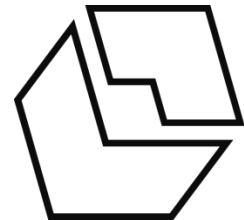


1

forschungsdaten
bildung **informiert**

Alexia Meyermann und Maike Porzelt

Hinweise zur Anonymisierung von qualitativen Daten



Version 1.0 // Dezember 2014

Impressum

forschungsdaten bildung informiert // Nr. 1 (2014)

Herausgeber

Forschungsdatenzentrum (FDZ) Bildung am DIPF
Deutsches Institut für Internationale Pädagogische Forschung
Schloßstraße 29
60486 Frankfurt am Main

Redaktion und Layout

Alexander Schuster

Rechte

Nachdruck, auch auszugsweise, nur mit Genehmigung des FDZ Bildung gestattet.

www.forschungsdaten-bildung.de

1	HINTERGRUND – WOZU ANONYMISIEREN?	4
2	Anonymisierung von Interviewtranskripten	6
2.1	WAS IST ZU ANONYMISIEREN?	6
2.2	VERFAHREN DER ANONYMISIERUNG UND PSEUDONYMISIERUNG	7
2.1.1	<i>Löschen von Merkmalen (und Ersetzen durch Platzhalter)</i>	7
2.2.2	<i>Pseudonymisierung: Ersetzen von Merkmalen durch Pseudonyme</i>	7
2.2.3	<i>Aggregation von Informationen</i>	8
2.2.4	<i>Ersetzen durch Merkmale mit vergleichbarer Bedeutung</i>	8
2.3	ZU WELCHEM ZEITPUNKT SOLLTE DIE ANONYMISIERUNG STATTFINDEN?	9
2.4	WER FÜHRT DIE ANONYMISIERUNG DURCH?	9
2.5	BEST PRACTICES / VORGEHEN BEI DER ANONYMISIERUNG	10
3	Anonymisierung von Audio- und Videodateien	12
3.1	VERFAHREN DER ANONYMISIERUNG	12
3.2	ALTERNATIVEN ZUR ANONYMISIERUNG	13
3.3	TECHNISCHE MÖGLICHKEITEN	14
4	Aufbewahrung von Namen und Adressen für Follow-up-Studien	14
5	Anonymisierung im Spannungsfeld zwischen Forschungsinteresse und Persönlichkeitsschutz	16
6	Literatur	16

HINWEISE ZUR ANONYMISIERUNG VON QUALITATIVEN DATEN

Datenschutzrechtliche Bestimmungen einzuhalten und Persönlichkeitsrechte von Untersuchungspersonen zu wahren, ist eine zentrale Anforderung an die Archivierung und Nachnutzung von personenbezogenen Forschungsdaten. Einen der Bausteine hierzu stellt die Anonymisierung der Forschungsdaten dar. Mit dem Begriff „Anonymisierung“ wird hier der Prozess bezeichnet, aus Forschungsdaten diejenigen Informationen zu entfernen, die eine Identifikation konkreter Personen ermöglichen würden.

Der vorliegende Leitfaden gibt einen Überblick über die Anonymisierung von qualitativen Interviewtranskripten sowie qualitativen Audio- und Videomaterialien, die in der Empirischen Bildungsforschung, vor allem in der Schul- und Unterrichtsforschung, vielfach verwendet werden.

1 Hintergrund – Wozu Anonymisieren?

Anonymisierung bezeichnet das Entfernen personenbezogener oder personenbeziehbarer Informationen aus den Forschungsdaten. In der Regel sind persönliche Daten, sobald es der Forschungszweck erlaubt, zu anonymisieren.

Anonymisierung: „Anonymisieren ist das Verändern personenbezogener Daten derart, 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 bestimmaren natürlichen Person zugeordnet werden können.“ §3 Abs.6 Bundesdatenschutzgesetz (BDSG)

Personenbezogene Daten: „Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmaren natürlichen Person (Betroffener).“ §3 Abs.1 BDSG

Personenbeziehbar: „(...) Einzelangaben, die eine bestimmte Person zwar nicht eindeutig oder unmittelbar identifizieren, die es aber erlauben, die Identität der Person mit Hilfe anderer Informationen festzustellen.“ (Metschke & Wellbrock 2002, S. 19)

In der Forschung werden drei Grade der Anonymisierung unterschieden:

- » die formale Anonymisierung,
- » die faktische Anonymisierung und
- » die absolute Anonymisierung.

