eLabour Datenschutzkonzept (Stand 10/2017)

Ziel

In diesem Dokument werden die Grundsätze zur sicheren Speicherung, Verarbeitung, Bereitstellung und Nutzung von qualitativen Forschungsdaten im Forschungsverbund eLabour im Sinne des Datenschutzes dokumentiert und fortentwickelt. Um die wissenschaftliche Nachnutzung qualitativer, arbeitssoziologischer Forschungsdaten zu ermöglichen, werden umfangreiche empirische Studien der datengebenden Forschungsinstitute in die eLabour Infrastruktur eingebracht, aufbereitet und für wissenschaftliche Sekundäranalysen zur Verfügung gestellt. Die Speicherung, Bereitstellung und Weitergabe qualitativer, sozialwissenschaftlicher Forschungsdaten stellt besondere Anforderungen an die Maßnahmen zum Datenschutz: Einerseits können qualitative Forschungsdaten aufgrund der prinzipiellen Offenheit der Erhebungsmethoden jederzeit sensible personenbezogene Informationen enthalten, d.h. die Daten müssen umfassend auf solche Informationen überprüft werden, wenn man die Weitergabe sensibler, personenbezogener Informationen unterbinden will. Andererseits ist die sinnvolle Nachnutzung und Analyse ohne solche sensiblen, personenbezogenen Informationen oft nicht sinnvoll. D.h. Maßnahmen zum Datenschutz müssen darauf ausgerichtet sein, auch personenbezogene Informationen bis zu einem gewissen Grad zu erhalten oder zu umschreiben. Daher ist Datenschutz hier eine Gratwanderung zwischen diesen widersprüchlichen Anforderungen. Dem entspricht der Weg, den wir hier einschlagen: Eine intensive Bewertung der Risiken für die Personen wird verbunden mit differenzierten Einschränkungen bei der Freigabe und Nutzung, die durch Nutzungsverträge und die IT-basierte Dokumentation von Zugriffen abgesichert wird.

Das vorliegende Konzept ist engem Austausch mit dem Datenschützer der Uni-Göttingen erarbeitet worden, muss aber noch einer abschliessenden rechtlichen Prüfung unterzogen werden.

Gliederung

Rechtliche Grundlagen

Die Verarbeitung und wissenschaftliche Nutzung personenbezogener Daten der datenhaltenden Forschungseinrichtungen erfolgt im Forschungszentrum eLabour nach der EU-Datenschutzgrundverordnung (EU-DSGVO). Zusätzlich gelten für das Soziologisches Forschungsinstitut Göttingen (Sofi Göttingen e.V.), das Institut für sozialwissenschaftliche Forschung München (ISF München) und die Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) das Bundesdatenschutzgesetz-Neu (BDSG-Neu), da sie keine öffentlichen Einrichtungen bzw. Körperschaften des öffentlichen Rechts sind. Für die Sozialforschungsstelle Dortmund (sfs) und das Forschungsdatenzentrum Betriebs- und Organisationsdaten der Universität Bielefeld (FDZ-BO) als öffentliche Einrichtungen bzw. Körperschaft des öffentlichen Rechts gilt zusätzlich zur EU-DSGVO das Landesdatenschutzgesetz-NEU (LDSG-Neu) Nordrhein-Westfalen. Ebenso gilt das LDSG-Neu Niedersachsen für die Niedersächsische Landes- und Universitätsbibliothek Niedersachsen (SUB) und das Forschungszentrum L3S sowie das LDSG-Neu Thüringen für das Institut für Soziologie der Friedrich-Schiller-Universität Jena.

Die Verarbeitung der Originalforschungsdaten der Forschungseinrichtungen umfasst gemäß Art. 4, Abs. 2 EU-DSGVO „das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung“ personenbezogener Daten. Die EU-DSGVO ist seit dem 4. Mai 2016 bekannt und wird ab 25. Mai 2018 die Datenschutzgesetze des Bundes und der Länder in Ihrer bisherigen Form ablösen. Das BDSG-Neu wurde am 27.04.2016 beschlossen und gilt flankierend zur EU-DSGVO, in dem es diese ergänzt und Lücken in der EU-DSGVO schließt bzw. diese konkretisiert. Die Datenschutzgesetze der Länder-Neu sind derzeit noch nicht bekannt. Inwiefern es durch LDSG-Neu zur Verschärfungen und Konkretisierung ist derzeit noch offen, eine Entscheidung wird noch im Jahr 2017 erwartet.

Trotz einer Vielzahl von Neuerungen wird es bei dem Grundsatz des „Verbotes mit Erlaubnisvorbehalt“ verbleiben, so dass personenbezogene Daten nur mit der Zustimmung der Betroffenen oder aufgrund eines besonderen Erlaubnistatbestands (Rechtsgrundlage) verarbeitet werden dürfen (Art. 6 EU-DSGVO). Zentral für die Forschung ist die Ausnahme vom strengen Zweckbindungsgrundsatz – wie noch im BDSG und in den LDSG – für die zu wissenschaftlichen Zwecken erhobenen Daten (§27 BDSG-Neu). Unter der Voraussetzung geeigneter Vorkehrungen (§22 BDSG-Neu), d.h. datenverändernder, technischer und organisatorischer Maßnahmen, zum Schutz persönlicher Daten der an einer Studie Teilnehmenden, wird auch eine vom ursprünglichen Forschungszweck abweichende Datenverarbeitung zu wissenschaftlichen Zwecken möglich sein (Schaar 2016). Es wird möglich, wissenschaftliche Daten zu Forschungszwecken mit abgestuften Konzepten bis hin zu Klardaten verarbeiten zu dürfen, unter der Voraussetzung, dass technische und organisatorische Maßnahmen zum Schutz der Befragten getroffenen werden. Die EU-DSGVO wird auch eine Beschränkung der Auskunftsrechte über die Befragten beinhalten, die sich insbesondere auf bestimmte Kategorien personenbezogener Daten bezieht, d.h. Ethnie, Gesundheitsdaten, Angaben zu Einstellungen, Gewerkschaftszugehörigkeit, politische Einstellungen. Doch auch in diesem Fall dürfen diese Daten zu wissenschaftlichen Zwecken – auch ohne Einwilligung der Befragten – verarbeitet werden, wenn die Interessen der wissenschaftlichen Nutzung die Interessen der Befragten erheblich überwiegen und die Rechte der Befragten durch technische und organisatorische Maßnahmen gesichert werden - Pseudonymisierung

in der Kombination mit einem technisch und organisatorisch kontrollierten Zugang zu den pseudonymisierten Daten.

Die Begründung des überwiegenden wissenschaftlichen Interesses gemäß der EU-DSGVO folgt dabei im Wesentlichen der Argumentation des überwiegend öffentlichen Interesses der Datennutzung gemäß dem BDSG (siehe Anhang 1.)

Das BDSG-Neu §40 erlaubt im Gegensatz zur bisherigen Regelung im BDSG-Alt §25 auch die Übermittlung von personenbezogenen Daten an private Einrichtungen, wenn geeignete Maßnahmen zum Schutz der betroffenen Personen ergriffen werden (§22 und §27 BDSG-Neu).

Um eine möglichst hohe wissenschaftliche Qualität der qualitativen Forschungsdaten für die wissenschaftliche Nachnutzung zu erhalten, wendet das Datenschutzkonzept von eLabour im wissenschaftlichen und öffentlichen Interesse bereits vorausschauend den erleichterten Zweckbindungsgrundsatz der EU-DSGVO an. Die personenbezogenen Forschungsdaten der datengebenden Institute werden im Zentrum eLabour im Auftrag der Datenhalter für die wissenschaftliche Nachnutzung vorbereitet. Die originalen Forschungsdaten der Datenhalter werden gemäß der EU-DSGVO im Hinblick auf die personenbezogenen und organisationsbezogenen Risiken klassifiziert und in Abhängigkeit vom bewerteten Risiko für Personen und Organisationen unterschiedlich weitgehend anonymisiert. Im anschließenden Freigabeprozess werden auf der Grundlage entsprechender Verträge Funktionen und Verwertungsrechte der Datenhalter an das Zentrum eLabour übertragen und durch organisatorische und technische Maßnahmen geschützt. Der Zugang für die wissenschaftliche Nachnutzung der freigegebenen Daten erfolgt nach dem 25. Mai 2018 durch eLabour auf der Grundlage von Nutzungsverträgen, die die wissenschaftlichen Nutzer mit eLabour abschließen.

