.

Anonymisierung Textbeispiel 2:

P001: Nein, mein Mann war bei (ei)ner anderen Firma. Ich weiß nicht, kennen Sie die **örtlichen** Geschäfte? Mein Mann war bei |Firma A, Unternehmen im Einzelhandel mit Nahrungs- und Genussmitteln, Getränken und Tabakwaren (in Verkaufsräumen) – WZ47.2| und ich bin bei |Firma B, Einzelhandel mit Textilien – WZ47.51| gewesen.

IV1: |Firma A|, das ist mir ein Begriff.

P001: Und da hab ich noch zehn Jahre immer +/.

IV1: Aber das heißt, das war dann immer Saisonarbeit. Also immer, wenn |Volksfest| +/.

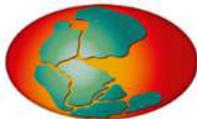
- P001: Nur noch |Volksfest| dann. Weil ich ja als Verkäuferin in |Stadtteil A in Stadt A, große Großstadt| gearbeitet hab, aber bis Mittag. Und dann fing ich nachmittags um drei auf (de)m Markt an, bis abends, noch zusätzlich.
- IV1: Ach, die Verkäuferinnenstelle, die hatten Sie hier in |Stadtteil A in Stadt A, große Großstadt|? Also, das ist noch (ei)ne andere.
- P001: (Ei)Ne andere Stelle wieder. Da in der Obst- und Gemüse-(Abteilung), da war ich zwölf Jahre als Verkäuferin.

Über Qualiservice.

Das Forschungsdatenzentrum Qualiservice archiviert qualitative sozialwissenschaftliche Forschungsdaten aus unterschiedlichen Disziplinen und stellt sie für die wissenschaftliche Nachnutzungen zur Verfügung. Unsere Services sind sicher, flexibel und forschungsorientiert. Sie beinhalten die persönliche und studienspezifische Beratung, die Kuration und Aufbereitung Ihrer Daten für die Nachnutzung und die Langzeitarchivierung ebenso wie die Bereitstellung archivierter Forschungsdaten und relevanter Kontextinformationen. Durch international anschlussfähige Metadaten werden Datensätze such- und findbar. Persistente Identifikatoren (DOI) sorgen für eine nachhaltige Zitierfähigkeit von Daten und Studienkontexten.

Im Juni 2019 wurde Qualiservice vom RatSWD akkreditiert und orientiert sich an dessen Kriterien zur Qualitätssicherung. Qualiservice fühlt sich den *DFG-Richtlinien zur Sicherung guter wissenschaftlicher Praxis* verpflichtet und berücksichtigt darüber hinaus die *FAIR Guiding Principles for Scientific Data Management and Stewardship* sowie die *OECD Principles and Guidelines for Access to Research Data from Public Funding*.

Das Qualiservice-Konsortium:



PANGAEA – zertifiziertes
Weltdatenzentrum



SOCIUM – Forschungszentrum
Ungleichheit und Sozialpolitik



Staats- und
Universitätsbibliothek
Bremen

Staats- und
Universitätsbibliothek
Bremen



GESIS – Leibniz Institut für
Sozialwissenschaften



Fachinformationsdienst Sozial- und
Kulturanthropologie (FID SKA)



Akkreditiert durch



Gefördert von