Bei der *formalen Anonymisierung* werden direkte Identifizierungsmerkmale (Eigennamen von Personen und Orten, Bilder, Stimmen) aus den Forschungsdaten entfernt. Die *faktische Anonymisierung* zielt darauf ab, die Daten so zu verändern, dass „die Person nur mit einem völlig unverhältnismäßigen Aufwand re-identifiziert werden kann“ (Metschke & Wellbrock 2002, S. 21). *Absolut anonymisierte* Daten sind so verändert, dass die Re-Identifikation einer Person unmöglich ist (vgl. Medjedovic & Witzel 2010, S. 75ff.). Absolut anonymisierte Daten unterliegen nicht dem Datenschutz (vgl. Metschke & Wellbrock, 2002, S. 20).

Die Anonymisierung von Forschungsdaten ist in all den Fällen notwendig, in denen keine Einwilligung der Probanden zur Verwendung ihrer persönlichen Daten vorliegt oder eine derartige Verwendung nicht durch eine entsprechende Rechtsvorschrift ermöglicht wird.

Weil mit der Anonymisierung notwendigerweise ein Verlust von Information einhergeht, sollten stets die Auswirkungen auf das Analysepotenzial bedacht werden, damit unnötige Einschränkungen vermieden werden.

Aufgrund des Restrisikos einer De-Anonymisierung ist zu empfehlen, den Schutz der persönlichen Daten nicht nur vermittels der Anonymisierung von Forschungsdaten zu gewährleisten, sondern diese mit weiteren Maßnahmen, etwa einer Zugangsbeschränkung, zu kombinieren (vgl. z. B. Häder 2009 S. 21f.; Metschke & Wellbrock 2002, S. 22).

2 Anonymisierung von Interviewtranskripten

Die Art und das Ausmaß der Anonymisierung hängen von den jeweiligen Forschungsdaten ab und können daher nicht allgemeingültig bestimmt werden. Die Anforderungen an den Grad der Anonymisierung variieren mit dem Grad der Sensibilität und der Höhe des Re-Identifikationsrisikos der jeweiligen Forschungsdaten sowie der Form der vorliegenden Einwilligungserklärungen der Probanden.

2.1 Was ist zu anonymisieren?

Gegenstände einer Anonymisierung sind insbesondere

- » personenbezogene und personenbeziehbare Merkmale,
- » Personennamen, Ortsangaben, Straßennamen, Bundesländer, Institutionen und Organisationen (z. B. Firmen, Schulen, Institute), Berufsangaben, Titel und Bildungsabschlüsse, Alter, Zeitangaben/kalendarische Daten, Bilder und Stimmen,
- » indirekte, aber spezifische Kontextinformationen,
- » Merkmale der Probanden wie auch solche von dritten, in den Interviews erwähnten, Personen (auch die Persönlichkeitsrechte der Interviewenden, Transkribierenden usw. sind zu berücksichtigen),
- » sensible Informationen: Informationen zur ethnischen Herkunft, politischen Meinung, religiösen oder philosophischen Überzeugung, Gewerkschaftszugehörigkeit, Gesundheit oder zum Sexualleben (§3 Abs.9 BDSG).

Personenbeziehbare Merkmale sind Merkmale, die es (häufig durch ihre Kombination) ermöglichen, eine Person zu identifizieren, also indirekt identifizierende Merkmale.

➤ *Beispiel für personenbeziehbare, das heißt indirekt identifizierende Merkmale:*

Beruf [Optiker] in Kombination mit Arbeitsort [Kahl am Main]. Gibt es in Kahl am Main nur einen einzigen Optiker, dann kann die betreffende Person bereits anhand dieser beiden Merkmale in ihrer Kombination eindeutig identifiziert werden.

Grundsätzlich ist bei der Einschätzung des Re-Identifikationsrisikos das „mögliche Zusatzwissen“ (Metschke & Wellbrock 2002, S. 21) der die Daten besitzenden Stelle bzw. eines potenziellen „Angreifers“ zu berücksichtigen. Dieses kann nicht nur aus den Forschungsdaten selbst, sondern auch aus sonstigen allgemein oder nicht allgemein zugänglichen Quellen stammen.

Auch *indirekte, aber spezifische Kontextinformationen* können zur Re-Identifizierung genutzt werden. Zu achten ist auf Merkmalsausprägungen, die – in dem spezifischen Kontext – *selten* vorkommen.

➤ *Beispiele für spezifische Kontextinformationen:*

- » Autobauer aus Niedersachsen,
- » Rektor der örtlichen Grundschule,
- » Person mit senegalesischer Abstammung in Kleinstadt X,
- » Familie mit fünf Kindern in Kleinstadt X.