Rechtliche Anforderungen

Faktisch anonymisierte Forschungsdaten unterliegen prinzipiell nicht mehr dem Datenschutz, sie dürfen automatisiert, d.h. ohne manuelle Prüfung weitergegeben werden. Die Prüfung, ob ein Datensatz als faktisch anonymisiert gelten kann, kann nicht allein IT-basiert erfolgen, sondern bedarf einer manuellen Prüfung und Freigabe durch eine dafür qualifizierte und vom jeweiligen Datenhalter autorisierte Person. Erst danach darf eine automatisierte Weitergabe erfolgen. D.h. jede Form der IT-basierten Anonymisierung erfordert im Einzelfall eine manuelle Überprüfung.

Alle Forschungsdaten, die nicht faktisch anonymisiert sind, sind personenbezogen und damit schutzwürdig, sie müssen sicher gehalten werden und unterliegen definierten Zugriffsbeschränkungen, die im Freigabeprozess festgelegt werden.

Schutz sensibler Daten: die Forschungsdaten enthalten neben dem gesetzlich Vorgeschriebenen Schutz auch sensible Informationen, die aus (forschungs-)ethischen Gründen als schutzwürdig angesehen werden. Diese Daten/Informationen werden am besten bereits von den Primärforschern oder im weiteren Prozess der Datenbearbeitung und Anonymisierung als sensible Daten gekennzeichnet und mit einer entsprechenden Schutz- oder Risikoklasse versehen. Letztere bestimmt darüber, ob und unter welchen Bedingungen darauf zugegriffen werden darf.

Organisationsbezogene Daten z.B. von Unternehmen sind schutzwürdig aufgrund der Wahrung von Betriebsgeheimnissen nach dem Informationsfreiheitsgesetz (IFG) bzw. dem Gesetz gegen den unlauteren Wettbewerb (UWG) und insbesondere aufgrund von schriftlichen oder mündlichen Absprachen der Primärforscher mit den beforschten Unternehmen. Letztere finden im Rahmen von Betriebsfall-

studien generell statt und ihre Einhaltung ist sehr kritisch für die datenhaltenden Institute (Betriebszugang als Geschäftsgrundlage).

Gleichzeitig sind organisationsbezogene Daten für die Sekundärforschung oft unabdingbar, ihre Anonymisierung kann das Forschungsziel verunmöglichen. *Bezogen auf den Umgang mit Organisationsdaten müssen weitergehende Regeln diskutiert und erstellt werden.*

Datenschutzkonzept des Zentrums eLabour

eLabour organisiert den Prozess der Speicherung, Aufbereitung und Bereitstellung der originalen qualitativen Forschungsdaten der beteiligten datenhaltenden Forschungseinrichtungen für die wissenschaftliche Nachnutzung. Das Datenschutzkonzept umfasst die im Folgenden beschriebenen Bearbeitungsprozesse der von den Forschungseinrichtungen eingebrachten Forschungsdaten und die Übergabe von Verwertungsrechten der datenhaltenden Einrichtungen an eLabour im Rahmen eines Freigabeprozesses. Drittens stellt eLabour diese Forschungsdaten für die wissenschaftliche Nachnutzung zur Verfügung, schließt Nutzungsverträge mit Wissenschaftlern ab und organisiert Beratung und Austausch bei der Nutzung.

eLabour stellt eine geeignete Forschungsinfrastruktur, Regeln, Prozesse, sowie organisatorische und technische Maßnahmen bereit, um die personenbezogenen Originaldaten der Datengeber für die wissenschaftliche Nachnutzung gemäß EU Grundverordnung und dem BDSG-neu aufzubereiten (siehe Protokoll des Gesprächs mit den Datenschutzbeauftragten der Uni Göttingen. a. 16.06.2017).

Die Forschungsdatenplattform und das Speicherkonzept mit den technischen Schutzmaßnahmen sind in einem weiteren Dokument im Einzelnen beschrieben (siehe Anhang).

Die Nutzungs- und Verfügungsrechte an den Forschungsdaten liegen bei den an eLabour beteiligten Partnerinstituten. Diese sind Datenhalter der eigenen Daten, unabhängig von den Urheberrechten der Primärforschenden, die durch Zitation zu wahren sind. Die Forschungsdaten werden von den datenhaltenden Instituten in einen internen Speicherbereich des Zentrums eLabour übertragen, der exklusiv dem jeweiligen Datenhalter zur Verfügung gestellt wird. Die Daten werden hier als Backup und Langzeitsicherung in einer sicheren Umgebung so archiviert, dass die Verfügung und Verantwortung der Datenhalter transparent gewährleistet ist. Im Weiteren werden sie im Auftrag des Datenhalters und in enger Zusammenarbeit in die Forschungsdateninfrastruktur des Zentrums eingespeist (Ingest-Prozess) und mit IT-Unterstützung für die Freigabe in das eLabour-Repositorium aufbereitet. Hierfür werden IT-basierte "Anonymisierungs- und Freigabewerkzeuge" und geeignete Workflows für unterschiedliche Stufen und Anforderungen entwickelt und bereitgestellt. Die Originaldaten, die anonymisierten bzw. pseudonymisierten Daten und die entsprechenden Schlüsseldateien sind hierfür auf physisch getrennten Servern bzw. auf getrennten Partitionen zu speichern. Die datenhaltenden Institute tragen in dieser Phase die Verantwortung für die Einhaltung des Datenschutzes und der von ihnen im Primärprojekt eingegangenen Verpflichtungen gegenüber den Betroffenen. Sie werden dabei aber vom Zentrum unterstützt, insbesondere nutzen sie die sichere Infrastrukturumgebung und Kompetenz der IT-Partner.

Das Datenschutzkonzept wird in drei Schritten (Datenübermittlung (Ingest), Datenbearbeitung und Datenfreigabe) umgesetzt, denen jeweils unterschiedlich schutzwürdige Forschungsdaten und Bearbeitungsstufen entsprechen. Ein ausführliches Ablaufmodell, das den gesamten Workflow des For-

schungsdatenmanagements im Rahmen von eLabour umfasst, ist im Anhang (muss noch erstellt werden).

1. Einbringen von Originalforschungsdaten durch die Datenhalter (Ingest)

Die datengebenden Institute speichern personenbezogene Originalforschungsdaten von qualitativen, empirischen Studien in einem eigenen, durch technische Maßnahmen hinreichend gesicherten Speicherbereich, der von eLabour zur Verfügung gestellt wird. Der Zugang zu diesem Speicherbereich ist auf autorisierte Personen des jeweiligen Instituts zum Zwecke der Datenbearbeitung beschränkt.

Die Entscheidung darüber, welche Originaldaten für eine Nachnutzung im Rahmen von eLabour erschlossen werden sollen, liegt bei den Datenhaltern. Im Interesse der späteren Nachnutzung qualitativ hochwertiger, historischer Daten ist es wünschenswert, dass möglichst wenig veränderte Originaldaten gesichert werden. D.h. Originaldaten sind i.d.R. nicht oder nur formal anonymisiert (ohne Klarnamen der Befragten).

Die Originaldaten werden in einen ebenso gesicherten Speicherbereich zur Bearbeitung übertragen und abgelegt. Dabei wird angestrebt, in Originaldaten verfügbare Klarnamen in gesicherter und dokumentierter Weise für zu erhalten, da diese insbesondere für zeithistorische Forschungszwecke von Interesse sind. Dies kann entweder durch Sicherung der Klarnamen oder Sicherung der Schlüsseldateien erfolgen. Diese sind separat von den bearbeiteten Forschungsdaten im gesicherten Bereich der VFU zu speichern, auf den ausschließlich autorisierte Personen Zugriff haben.