2.2 Verfahren der Anonymisierung und Pseudonymisierung

2.2.1 Löschen von Merkmalen (und Ersetzen durch Platzhalter)

Werden Merkmale gelöscht, dann sind diese zumindest durch Platzhalter zu ersetzen. Bei dieser einfachsten Form der Anonymisierung erfolgt eine starke Abstraktion der ursprünglichen Information.

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

2.2.2 Pseudonymisierung: Ersetzen von Merkmalen durch Pseudonyme

Neben der Anonymisierung durch das Löschen von Informationen besteht die Möglichkeit der *Pseudonymisierung* von Forschungsdaten.

Pseudonymisierung: „Pseudonymisieren ist das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.“ §3 Abs.6a BDSG

Bei einer Pseudonymisierung werden Merkmale auf der Grundlage einer zu bildenden *Zuordnungsvorschrift* durch andere ersetzt. Wenn diese Zuordnungsvorschrift bekannt ist, können die Originalangaben wieder hergestellt werden. Es wird empfohlen, die Zuordnungsvorschrift durch einen „Datentreuhänder“ verwahren zu lassen und nicht beim Forschenden aufzubewahren (vgl. Metschke & Wellbrock 2002, S. 24). Die Aufbewahrung einer Zuordnungsvorschrift ist aber nur in bestimmten Fällen, beispielsweise bei Follow-up-Studien, erforderlich (vgl. Kapitel 4).

In der Praxis wird mit Pseudonymisierung auch das Ersetzen von Merkmalen *durch Merkmale vergleichbaren Informations- und Sinngehalts* bezeichnet, die jedoch keinen Rückschluss auf den Probanden/die Probandin erlauben. Damit lässt sich das Analysepotenzial der Daten erhalten. Von Vorteil ist dabei auch, dass der Lesefluss erhalten bleibt, wohingegen der Einsatz von (so genannten nicht sprechenden) Ziffern oder Buchstabenkombinationen diesen erheblich stört. Ein Vorname sollte also durch einen vergleichbaren Vornamen und ein Nachname durch einen vergleichbaren Nachnamen (wenn möglich mit gleich vielen Silben und aus dem gleichen Kulturkreis) ersetzt werden.

➔ *Beispiele von Pseudonymen* (übernommen von www.wisdom.at/pdf/TemplateAnonymisierung_WISDOM.pdf, abgerufen am 29. 09. 2014):

- » **Eigename:** Aus einer Frau *Melanie Bauer* wird keine *Beatrix Fürstenberg*, sondern eine *Melissa Landmann*.

- » **Beruf:** Aus einer *Friseurin* wird keine *Bürokauffrau*, sondern eine *Kosmetikerin*.
- » **Herkunft:** Jemandem, der aus einem *Wiener Arbeiterbezirk* stammt, wird als Herkunftsort nicht ein *Wiener Nobelbezirk* zugeordnet, sondern ein *anderer Wiener Arbeiterbezirk*.

Originalangabe, Identifikator		(2) Pseudonym
Name	Aishe	Nesli
	Christian	Thomas
Arbeitgeber	Bäckerei Schmidt	Metzgerei
Beruf	Friseur	Kosmetiker

2.2.3 Aggregation von Informationen

Eine dritte Art der Anonymisierung von Forschungsdaten besteht in der Vergrößerung oder Aggregation von Informationen durch Bildung von Klassen oder Kategorien. Hierzu gehören beispielweise das Ersetzen des konkreten Alters durch Altersklassen, das Ersetzen des konkreten Arbeitgebers durch die Angabe der Branche oder der Firmengrößenklasse und das Ersetzen eines spezifischen Mädchennamens durch „Schülerin“.

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

2.2.4 Ersetzen durch Merkmale mit vergleichbarer Bedeutung

Eine weitere Möglichkeit der Anonymisierung besteht darin, die vorliegenden Informationen durch solche zu ersetzen, welche die Bedeutung für die und die Beziehung zur Untersuchungsperson beinhalten (vgl. University of Leeds 2008: www.timescapes.leeds.ac.uk/assets/files/timescapes/Timescapes-Anonymisation-Guidelines-18Aug08-in-use.doc).

- » Personen können umschrieben werden aufgrund ihrer Bedeutung für die Untersuchungsperson, beispielsweise Freundin, Mutter, Lehrer.
- » Ortsnamen können umschrieben werden anhand der Bedeutung des Orts für die Untersuchungsperson, etwa Geburtsort, Wohnort, Arbeitsort.
- » Ländername: je nach Forschungszusammenhang, wie Land mit hoher Jugendarbeitslosigkeit oder Wohlfahrtsstaat.

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

2.3 Zu welchem Zeitpunkt sollte die Anonymisierung stattfinden?

Die Anforderungen des Datenschutzes sollten bereits *zu einem frühen Zeitpunkt* des Forschungsprozesses berücksichtigt werden. Geklärt werden sollte dann, was im Vorfeld der Datenerhebung, während der Datenerhebung und im Anschluss an die Datenerhebung zu beachten ist:

- » Die Art der Einwilligungserklärungen, die von den Probanden unterzeichnet werden, ist entscheidend für die späteren Verwendungsmöglichkeiten der Forschungsdaten.
- » Der Aufwand der Anonymisierung lässt sich erheblich verringern, wenn im Vorfeld eines Interviews mit dem Probanden vereinbart wird, identifizierende Merkmale, insbesondere die Eigennamen von Personen, Orten und Organisationen, im Gespräch nicht zu erwähnen – sofern dies die Aussagefähigkeit nicht beeinträchtigt.
- » Bereits während der Transkription kann es sinnvoll sein, Forschungsdaten zu anonymisieren durch a) das Ersetzen von Eigennamen und anderen direkten Identifikatoren und b) die Kennzeichnung indirekter Identifikatoren und sensibler Textpassagen.

Die während der Transkription durchgeführten Anonymisierungsschritte ersetzen in der Regel nicht eine abschließende Beurteilung des gesamten vorhandenen Materials. Die Einschätzung von Merkmalen als sensibel oder als personenbeziehbar ergibt sich erst in der Gesamtschau der Transkripte. Um einschätzen zu können, welche Informationen sensibel sind und welche nicht, wird empfohlen, das Material vorher vollständig anzuhören.

2.4 Wer führt die Anonymisierung durch?

Die Anonymisierung kann sowohl von MitarbeiterInnen von Datenzentren als auch von den PrimärforscherInnen selbst durchgeführt werden. Eine gegenseitige Abstimmung über die verwendeten Anonymisierungsregeln ist in der Regel sinnvoll. Die PrimärforscherInnen besitzen das größte Expertenwissen hinsichtlich der Forschungsdaten (vgl. Liebig et al. 2014). Sie können am besten die vorhandenen Re-Identifikationsrisiken sowie die Sensibilität der Daten einschätzen und die durch die Anonymisierung erreichten Eingriffe in das Analysepotenzial beurteilen. Auch bei der Wahl der Pseudonyme sind fachspezifische Kenntnisse erforderlich. Demgegenüber können ArchivarInnen das erforderliche datenschutzrechtliche Wissen beisteuern und die vorgenommene Anonymisierung mit weiteren Maßnahmen zur Einhaltung des Datenschutzes abstimmen.¹

Das österreichische Archiv WISDOM empfiehlt „DatengeberInnen außerdem, qualitativen Datensätzen bei der Archivierung zusätzlich zu einer Anonymisierung ein Merkblatt beizufügen, in dem auf heikle Passagen in Interviews oder aus zukünftigen Veröffentlichungen resultierende potenzielle Gefahren und/oder Unannehmlichkeiten für InterviewpartnerInnen hingewiesen wird“ (www.wisdom.at/ma_phase5.aspx, abgerufen am 29.09.2014).

¹ Das deutsche Archiv Qualiservice arbeitet derzeit an einem Tool zur (teil-)automatischen Anonymisierung von Transkripten, dem so genannten eAnonymizer.

2.5 Best Practices / Vorgehen bei der Anonymisierung

1. Wenn es möglich ist, sollten bei Interviews bereits vor deren Durchführung der Proband/die Probandin darauf hingewiesen werden, dass sie *im Interview keine Namen oder identifizierenden Merkmale (z. B. Ortsangaben) nennen*, die ihn/sie selbst oder Dritte betreffen.
2. Die Anonymisierung sollte *bereits bei der Transkription durchgeführt* bzw. sensible Informationen sollten *bereits hier für eine spätere Anonymisierung gekennzeichnet* werden (z. B. durch die Verwendung von eckigen Klammern). Um einschätzen zu können, welche Informationen sensibel sind und welche nicht, wird empfohlen, das Material vorher vollständig anzuhören.
3. Anonymisierte und sensible Textpassagen sind klar zu *kennzeichnen*.