Umgang mit originalen Forschungsdaten in der Forschungsumgebung eLabour (VFU):

- die Speicherung sowie Backups erfolgen in verschlüsselter Form
- die Datenübertragung zur Bearbeitung der Daten erfolgt verschlüsselt
- Vergabe der Berechtigung für einen Zugriff auf die Daten erfolgt durch die Datenhalter

Im ersten Schritt sind die Einwilligungserklärungen der Primärstudie zu prüfen. Informationen über das Einwilligungsverfahren der Primärforscher und ggf. vorhandene schriftliche oder mündliche Erklärungen sind sorgfältig zu dokumentieren und den Daten (im Ordner Studienbeschreibung und ggf. auf der Fallebene unter Fallerhebungsinstrumenten) beizufügen. Wenn die von den befragten Personen und Organisationen gegebenen **Einwilligungserklärungen** eine Archivierung und Sekundärnutzung explizit ausschließen (z.B. Zusicherung der Nutzung nur im Primärprojekt und anschließende Löschung), ist eine Archivierung und Sekundärnutzung der Originaldaten nicht möglich.

Um zu entscheiden welche Schutzmaßnahmen und Anonymisierungsschritte notwendig sind, werden die Daten einer *Risikoklassifikation* (siehe Abschnitt 3.) unterzogen, die ein zentrales Instrument des Datenschutzkonzepts darstellt. Die Risikoklassifikation soll zukünftig bereits bei der Datenerhebung durch die Primärforscher erfolgen, bei bereits erzeugten Daten muss sie nachträglich vorgenommen werden. Sie ist eine Aufgabe der Datenhalter, die durch eLabour unterstützt wird.

2. Bearbeitung der Forschungsdaten in eLabour (Datenverarbeitung)

Die originalen Forschungsdaten werden im nächsten Schritt im Rahmen von eLabour im Auftrag und in enger Kooperation mit dem Datenhalter bearbeitet. Gestützt auf IT-Tools werden sie anonymisiert oder pseudonomysiert und mit Metadaten versehen.

Der Zugang zu den personenbezogenen Forschungsdaten (des eigenen oder anderer Institute) im Rahmen der Datenverarbeitung in eLabour kann unterschiedlichen Zwecken dienen:

- der Datensicherung und Langzeitarchivierung
- Durchführung von Freigabe und Anonymisierungs-Workflow für autorisierte MitarbeiterInnen des Datenhalters (Schreibrechte für Originaldaten)
- Anreicherung mit Metadaten, Dokumentation, Ergänzung - Workflow für autorisierte MitarbeiterInnen des Datenhalters und andere (Schreibrechte für Originaldaten)
- Experimentelle Anwendung mit hierfür individuell freigegebenen (personenbezogenen) Arbeitsdaten, die den gleichen Restriktionen unterliegen, wie die Originaldaten; (keine Schreibrechte auf die Originaldaten); sie werden in eigenen Arbeitsspeicherbereichen der bearbeitenden WissenschaftlerInnen gehalten

Diese Bearbeitungsschritte erfordern i.d.R. den Zugriff auf eine Kopie der originalen Forschungsdaten. Dieser Zugriff wird durch Vertraulichkeitserklärungen mit den individuellen Wissenschaftlern geregelt. Die Verarbeitung dieser Daten erfolgt in einem privaten/abgeschirmten Bereich der VFU (privater Bereich):

Umgang in der zentralen eLabour-IT mit diesen originalen Forschungsdaten erfordert:

- Rechte- und Rollenmodell, das den Zugriff auf die Datensätze regelt und protokolliert.
- Speicherung und Backups erfolgen ausschließlich verschlüsselt
- die Datenübertragung erfolgt ausschließlich verschlüsselt
- Datenderivate, wie Suchindexe und Stichwortlisten aus diesem Datenbestand unterliegen einer gleichen Risikostufe der (erweiterten /bearbeiteten) Originaldaten.
- Innerhalb von eLabour wird die Verarbeitung von Klartexten mittels Vertraulichkeitserklärungen ermöglicht
- die Bearbeitung von Klartexten erfolgt in geschützten Bereichen der eLabour VFU.

Auch die im Rahmen der Bearbeitung der Forschungsdaten erzeugten *Metadaten* können in Bezug auf Datenschutz sensible Informationen enthalten. Dies betrifft insbesondere die Bezeichnungen von untersuchten Organisationen (typischerweise z.B. Firmennamen). Im Rahmen der VFU werden diese Daten durch geeignete organisatorische (Nutzungsverträge) und technische Maßnahmen (sensible Metadaten werden abgestimmt nach Nutzerrollen aus- oder eingeblendet) geschützt. Im Rahmen der Anonymisierung werden auf dieser Grundlage die notwendigen Maßnahmen durchgeführt und dokumentiert. Im Freigabeprozess wird festgelegt, welche Nutzergruppen mit welchen Auflagen Zugang zu den Daten erhalten können. Diese Festlegungen werden IT-basiert umgesetzt und die Zugriffe dokumentiert.

Für die wissenschaftliche Nutzung in Sekundärprojekten müssen die Forschungsdaten den Anforderungen des Datenschutzes für die Weitergabe entsprechen, d.h. anonymisiert werden. Nur faktisch anonymisierte Daten enthalten keine personenbezogenen Daten mehr und können für wissenschaftliche Zwecke uneingeschränkt genutzt werden. Allerdings sind sie nicht für alle Fragestellungen von hinreichender Qualität. Daher streben wir eine Form der Anonymisierung an, die den erforderlichen Schutz Personen und Organisationen vom konkreten Schadensrisiko ableitet. Denn formal anonymisierte und/oder pseudonymisierte Forschungsdaten ermöglichen analytisch gehaltvollere Sekundäranalysen, erfordern allerdings zusätzliche organisatorischen und technischen Maßnahmen und Nutzungsauflagen. Diese werden in einem Nutzungsvertrag festgehalten, den alle Nutzer abschliessen müssen.

Die Risikoklassifikation erfolgt in einem zweistufigen Prozess, in dem sowohl das mögliche Schadensrisiko (Welcher Schaden kann durch die konkreten personenbezogen Daten für die betroffenen Personen bzw. Organisationen entstehen?) als auch das Risiko einer De-Anonymisierung bzw. Re-Identifikation (Wie wahrscheinlich ist es, dass die betroffenen Personen bzw. Organisationen identifiziert werden?) eingeschätzt werden. Die erste Risikoklassifikation zielt darauf ab, vorläufige Freigabeempfehlungen auszusprechen und die dafür ggf. notwendigen Anonymisierungsschritte festzulegen. (siehe Abschnitt 3).

In der Regel werden die von den Datenhaltern übermittelten Originaldaten nicht hinreichend anonymisiert sein. Entsprechend der Risikoklassifikation werden skalierbare Anonymisierungsmaßnahmen durchgeführt, für die eLabour technische Tools und Prozesse und Unterstützung bereitgestellt. Dieser Bearbeitungsprozess erfolgt in der Forschungsumgebung von eLabour durch oder im Auftrag der Datenhalter, die ihre originalen, personenbezogenen Forschungsdaten in eine eigene, sichere Speicherumgebung in eLabour einbringen. Hierbei ist zudem zu prüfen, ob im Rahmen von Einwilligungserklärungen den befragten Personen oder untersuchten Organisationen konkrete Zusagen hinsichtlich der Form der Anonymisierung gemacht wurden.