Vorschlag zur Kennzeichnung bereits anonymisierter Textstellen:

- » `<pseudonym> [pseudonym] @@pseudonym##`

Vorschlag zur Kennzeichnung noch zu anonymisierender Textstellen:

- » `$$sensible Textstellen##`

Die University of Leeds (2008) empfiehlt die Verwendung von `@@pseudonym##`, da diese Sonderzeichen üblicherweise nicht mit anderen Bedeutungen belegt sind. Durch die Verwendung von unterschiedlichen Start- und Endzeichen beugt man zudem dem Effekt vor, dass markierte mit nicht markierten Textstellen verwechselt und die Endzeichen vergessen werden.

4. Es sollte ein *Anonymisierungsprotokoll angefertigt* werden, in dem alle Ersetzungen, Zusammenfassungen (Aggregationen) und Löschungen dokumentiert sind, und es sollte festgehalten werden, welche Personen die Anonymisierung durchgeführt haben.

Das Anonymisierungsprotokoll ist erforderlich

- » für eine konsistente und eindeutige Verwendung von Pseudonymen,
- » falls eine spätere De-Anonymisierung vorgenommen werden soll, beispielsweise für Nachfolgestudien (spätere Erhebungswellen),
- » für die nachhaltige Dokumentation des Vorgehens.

➤ **Beispiel Anonymisierungsprotokoll** (Quelle: In Anlehnung an <http://ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx>, abgerufen am 29.09.2014)

Interviewnummer und Seitenzahl	Original	Veränderung
<i>Interview 1</i>		
S. 1	32 Jahre alt	Zwischen 25 und 35
S. 2	Gelsenkirchen	Stadt in Nordrhein-Westfalen / im Ruhrgebiet / Stadt mit Einwohnerzahl von 200T bis 300T
S. 2	18. 02. 2014	Februar
S. 3	Meine Freundin Angelika	Meine Freundin Jessica
S. 7	Walter-Eber-Grundschule Bad Neuheim	Grundschule
S. 8	Deutsche Bank	Branche Finanzen und Versicherungen

Wenn es von Seiten des Datenschutzrechts erlaubt ist, sollten das Anonymisierungsprotokoll und der Anonymisierungsschlüssel aufbewahrt werden – allerdings stets *getrennt* von den anonymisierten Dateien (z. B. beim institutsinternen Datenschutzbeauftragten).

5. Das *Verwenden von automatischen „Suchen und Ersetzen“-Funktionen* (z. B. in MS Word) ist zeitsparend. Aufgrund der damit einhergehenden Fehleranfälligkeit sollte jedoch zusätzlich eine manuelle Kontrolle erfolgen. Tippfehler beispielsweise führen bei der Transkription dazu, dass die entsprechenden Wörter bei der automatischen Suche nicht gefunden werden. Empfohlen wird daher der Einsatz der „Suchen und Ersetzen“-Funktion mit Abfrage bei jedem Wort, *nicht* jedoch für alle im Text gefundenen Wörter durch die Funktion „alle ersetzen“.
6. Pseudonyme sollten *konsistent* verwendet werden, und zwar sowohl innerhalb eines Transkripts als auch in allen Transkripten einer Erhebung, innerhalb des Projektteams, in den verschiedenen Publikationen und in etwaigen Folgeerhebungen.
7. Wenn es von Seiten des Datenschutzrechts erlaubt ist, sind die Rohdaten aufzubewahren. Eine Kopie der Originaldateien ist anzufertigen und die Anonymisierung ist in der Kopie durchzuführen. Anonymisierungsprotokolle sind separat von den Daten aufzubewahren.

(Quelle: In Anlehnung an <http://ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx>, abgerufen am 29.09.2014)

3 Anonymisierung von Audio- und Videodateien

3.1 Verfahren der Anonymisierung

- » Akustische Veränderung (Löschen, Überdecken, Ersetzen, Verfremden)
- » Visuelle Manipulation (Überdecken, Verpixeln, Unkenntlich machen)

Um Audio- und Videomaterialien zu anonymisieren, sind diese *akustisch zu verfremden* und/oder *visuell zu manipulieren*. Bei der Anonymisierung von Audiomaterialien (vgl. Pätzold 2005) sollten zunächst alle Informationen, die Rückschluss auf die interviewte Person oder die im Interview genannten Personen ermöglichen, mit einem entsprechenden Programm *entfernt* oder *akustisch überdeckt* werden. Alternativ können mit einem solchen Programm auch Passagen mit persönlichem Inhalt *ersetzt* statt gelöscht werden, damit der Informationsgehalt der entsprechenden Daten zumindest sinngemäß erhalten bleibt.

Um zu vermeiden, dass die Untersuchungspersonen an ihren Stimmen erkannt werden können, sollten in einem nächsten Schritt die Materialien *akustisch verfremdet werden*. Das lässt sich entweder durch Rauschsignale, mit denen sich eine hohe Verfremdung erzielen lässt, oder durch eine weniger stark verfremdende Tonhöhenveränderung (pitch shifting) erreichen. Dabei bleiben der Abstand zwischen den Tonhöhen der verschiedenen Stimmen sowie weitere Charakteristika in einem größeren Umfang erhalten.

Bei audio-visuellen Materialien müssen für eine Anonymisierung zusätzlich visuelle Manipulationen durchgeführt werden. Beispielsweise kann man die Gesichter der aufgenommenen Personen mit Hilfe von Videoschnittprogrammen und den darin enthaltenen Maskierungsfunktionen teilweise oder vollständig *überdecken*, *verpixeln* oder in einer anderen Weise *unkennlich machen*.

Anonymisierung von Audiodaten und audio-visuellen Materialien (AV-Medien):

Akustische Veränderung

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

Visuelle Manipulation

- » Augen mit schwarzen Balken verdecken,
- » Verpixelung oder sonstige Verfremdung von Gesichtern.

Mit diesen Formen der Anonymisierung von AV-Materialien geht jedoch ein teilweise erheblicher Verlust des Analysepotenzials einher. Analysen, bei denen Gesichtsausdrücke, Blickrichtungen oder Stimmauswertungen im Mittelpunkt stehen, sind dann nicht mehr durchführbar, wenn Gesichter unkenntlich gemacht wurden (vgl. Andersson & Sørvik 2013). Darüber hinaus ist mit solchen Anonymisierungsmaßnahmen immer ein Eingriff in die technische Datenqualität verbunden. Die Anforderungen an die technische Qualität und das Format der Ausgangsdateien sind dementsprechend hoch.

Der Arbeitsaufwand bei diesen Formen der Anonymisierung ist hoch, und ihre Anwendung bedarf eines entsprechenden technischen Könnens.

Nachteile der technischen Eingriffe zur Anonymisierung mittels akustischer oder visueller Eingriffe in die Ausgangsmaterialien:

- » erheblicher Arbeitsaufwand,
- » hohe Kosten,
- » Qualitätsverlust aufgrund des Eingriffs in die technische Datenqualität,
- » erheblicher Verlust des Analysepotenzials,
- » Daten sind für Analysen, die auf Gesichtsausdrücke, Blickrichtungen oder Stimmauswertungen angewiesen sind, nicht mehr brauchbar,
- » hohe Anforderungen an die technische Qualität und das Format der Dateien: Technisch hochwertige Ausgangsdateien sind erforderlich, damit die mit der Veränderung einhergehenden Qualitätsverluste das Analysepotenzial nicht zu sehr einschränken,
- » hohe Anforderungen an das technische Know-how der Forscher/-innen.

3.2 Alternativen zur Anonymisierung

- » Einverständnis der Probanden, die Daten unverändert nutzen zu können, einholen
- » Zugangsbeschränkte Bereitstellung
- » Das Erheben von personenbezogenen Daten im Vorfeld durch entsprechende Gestaltung der Aufzeichnung vermeiden

Da mit der Anonymisierung der AV-Medien Nachteile (Einschränkungen des Analysepotenzials, Qualitätsverlust, Arbeitsaufwand) verbunden sind, sind – insofern dies von Seiten des Datenschutzrechts erlaubt ist – Alternativen in Betracht zu ziehen.

Die optimale Lösung ist, das *Einverständnis* der Untersuchungspersonen für die Verwendung der (unveränderten) Originalmaterialien einzuholen. Dieses Einverständnis sollte auch die Weiterverwendung der Daten durch SekundärforscherInnen einschließen. Eine Arbeitsgruppe des Ständigen Ausschusses Forschungsdateninfrastruktur hat Muster für eine Einverständniserklärung zur Vorlage bei qualitativen Interviews herausgegeben (vgl. Liebig et al. 2014).