Regeln und Vorschläge wie genau faktisch anonymisiert bzw. pseudonymisiert werden soll, werden den Regeln zur Anonymisierung im Anhang festgelegt. Dabei soll der Grundsatz gelten: *So viel Informationen wie möglich für die qualitative Sekundäranalyse erhalten und so viel Informationen wie nötig durch Pseudonymisierung oder Anonymisierung schützen.* Bei der Festlegung von Form und Reichweite der Anonymisierung ist die notwendige Datenqualität für die wissenschaftliche Forschung und das Risiko der Identifikation von Personen und Organisationen so gegeneinander abzuwägen, dass das Schadensrisiko gering ist.

Im Rahmen der Datenbearbeitung in eLabour wird it-gestützt eine formale Anonymisierung durchgeführt (Klarnamen von Personen werden entfernt). Für Daten, die älter als 30 Jahre sind und unter Archivrecht fallen, kann auf eine formale Anonymisierung verzichtet werden, sofern keine anderen Gründe (Ansprüche der Befragten oder ihrer Erben) dagegensprechen. Die Bearbeitungsschritte zur Sicherung des Datenschutzes erfolgt im geschützten Speicherbereich, es sollen aber grundlegende Informationen zur Existenz von nicht weitergegebenen Daten in standardisierter Weise z.B. in Metadaten verfügbar sein (wobei festgelegt werden muss, wer was sehen darf).

Nach der Risikoklassifikation und ggf. weitergehenden Bearbeitungsschritten kann der Freigabeprozess erfolgen. Aus der Kombination der Risikoklassifikation, dem Grad der Anonymisierung und den technischen und organisatorischen Maßnahmen zum Schutz personenbezogener Daten ergeben sich konkrete Freigabemöglichkeiten. Im Freigabeprozess werden die Datenschutzanforderungen überprüft und Festlegungen zu Nutzungsmöglichkeiten bzw. deren Einschränkung getroffen.

2. Risikoklassifikation

Alle Forschungsdaten und deren Derivate werden in eLabour einer Risikobewertung unterzogen und dentsprechend klassifiziert, sowie anschliessend in Abhängigkeit den konkreten Risikomerkmalen anonymisiert und im Rahmen des Freigabeprozesses mit sich daraus ergebenden Maßnahmen geschützt. Hierzu wird in eLabour eine mehrstufige Risikobewertung der Daten vorgenommen und im Freigabeprozess in Zugangsrechte und Nutzungsauflagen umgesetzt.

Die Risikobewertung erfolgt hierzu zum ersten bei der Überführung der Daten von den Primärforschungen in eLabour und zum zweiten nach erfolgter der Durchführung von Anonymisierungsmaß-

nahmen (siehe A4). Ausgangsgrundlage der ersten Risikobewertung sind die Originalforschungsdaten bzw. formal anonymisierte Forschungsdaten (wenn nicht datenschutzrechtliche oder andere Gründe aus Sicht des Datenhalters dagegen sprechen.)

Ziel ist einerseits die Bewertung der Schadensrisiken für die untersuchten Personen und Organisationen, die entstehen würden, wenn der Inhalt der Dokumente mit der Person verknüpft (de-anonymisiert) und weitergegeben würde. Dabei ist zu unterscheiden zwischen der Weitergabe für wissenschaftliche Zwecke und der öffentlichen Verbreitung (Veröffentlichung). Andererseits wird zudem das Risiko (bzw. die Wahrscheinlichkeit) einer Re-Identifikation bzw. De-Anonymisierung bewertet. Dieses Risiko kann anschließend durch datenverändernde Maßnahmen zum Zwecke der Anonymisierung reduziert werden.

Der erste Schritt der Risikobewertung erfolgt durch das datenhaltende Institut, im optimalen Fall unmittelbar durch die Primärforscher, oder durch die Wissenschaftler, die die Daten sichten und für die Sekundärnutzung aufbereiten. Hierfür wird ein Fragebogen verwendet, der standardisierte und offene Angaben enthält (siehe Anhang 2). Die ausgefüllten Fragebögen sind analog zu den Daten selbst zu schützen und dienen im weiteren den für die Anonymisierung und Freigabe verantwortlichen Personen als Entscheidungsgrundlage. Zusätzlich wird eine standardisierte Bewertung vorgenommen und dokumentiert, diese ist mit den Metadaten für Nutzer zugänglich, damit diese Daten auswählen können, die ihnen zugänglich sind und ggf. den Zugang zu besonders geschützten Daten beantragen können.

Die Risikobewertung ist zwingend für jede Studie vorzunehmen, sie gilt für alle Dokumente der Studie, die keine eigene, höhere Risikobewertung haben. Zusätzlich kann auch auf der Ebene der Fälle und/oder für Typen von Dokumenten (z.B. Expertengespräche) ein höheres Risiko definiert werden, wenn nötig können auch einzelne Dokumente eine von der Gesamtstudie abweichende Risikobewertung erhalten (Granularität der Bewertung). Das höhere Risiko ist immer ausschlaggebend. D.h. die Fälle, Dokumenttypen oder Dokumente, für die ein höheres Risiko festgestellt wird, werden aus der allgemeinen Bewertung der Studie herausgenommen und gesondert bewertet.

Gegenstand der Risikobewertung sind die folgenden Merkmale

S1.1 Grad des Schadensrisikos für die Person

1. kein (besonderes) Schadensrisiko
2. mittleres Schadensrisiko
3. hohes Schadensrisiko
4. sehr hohes Schadensrisiko

S1.2 Art des Schadensrisikos für die Person (Mehrfachantworten möglich)

1. Beeinträchtigungen organisationsinterner Beziehungen bzw. des unmittelbaren sozialen Nahbereichs der Person (bspw. Kritik an Kolleginnen und Kollegen oder an Vorgesetzten, Äußerungen zum Lebenspartner, Sachbearbeiter)
2. Beeinträchtigung der gesellschaftlichen und öffentlichen Stellung; **“Ansehen“** der Person (Medien, Öffentlichkeit)
3. Beeinträchtigung der wirtschaftlichen Verhältnisse und organisationale Sanktionen; „Existenz“
4. staatliche Sanktionen, Gefahr für Freiheit

S1.3 offene Begründung des Schadensrisikos für Personen

S2.1 Grad des Schadensrisikos für die Organisation

1. kein (besonderes) Schadensrisiko
2. mittleres Schadensrisiko
3. hohes Schadensrisiko
4. sehr hohes Schadensrisiko

S2.2 Art des Schadensrisikos für die Organisation (Mehrfachantworten möglich)

1. Beeinträchtigung des Organisationsfrieden (bspw. Bekanntwerden der Bevorteilung einzelner, Restrukturierungen)
2. Beeinträchtigung des politische und öffentlichen „Ansehens“ der Organisation
3. Verrat von Betriebs- und Geschäftsgeheimnissen Beeinträchtigung der (wirtschaftlicher Schaden)

S.2.3 offene Begründung des Schadensrisikos für die Organisation

S3 besonders schutzwürdige personbezogenen Daten

Rechtlich relevante Kategorien besonders schutzwürdiger personenbezogener Informationen sind: rassische, ethnische Herkunft, politische Meinungen, religiöse, weltanschauliche Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheitsdaten, sexuelle Orientierung

1. keine Daten der schutzwürdigen personenbezogenen Kategorien enthalten
2. besonders schutzwürdige personenbezogene Daten enthalten

S4.1 Erhebungszeitraum: Jahr(e) der Erhebung)

S4.2 Bewertung des Erhebungszeitraums (wird in Zeitintervall od. bei Nutzungsanfrage angepasst)

1. laufende Studie
2. bis 10 Jahre (erhöhtes Re-Identifikationsrisiko)
3. jünger als 20 bis 11 Jahre
4. jünger als 30 bis 20 Jahre (geringes Re-Identifikationsrisiko)
5. 30 Jahr und älter - zu (vernachlässigendes Risiko (Archivrecht)

Grundsätzlich gilt: Sollten Ausprägungen einzelner Risikomerkmale oder die gesamte Risikobewertung unklar sein, ist stets von einem sehr hohen Risiko auszugehen.