Zusätzlich müssen die AV-Materialien aufgrund ihres besonderen Re-Identifikationsrisikos und der damit verbundenen ethischen Verantwortung entsprechend sicher aufbewahrt und dürfen nur unter restriktiven Bedingungen *zugänglich gemacht* werden (vgl. Andersson & Sørvik 2013). Der Zugang zu Daten kann on-site, in den Räumen des Datenzentrums, oder off-site, bei den Sekundärnutzenden (z. B. auf CD-ROM oder per Remote Access), erfolgen. Es wird empfohlen, den Zugang immer nur auf der Grundlage eines schriftlichen Nutzungsvertrages zu gewähren, da die NutzerInnen durch diesen Vertrag explizit zur Einhaltung des Datenschutzes und damit zur Wahrung der Persönlichkeitsrechte der Betroffenen verpflichtet werden können. In der Praxis wird der Zugang zudem häufig nur zu wissenschaftlichen Zwecken gestattet. Das FDZ Bildung überprüft außerdem die Identität der Antragsteller mithilfe eines Post-Ident-Verfahrens.

Einen besonderen Fall stellt die Videoaufnahme einer Gruppe von Personen dar, wenn nur Teile der Gruppe einer Weiterverwendung des unveränderten Materials zugestimmt haben:

Beispiel aus der Unterrichtsforschung

Sind nicht alle SchülerInnen einer Schulklasse, die aufgezeichnet werden soll, mit der Weiterverwendung der Forschungsdaten einverstanden, dann lässt sich durch eine entsprechende Aufstellung der Kamera und die Wahl der Sitzordnung vermeiden, dass diejenigen SchülerInnen, die nicht einverstanden sind, aufgenommen werden.

Eine etwaige spätere Archivierung und Nachnutzung der Daten sollte also schon frühzeitig bei der Planung und Vorbereitung einer Aufzeichnung bedacht werden.

3.3 Technische Möglichkeiten

Die meisten professionellen Videobearbeitungsprogramme bieten die Möglichkeit (z. B. durch Verpixe- lung), Gesichter oder andere persönliche Kennzeichen in Videoaufzeichnungen zu anonymisieren. Sol- che kostenpflichtigen und teilweise teuren Programme sind beispielsweise:

- » Adobe Premiere
- » Sony Movie Studio
- » Corel VideoStudio
- » Magix Video

Beispiele für Freeware sind:

- » VideoPad Video Editor
- » Avidemux

Für Veränderungen an Audiodateien bieten sich Programme an, welche die Stimme(n) der aufgenom- menen Person(en) verfremden oder bestimmte Teile davon vollständig entfernen. Nach Pätzold (2005) bieten sich hierfür beispielsweise folgende Freeware-Programme an:

- » Zerius Vocoder (Vocodersoftware für das Verfremden der Stimme durch Rauschsignale)
- » Audacity (Software für das Verfremden der Stimme durch Tonhöhenveränderung sowie für das Entfernen einzelner Abschnitte innerhalb der Tondatei (pitch shifting))

4 Aufbewahrung von Namen und Adressen für Follow-up- Studien

Sind Follow-up-Studien geplant, beispielsweise Panelbefragungen, dann ist es erforderlich, die Namen und Adressen der Untersuchungspersonen aufzubewahren, um sie erneut kontaktieren zu können.

Für die Auswertung von Panelbefragungen ist es erforderlich, die Angaben der Untersuchungspersonen der verschiedenen Wellen miteinander verknüpfen zu können.

Dabei sind folgende Aspekte zu beachten:

- » Die Namen und Adressen sollten getrennt von den Forschungsdaten aufbewahrt werden.

- » Für die Aufbewahrung kann auf einen Treuhänder zurückgegriffen werden, beispielsweise den Datenschutzbeauftragten der wissenschaftlichen Einrichtung (vgl. Metschke & Wellbrock 2002, S. 41ff.). Dieser kann auch die Schlüsseldatei aufbewahren, anhand derer die Erhebungsdaten der verschiedenen Messzeitpunkte miteinander verknüpft werden können (vgl. Abbildung 1).
- » Die Zuordnung der Daten aus der erneuten Erhebung (z. B. des zweiten Messzeitpunkts) zu denen der vorherigen Erhebung (z. B. des ersten Messzeitpunkts) sollte nicht mithilfe der Namens- und Adressinformationen der Untersuchungspersonen erfolgen, sondern vermittels von dafür zu erstellenden nichtsprechenden Identifikatoren.

Außerdem ist es sinnvoll, *Löschfristen* für die Aufbewahrung der Adressen festzulegen.

Generell wird empfohlen, bei der Arbeit im Feld andere Identifikatoren zu benutzen als in den von den Forschenden bei der Analyse verwendeten Daten.