Bewertung des Re-Identifikationsrisikos

R1 Re-Identifikationsrisiko für die untersuchten Personen

1. Kein Risiko
2. Gering
3. Mittel
4. Hoch, identifizierende persönliche Merkmale (z.B. Ortsnamen, Rollen)
5. Hoch, organisationsspezfische Merkmale (Funktionen, andere Personen)
6. Hoch, persönliche Merkmale sind öffentlich (Medien, Netz)
7. Hoch, unspezifisch

R2 Re-Identifikationsrisiko für die untersuchte Organisation

1. kein Risiko
2. gering
3. mittel
4. hoch

Aus den jeweils vorliegenden Ausprägungen der Risikomerkmale und ihrer Kombination ergeben sich die Notwendigkeit von ggf. weiteren Anonymisierungsschritten sowie die zu treffenden organisatorischen und technischen Maßnahmen des Datenschutzes im Rahmen der Freigabe. Dabei gilt grundsätzlich, dass ein bestehendes Re-Identifikationsrisiko umso problematischer anzusehen ist, je höher das Schadensrisiko ist (dementsprechend sind auch umso weitgehendere Anonymisierungsmaßnahmen umzusetzen).

3. Freigabeklassen

Aus der Risikobewertung ergibt sich ein Spektrum von Freigabemöglichkeiten. Für jede Studie und ggf. für Fälle, Dokumententypen und Dokumente (je nach Granularität der Risikobewertung) werden bestimmte Freigaben festgelegt. Der Nutzungsvertrag beinhaltet über alle Freigabeklassen ein striktes Verbot der Re-Identifikation und der Veröffentlichung persönlicher Merkmale der Personen in der Studie. Durch den Nutzungsvertrag hat das Repositorium eLabour die Kontrolle darüber, wer für welche Forschungszwecke das Material nutzt. Mit dem Nutzungsvertrag und der damit verbundenen kontrollierten Datennutzung wird das Restrisiko der nicht faktisch anonymisierten Forschungsdaten und deren unbefugter Verwendung abgefangen.

Es sind sechs differenzierte Freigabeklassen mit unterschiedlichem Anonymisierungsgraden und Zugangsmöglichkeiten mit definierten technischen und organisatorischen Maßnahmen vorgesehen. Die Reihenfolge der sechs Freigabeklassen und der darin erfassten Risikomerkmale ergibt sich aufsteigend aus dem Schadensrisiko für Personen und Organisationen.

Freigabeklasse	*Anonymisierungsgrad*	*Zugang mit Nutzungsvertrag*
I. Archivrecht > 30 Jahre	formal (od. original)	Volltext, Download
II. geringes Risiko, > 10J.	formal	Volltext, Download
III. mittleres Risiko, > 10J.	formal	Volltext, Download mit Auflagen
IV. mittleres Risiko, < 10J.	pseudonymisiert	Volltext, Download mit Auflagen
V. hohes Risiko	pseudonymisiert	Volltext, Download mit hohen Auflagen und Kontrolle
Vi. sehr hohes Risiko	kein Zugang od. faktisch anonymisiert	

Freigabeklasse I

Daten in Freigabeklasse I können zu wissenschaftlichen Zwecken frei weitergegeben werden. Es handelt sich um Material, das älter als 30 Jahre und formal anonymisiert ist oder jünger als 30 Jahre und faktisch anonymisiert ist.

Ausgenommen hiervon sind Daten, die 30 Jahre und älter sind, aber für die betroffenen Person oder Organisationen ein hohes oder sehr hohes Schadensrisiko in sich tragen. Diese Daten unterliegen den Zugangsbeschränkungen der Freigabeklasse IV. Zudem sind hiervon ausgenommen Daten die 30 Jahre oder älter sind, nicht formal anonymisiert sind und keine besonderen Risiken enthalten. In diesem Fall können die Originaldaten mit Klarnamen zugänglich gemacht werden, wenn die formale Anonymisierung mit hohem Aufwand verbunden ist und der Datenhalter einer Freigabe der Originaldaten zustimmt.

Tabelle 1: Freigabeklasse I

Freigabebedingungen		
	Freigabeklasse Ia	**Freigabeklasse Ib**
Zugangsmöglichkeit	Im Volltext in der VFU durchsuchbar, Download	Im Volltext in der VFU durchsuchbar, Download
Datenverändernde Maßnahmen	Faktische Anonymisierung	Formale Anonymisierung
Organisatorische Maßnahmen	Nutzungsvertrag (Nutzung nur zu wiss. Zwecken)	Nutzungsvertrag (Nutzung nur zu wiss. Zwecken)
Zugrundliegende Risikobewertung		
	Freigabeklasse Ia	**Freigabeklasse I1b**
Risikomerkmale	**Ausprägung**	**Ausprägung**
S1.1 Grad des Schadensrisikos für die Person		2 mittleres Schadensrisiko
S2.1. Grad des Schadensrisikos Organisation		2 mittleres Schadensrisiko
S3 besonders schutzwürdige personenbezogene Daten		nicht vorhanden oder vorhanden
S4 Erhebungszeitraum	2-5, jünger als 30 Jahre	1 (30 Jahre und älter)
R1 Re-Identifikationsrisiko der Person		3 mittleres Re-Identifikationsrisiko
R1 Re-Identifikationsrisiko für die Organisation		3 mittleres Re-Identifikationsrisiko

Freigabeklasse II

Freigabeklasse II umfasst Daten, die älter als 10 Jahre und bis 29 Jahre sind. Diese Daten beinhalten kein bis ein geringes Re-Identifizierungsrisiko für Personen und Organisationen und kein (besonderes) Schadensrisiko für die befragten Personen und Organisationen. Diese Daten können formal anonymisiert zur Nutzung zu wissenschaftlichen Zwecken weitergegeben werden, mit den Auflagen Verbot der Datenweitergabe an unberechtigte Dritte und Verbot der Identifikation der Befragten.

Tabelle 2: Freigabeklasse II

Freigabebedingungen	
Zugangsmöglichkeit	Im Volltext durchsuchbar in der VFU, Download
Datenverändernde Maßnahmen	Formale Anonymisierung
Organisatorische Maßnahmen	Nutzungsvertrag (Nutzung nur zu wiss. Zwecken, Verbot der Weitergabe, Re-Identifikationsverbot)

Zugrundliegende Risikobewertung	
Risikomerkmale	**Ausprägung**
S1.1 Grad des Schadensrisikos für die Person	1 kein (besonderes) Schadensrisiko
S2.1. Grad des Schadensrisikos Organisation	1 kein (besonderes) Schadensrisiko
S3 besonders schutzwürdige personenbezogene Daten	nicht vorhanden oder vorhanden
S4 Erhebungszeitraum	3-5 (11 bis 29 Jahre)
R1 Re-Identifikationsrisiko der Person	1 kein Risiko oder 2 gering
R1 Re-Identifikationsrisiko für die Organisation	1 kein Risiko oder 2 gering

Freigabeklasse III

Freigabeklasse 3 umfasst Daten die älter als 10 Jahre sind. Diese Daten beinhalten ein mittleres Re-Identifizierungsrisiko für Personen und Organisationen und ein mittleres Schadensrisiko für die befragten Personen und Organisationen. Diese Daten können formal anonymisiert zur Nutzung zu wissenschaftlichen Zwecken weitergeben werden, mit den Auflagen Verbot der Datenweitergabe an unberechtigte Dritte und dem Verbot der Identifikation der Befragten. Zudem sind illustrierende Interviewausschnitte vorab einer Veröffentlichung eLabour zur Prüfung vorzulegen. Auch sind Insider (z.B. Personen aus dem Umfeld der Befragten bzw. Organisation) von der Datennutzung ausgeschlossen.

Tabelle 3: Freigabeklasse 3

Freigabebedingungen	
Zugangsmöglichkeit	Im Volltext durchsuchbar in der VFU, Download
Datenverändernde Maßnahmen	Formale Anonymisierung
Organisatorische Maßnahmen	Nutzungsvertrag (Nutzung nur zu wiss. Zwecken, Verbot der Weitergabe, Re-Identifikationsverbot) – Kontrolle illustrierender Interviewausschnitte vor Veröffentlichung – Ausschluss von Insidern
Zugrundliegende Risikobewertung	
Risikomerkmale	**Ausprägung**
S1.1 Grad des Schadensrisikos für die Person	2 mittel
S1.2 Art des Schadensrisikos für die Person	1 Beeinträchtigungen organisationsinterner Beziehungen bzw. des unmittelbaren sozialen Nahbereichs der Person
S2.1. Grad des Schadensrisikos für die Organisation	2 mittel
S2.2 Art des Schadensrisikos für die Organisation	1 Beeinträchtigung des Organisationsfriedens
S3 besonders schutzwürdige personenbezogene Daten	vorhanden
S4 Erhebungszeitraum	11 bis 30 Jahre
R1 Re-Identifikationsrisiko der Person	3 mittel
R1 Re-Identifikationsrisiko für die Organisation	3 mittel

Freigabeklasse IV

Freigabeklasse 4 umfasst Daten bis 10 Jahre. Diese Daten beinhalten ein mittleres Re-Identifizierungsrisiko für Personen und Organisationen und ein mittleres Schadensrisiko für die befragten Personen und Organisationen. Diese Daten können pseudonymisiert zur Nutzung zu wissenschaftlichen Zwecken weitergeben werden, mit den Auflagen Verbot der Datenweitergabe an unberechtigte Dritte und dem Verbot der Identifikation der Befragten. Die besondere Schutzwürdigkeit der Daten, der durch die Pseudonymisierung Rechnung getragen wird, ergibt sich aus dem geringen Alter der Daten. Zudem sind illustrierende Interviewausschnitte vorab einer Veröffentlichung eLabour zur Prüfung vorzulegen. Auch sind Insider von der Datennutzung ausgeschlossen.

Tabelle 4: Freigabeklasse IV

Freigabebedingungen	
Zugangsmöglichkeit	Im Volltext durchsuchbar in der VFU, Download
Datenverändernde Maßnahmen	Pseudonymisierung
Organisatorische Maßnahmen	Nutzungsvertrag (Nutzung nur zu wiss. Zwecken, Verbot der Weitergabe, Re-Identifikationsverbot) – Kontrolle illustrierender Interviewausschnitte vor Veröffentlichung – Ausschluss von Insidern
Zugrundliegende Risikobewertung	
Risikomerkmale	**Ausprägung**
S1.1 Grad des Schadensrisikos für die Person	2 mittel
S1.2 Art des Schadensrisikos für die Person	1 Beeinträchtigungen organisationsinterner Beziehungen bzw. des unmittelbaren sozialen Nahbereichs der Person
S2.1. Grad des Schadensrisikos für die Organisation	2 mittel
S2.2 Art des Schadensrisikos für die Organisation	1 Beeinträchtigung des Organisationsfriedens
S3 besonders schutzwürdige personenbezogene Daten	vorhanden
S4 Erhebungszeitraum	1 bis 10 Jahre
R1 Re-Identifikationsrisiko der Person	3 mittel
R1 Re-Identifikationsrisiko für die Organisation	3 mittel

Freigabeklasse V

Freigabeklasse V umfasst Daten bis 30 Jahre, die ein hohes Schadensrisiko für die befragten Personen und Organisationen bei deren Re-Idenifikation beinhalten. Diese Daten können pseudonymisiert zur Nutzung zu wissenschaftlichen Zwecken weitergeben werden, mit den Auflagen Verbot der Datenweitergabe an unberechtigte Dritte und dem Verbot der Identifikation der Befragten. Auch ist von den Nutzenden ein Datensicherungskonzept vorzulegen, in dem die technischen Datensicherungsmaßnahmen bei den Nutzenden gegenüber eLabour dokumentiert und zugesichert werden. Zudem muss vorab der Datennutzung eine Datenschutzschulung durch eLabour erfolgen. Auch sind illustrierende Interviewausschnitte vorab einer Veröffentlichung eLabour zur Prüfung vorzulegen und sind Insider von der Datennutzung ausgeschlossen.

Tabelle 5: Freigabeklasse V

Freigabebedingungen	
Zugangsmöglichkeit	Im Volltext durchsuchbar in der VFU, Download
Datenverändernde Maßnahmen	Pseudonymisierung
Organisatorische Maßnahmen	Nutzungsvertrag: Nutzung nur zu wiss. Zwecken, Verbot der Wetergabe, Re-Identifikationsverbot; – Datensicherheitskonzept beim Sekundärnutzer = Nutzung nur in gesicherter Umgebung; – Vorlage des Datensicherheitskonzeptes durch die Nutzenden, – Kontrolle von Interviewausschnitten vor Veröffentlichung, – Ausschluss von Insidern – Schulung zum Datenschutz
Zugrundliegende Risikobewertung	
Risikomerkmal	**Ausprägung**
S1.1 Grad des Schadensrisikos für die Person	3 hoch, 4 sehr hoch
S1.2 Art des Schadensrisikos für die Person	2 Beeinträchtigung gesellschaftliches Ansehen oder 3 Beeinträchtigung wirtschaftlicher Verhältnisse
S2.1. Grad des Schadensrisikos Organisation	4 Sehr hoch
S2.2 Art des Schadensrisikos für die Organisation	1 Beeinträchtigung des Organisationsfriedens oder 2 Beeinträchtigung des politische und öffentlichen „Ansehens“ der Organisation
S3 besonders schutzwürdige personenbezogene Daten	vorhanden
S4 Erhebungszeitraum	2-4 1 bis 30 Jahre
R1 Re-Identifikationsrisiko der Person	1 kein Risiko oder 2 gering (nach datenverändernden Maßnahmen)
R1 Re-Identifikationsrisiko für die Organisation	1 kein Risiko oder 2 gering (nach datenverändernden Maßnahmen

Freigabeklasse VI

Freigabeklasse 6 umfasst Daten die nach der Risikobewertung ein hohes bzw. sehr hohes Schadensrisiko für die befragten Personen und Organisationen sowie ein hohes Re-Identifizierungsrisko in sich tragen. Eine Einstufung in Freigabeklasse VI und damit ein Verzicht auf jegliche Freigabe erfolgt dann, wenn die auf Basis der Risikobewertung notwendigen datenverändernden Maßnahmen (Pseudonymisierung) nicht erfolgen können, da diese eine sinnvolle Analyse und Interpretation des Materials unmöglich machen würden. Oder wenn sie mit einem sehr hohen Aufwand verbunden sind, der mit den gegebenen Ressourcen nicht leistbar ist. Daten der Freigabeklasse 6 werden für eine Freigabe außerhalb von eLabour nicht freigegeben.

4. Freigabe und Übertragung von Verwertungsrechten an den Forschungsdaten von den Datenhaltern an eLabour

Am Ende des Freigabeprozess übergeben die datenhaltenden Institute Verwertungsrechte und Funktionen und Datenschutzverpflichtungen an eLabour, damit eLabour die freigegebenen Daten für die wissenschaftliche Nachnutzung zur Verfügung stellen kann. Diese **Funktionsübertragung** von den

Datenhaltern an eLabour wird vertraglich geregelt und erfolgt im Rahmen eines durch die Infrastruktur unterstützten, gesicherten und dokumentierten Freigabe-Prozesses.

Vor der endgültigen Freigabe der Datenhalter wird die Risikoklassifikation noch einmal geprüft und ggf. verändert. Die Prüfung bezieht sich auf die Frage, ob die bezogen auf die Originaldaten festgestellten Schadens- und Identifikationsrisiken durch die durchgeführten Anonymisierungsschritte beseitigt werden konnten. D.h. die Gesamtbewertung wird unter Berücksichtigung der durchgeführten datenveränderten Maßnahmen zur Anonymisierung erneut durchgeführt. Auf der Grundlage dieser zweiten Risikoklassifikaiton wird über Freigabemöglichkeiten, sowie über notwendige technische und organisatorische Schutzmaßnahmen für personenbezogene Daten entschieden.