Abbildung 1: Aufbewahrung von Adressdaten für eine Follow-up-Studie

5 Anonymisierung im Spannungsfeld zwischen Forschungsinteresse und Persönlichkeitsschutz

Den datenschutzrechtlich und forschungsethisch begründeten Ansprüchen an eine Anonymisierung der Daten im Interesse der Untersuchungspersonen stehen von Seiten der Forschung der Wunsch nach dem Erhalt von Informationen und der Sicherstellung des Analysepotenzials entgegen. Dieses Dilemma kann nur von Fall zu Fall, und zwar in Abhängigkeit vom konkreten Datenmaterial, den damit einhergehenden Schutzbedürfnissen der Betroffenen sowie den Interessen der Forschung, aufgelöst werden. Bei der Beurteilung der Risiken einer Re-Identifikation sind die potenziellen Schäden und Nachteile für die Untersuchungspersonen ebenso zu berücksichtigen wie die Wahrscheinlichkeit und die Komplexität eines Re-Identifizierungsversuchs durch einen „Angreifer“ (vgl. Häder 2009, S. 21).

Auf jeden Fall ist eine formale Anonymisierung durchzuführen, das heißt, Eigennamen von Personen und Orten müssen entfernt werden. Welche Anonymisierungsmaßnahmen darüber hinaus nötig sind, hängt nicht zuletzt vom jeweiligen Verwendungszweck ab. Denn mit unterschiedlichen Verwendungszwecken (vollständige/teilweise Bereitstellung von Transkripten, Sekundärnutzung on-site oder off-site, Restriktionen hinsichtlich der Verwendung des Materials usw.) sind unterschiedliche Re-Identifikationsrisiken verbunden.

Es sind daher folgende Fragen zu stellen:

- » Sollen Transkripte für eine Sekundäranalyse verwendet werden oder dient ihre Aufbewahrung nur der Nachvollziehbarkeit der eigenen Ergebnisse in der Zukunft?
- » Falls Transkripte der Sekundärnutzung verfügbar gemacht werden sollen, welche Nutzung wird angestrebt? (Mit der Herausgabe von Transkripten gehen höhere Risiken einher als mit ihrer Bereitstellung ausschließlich für die Vor-Ort-Nutzung.)
- » Sollen Transkripte vollständig oder teilweise (ausschnitthaft) verfügbar gemacht werden?
- » Können die Auflagen hinsichtlich der Verwendung des Materials in Publikationen implementiert und kontrolliert werden?
- » Können Zugangsbeschränkungen für SekundärnutzerInnen implementiert und kontrolliert werden?

Da selbst bei einer *faktischen Anonymisierung* – wie sie gesetzlich vorgeschrieben ist (Personen sind so zu anonymisieren, dass sie nur mit einem „unverhältnismäßig hohen Aufwand“ re-identifiziert werden könnten) – das Restrisiko einer Re-Identifikation nicht auszuschließen ist, sollte zusätzlich zur Anonymisierung auch eine Zugriffskontrolle implementiert werden, um die Schutzrechte der Untersuchungspersonen zu gewährleisten.

6 Literatur

Andersson, E. & Sørvik, G. O. (2013). *Reality Lost? Re-Use of Qualitative Data in Classroom Video Studies*, Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 14 (3), Art. 1. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:0114-fqs130313> (abgerufen am 27.10.2014).

- Häder, M. (2009). *Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland* (RatSWD Working Paper Series. 90). Verfügbar unter: http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf (abgerufen am 27.10.2014).
- Liebig, S., Gebel, T., Grenzer, M., Kreusch, J., Schuster, H., Tscherwinka, R., Watteler, O. & Witzel, A. (2014). *Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten* (RatSWD Working Paper Series. 238). Verfügbar unter: http://www.ratswd.de/dl/RatSWD_WP_238.pdf (abgerufen am 27.10.2014).
- Medjedovic, I. & Witzel, A. (2010). *Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte*. Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Metschke, R. & Wellbrock, R. (2002). *Datenschutz in Wissenschaft und Forschung* (Materialien zum Datenschutz. 28). Verfügbar unter: <http://www.datenschutz-berlin.de/attachments/47/Materialien28.pdf?1166527077> (abgerufen am 27.10.2014).
- Pätzold, H. (2005). Sekundäranalyse von Audiodaten. Technische Verfahren zur faktischen Anonymisierung und Verfremdung, *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6 (1), Art. 24. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501249> (abgerufen am 27.10.2014).
- University of Leeds (2008). *Timescapes Anonymisation Guidelines*. Version in use 18 Aug 08 lb. Verfügbar unter: www.timescapes.leeds.ac.uk/assets/files/timescapes/Timescapes-Anonymisation-Guidelines-18Aug08-in-use.doc (abgerufen am 27.11.2014).