Mit der Übertragung der Daten an eLabour und deren Freigabe zur wissenschaftlichen Nutzung übertragen die beteiligten Partnerinstitute dem Zentrum eLabour Verwertungsrechte an den Daten, die mit den einzelnen Partnerinstituten in gesonderten **Übertragungsverträgen** geregelt werden. Sie bleiben aber weiterhin Datenhalter, behalten daher dauerhafte, grundlegende Rechte an der Verwertung ihrer Forschungsdaten. Im Interesse einer verlässlichen Nutzung sollen allerdings die Möglichkeiten der Datenhalter, die vergebenen Nutzungsrechte einseitig zurück zu nehmen, vertraglich geregelt und beschränkt werden.

Der Zugang der wissenschaftlichen Nutzer zu den Forschungsdaten wird im Weiteren durch eLabour organisiert. Gestützt auf die eLabour Infrastruktur wird dabei sichergestellt, dass die in der Risikoklassifikation und im Freigabeprozess festgelegten Zugangsmöglichkeiten und die technischen und organisatorischen Maßnahmen eingehalten werden.

Die Datengeber schließen mit dem Forschungszentrum elabour einen Vertrag zur Nutzung der freigegebenen Daten ab, der die Verfügungs- und Nutzungsrechte von eLabour regelt. Mit der Freigabe werden die Daten an das Forschungszentrum eLabour im Rahmen dieser Vereinbarung übergeben (Information Storage T-II). Der Freigabeprozess wird IT-basiert, transparent und dokumentiert im Rahmen der gesicherten eLabour Infrastruktur in der Verantwortung des Datenhalters durchgeführt. Die freigegebenen Daten werden mit Nutzungsauflagen versehen, die sich aus der Risikobewertung ergeben. Das Forschungszentrum eLabour sichert die strikte Einhaltung dieser Nutzungsauflagen durch geeignete IT-basierte Prozesse. Mit der Freigabe geht die Verantwortung für die Einhaltung des Datenschutzes an das Forschungszentrum eLabour über.

Im Freigabeprozess werden die Kriterien und Regeln auf die bearbeiteten Forschungsdaten der datengebenden Institute angewandt, um diese an das Zentrum eLabour zur Bereitstellung für die wissenschaftliche Nutzung zugänglich zu machen. Die freizugebenden Forschungsdaten sind formal anonymisiert und in sensiblen Merkmalen pseudonymisiert, aber nicht umfassend faktisch anonymisiert, d.h. sie dürfen nur mit Einschränkungen (siehe Freigabemodell und Nutzungsauflagen), Zugangskontrolle und Dokumentation zur eingeschränkten, wissenschaftlichen Nachnutzung an vertraglich zur Einhaltung von Anonymität verpflichtete Nutzer weitergegeben werden. Die erforderlichen Zugangs- und Nutzungseinschränkungen sowie die Freigabemöglichkeiten ergeben sich aus der Bewertung des Risikos für die in den empirischen Studien untersuchten Personen und Organisationen.

Anhang

Rechtliche Probleme der Weitergabe der Forschungsdaten für die wissenschaftliche Nachnutzung und öffentliche Interesse an der wissenschaftlichen Nachnutzung der qualitativen, arbeitssoziologischen Forschungsdaten

Die Weitergabe besonderer personenbezogener Daten für neue wissenschaftliche Forschungszwecke ist wird nach der EU-DSGVO und dem BDSG-Neu, vergleichbar dem BDSG-Alt aus zwei Gründen möglich sein:

1. Die Befragten geben ihre Einwilligung das Material auch über das originäre Forschungsvorhaben hinaus für neue wissenschaftliche Forschungszwecke nutzen zu dürfen (Einwilligungsgrundsatz
2. In Abwägung mit den Schutzinteressen der Befragten besteht ein überwiegendes Interesse an der Datennutzung, das eine Zweckänderung der Datennutzung erlaubt.

Im Überblick für die in eLabour eingebrachten Forschungsdaten liegen i.d.R. keine Einwilligungserklärung der Befragten vor (Studien die zeitlich vor 1975 und damit vor Inkrafttreten des Bundesdatenschutzgesetzes erfolgten) bzw. schließen die vorliegenden Einwilligungen keine Nutzung für andere als den originären Forschungszweck ein.

Begündung des öffentlichen Interesses an der Nachnutzung qualitativer Forschungsdaten aus arbeitssoziologischen, mit öffentlichen Mitteln geförderten Studien

Nach eingehender Prüfung besteht an der Datennutzung und nachhaltigen Bereitstellung der in eLabour eingebrachten und aufbereiteten Daten ein überwiegend öffentliches Interesse, d.h. die Nutzbarmachung der in eLabour eingebrachten Daten und Untersuchung des Wandels von Arbeit hat eine übergeordnete gesellschaftliche und wissenschaftliche Bedeutung, die sich auf fünf zentrale Argumente stützt:

1. Zielsetzung des BMBF ist es, Forschungsdaten der Allgemeinheit der Wissenschaft für die Weiternutzung verfügbar zu machen. Im Kontext dieser Zielstellung erfolgt auch die Förderung des Projektvorhabens eLabour durch das BMBF (https://www.bmbf.de/foerderungen/bekanntmachung-804.html).
2. Die Öffnung und Bereitstellung von Forschungsdaten für die allgemeine wissenschaftliche Nutzung (auch qualitativer Forschungsdaten) ist Zielsetzung der Deutschen Forschungsgemeinschaft (DFG)[1], des Bundesministeriums für Bildung und Forschung (BMBF)[2], der zwei größten Forschungsförderer in den Sozialwissenschaften, und in ähnlicher Form auch kleiner Forschungsförderer wie der Hans-Böckler-Stiftung (HBS).[3]
3. Der Öffnung wissenschaftlicher Daten für die allgemeine wissenschaftliche Öffentlichkeit ist fachübergreifender Konsens der gesamten wissenschaftlichen Gemeinschaft und festgeschrieben in der Erklärung der „Guten wissenschaftlicher Praxis“ der DFG

[1] Deutsche Forschungsgemeinschaft (DFG) (2015): Leitlinien zum Umgang mit Forschungsdaten.

[2] Bundesministerium für Bildung und Forschung (BMBF). 2012. Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschung im Bereich „Sprachliche Bildung und Mehrsprachigkeit“.

[3] Hans-Böckler-Stiftung (HBS). 2016. Hinweise für die Einreichung von Anträgen auf Forschungsförderung bei der Hans-Böckler-Stiftung.

(http://www.dfg.de/sites/flipbook/gwp/#/6/). Unterstützung für die nachhaltige Archivierung und Bereitstellung von Forschungsdaten kommt zudem auch von Seiten des deutschen Wissenschaftsrats[4], Allianz der deutschen Wissenschaftsorganisationen[5] und dem BMBF geförderten Rat für Sozial- und Wirtschaftsdaten (RatSWD).

4. Die Öffnung wissenschaftlicher Daten für die allgemeine Wissenschaftsöffentlichkeit wird auch auf europäischer Ebene, durch die Europäische Kommission gefördert im Rahmen des europäischen Programms „Horizon 2020“ (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

5. Das in eLabour verfolgte Forschungsinteresse der Untersuchung der Entwicklung von Arbeitsbedingungen durch den Wandel von Arbeit ist nicht vor einem wissenschaftlichen Interesse, sondern auch von einem gesamtgesellschaftlichen Interesse, da Arbeitsbedingungen und der Wandel von Arbeit auch auf Lebensbedingungen der Menschen insgesamt wirken. Denn Erwerbsarbeit ist ein wesentlicher Faktor für große Teile der Bevölkerung und damit für die Öffentlichkeit insgesamt von Interesse.

Begriffe

1. **Forschungsdaten:** Als Forschungsdaten werden alle Daten bezeichnet, die während des Forschungsprozesses entstehen oder sein Ergebnis sind. Sie werden abhängig von der Forschungsfrage und unter Anwendung verschiedener Methoden erzeugt bzw. gesammelt, bearbeitet, analysiert und schließlich publiziert und/oder archiviert. Für die Bereitstellung und Nachnutzung von Forschungsdaten ist es notwendig, den Entstehungskontext und die benutzten Werkzeuge zu dokumentieren.
   Qualitative Forschungsdaten, d.h. Daten, die mit überwiegend offenen qualitativen sozialwissenschaftlichen Methoden erzeugt werden, zeichnen sich dadurch aus, dass die untersuchten Personen selbst entscheiden, welche sensiblen Informationen sie, wann und wie einbringen. Dies macht u.a ihre besondere Qualität aus. Daher kann das Risiko nicht routinemäßig festgestellt, sondern muss im Einzelfall geprüft werden.

2. **Personenbezogene Daten:** Gemäß § 3 Abs. 1 BDSG sind personenbezogene Daten Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person. Natürliche Personen sind rechtsfähige Menschen, d.h. jeder Mensch ab Vollendung seiner Geburt. Persönliche Verhältnisse sind Merkmale und Charaktereigenschaften einer natürlichen Person. Dies sind bspw. Namen und Adressdaten, äußere Merkmale (Geschlecht, Augenfarbe, Größe, Gewicht) und innere Merkmale (Meinungen, Motive, Wünsche, Überzeugungen, Werturteile) sowie Angaben wie Familienstand, Geburtsdatum, Staatsangehörigkeit, Beruf, Ausbildungsstand, Erscheinungsbild, Leistungen, Arbeitsverhalten. Einen besonderen Schutz definiert das Datenschutzgesetz für *sensible Daten*, wie rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen oder die Gewerkschaftszugehörigkeit, sowie für Daten die Angaben über die Gesundheit oder Sexualleben einer natürlichen Person beinhalten. Angaben zu sachliche Verhältnisse beziehen sich insbesondere auf Informationen zu den Vermögens- und Eigentumsverhältnisse, Kommunikations- und Vertragsbeziehungen und alle sonstigen Beziehungen einer natürlichen Person zu ihrer Umwelt.

[4] Wissenschaftsrat. 2011. Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften.

[5] Allianz der deutschen Wissenschaftsorganisationen. 2010. Grundsätze zum Umgang mit Forschungsdaten.

3. **Sensible personenbezogene Daten** (Offenlegung = Schaden für Person): Diese Daten besitzen einen Bezug zu Personen. Eine Offenlegung dieser Daten kann Schäden (Vermögen, Ansehen, etc.) für einen Personenkreis bedeuten. Diese Daten sind mit erhöhtem Schutzbedarf zu speichern und zu verbreiten.

4. **Original Forschungsdaten:** Forschungsdaten, in der Form, wie sie von den datengebenden Instituten in den internen, institutseigenen Speicherbereich eingespeist werden. Diese Daten sind bezogen auf den Gehalt an sensiblen Daten sehr unterschiedlich, es werden keine Auflagen in Bezug auf die vorangehende Anonymisierung gemacht. Häufig werden sie formal anonymisiert sein. Hierunter befinden sich Aufzeichnungen und Daten, die Personenbezug enthalten, wie gescannte Originaldateien oder auch digital erzeugte Daten. Allerdings sollten die Daten vorstruktriert sein ("Ordnerstruktur") und es sollte eine erste Risikobewertung (Risikofragebogen) vorgenommen werden (die am besten von oder mit Primärforschern erstellt werden kann). Die Original Forschungsdaten werden geschützt gesichert und sind im Weiteren nicht direkt zugänglich. Wenn später auf sie zugegriffen werden soll, bedarf es der ausdrücklichen Zustimmung des Datengebers.

5. **Bearbeitete „originale" Forschungsdaten im Prozess der Datenverarbeitung:** Forschungsdaten, die in unterschiedlichen Bearbeitungsprozessen (s.o.) erzeugt werden. Sie werden dokumentiert und gesichert gespeichert (T-I oder T-II, je nach Anonymisierung und Risikomerkmalen). Um redundante Bearbeitungen der Dokumente zu vermeiden, werden die einzeln erstellten Dateien/ Dokumente mit dem „Status" der Bearbeitung gespeichert. Diese sind ungeprüft, teilgeprüft und geprüft.

6. An das Zentrum eLabour freigegebene/übergebene Forschungsdaten: Diese Daten sind formal anonymisiert, risikobewertet und mit definierten Freigabe- und Nutzungsmöglichkeiten versehen. Sie dürfen nur im Rahmen dieser Möglichkeiten verwendet und/oder zugänglich gemacht werden. Dies wird mit Hilfe der Benutzerrollen kontrolliert und dokumentiert.

7. Für die wissenschaftliche Nutzung freigegebene Forschungsdaten: (mit Nutzungsverträgen)

Anonymisierung - Begriffe

Anonymisieren ist das Verändern personenbezogener Daten durch datenverändernde Maßnahmen, d.h. direkte, so dass die Einzelangaben über persönliche und sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können (§ 3 Abs. 6 BDSG).

Hierzu existiert ein aktuelles Urteil des Europäischen Gerichtshofs (EuGH), das das Verfügbarmachen von Forschungsdaten erleichtert. Der EuGH stellte fest, dass Daten dann für einen Datenverarbeiter personenbezogen sind, wenn er über rechtliche Mittel verfügt, die es ihm erlauben, das Zusatzwissen eines Dritten zu erlangen und die betroffene Person zu identifizieren. Ist eine Identifizierung dem Datenverarbeiter aber gesetzlich verboten oder für diesen angesichts des dafür erforderlichen Aufwands an Zeit, Kosten und Arbeitskraft praktisch nicht durchführbar, so dass das Risiko einer Identifizierung de facto vernachlässigbar erscheint, handelt es sich nicht um personenbezogene Daten für den Datenverarbeiter. Der EuGH vertritt somit zwar einen im Lichte des deutschen Meinungsstreits relativen Ansatz, knüpft die effektive Pseudonymisierung von Daten aber zugleich an hohe Voraussetzungen.

Formal anonymisierte Daten: Die formale Anonymisierung ist der geringste Grad der Anonymisierung, der lediglich die Entfernung der direkten Identifikatoren (Namen und Adressen) umfasst. Der Merkmalsumfang und die fachliche Gliederung der Daten bleiben dagegen vollständig erhalten. Die Bearbeitung oder Nutzung formal anonymisierte Daten wird von den Nutzungsbedingungen der VFU und dem Rahmenvertrag des eLabour-Zentrums geregelt.

Formal anonymisierte und bezogen auf zusätzliche Merkmale pseudonymisierte Daten: Diese erweiterte Form der Anonymisierung ist geeignet /wird eingesetzt, um identifizierende Merkmale zu entfernen, insbesondere besondere Arten personenbezogener, die bei Bekanntwerden ein hohes Schadenrisiko für die Befragten enthalten. Dabei handelt es sich um Informationen zu rassischer und ethnischer Herkunft, politischen Meinungen, religiöse oder weltanschauliche Überzeugungen, Gesundheitsdaten, Angaben zum Sexualleben und Angaben zur Gewerkschaftszughörigkeit (BDSG §3 Abs. 9; DSGVO Artikel 9 Abs.1). Mit der neuen DSGVO werden die besonderen Arten personenbezogener Daten auf genetische und biometrische Daten einer natürlichen Person erweitert (DSGVO Artikel 9 Abs.1).

Diese Art der Pseudonymisierung bietet - je nach Intensität der Pseudonymisierung - bereits einen guten Schutz vor Re-Identifikation. Der Personenbezug ist jedoch nicht gänzlich ausgeschlossen (insbesondere für "Insider"), sodass Nutzer vertraglich der Versuch der Re-Identifikation untersagt und Nutzungseinschränkungen (z.B. Vor-Ort-Nutzung, Remote-Zugänge) vorgenommen werden müssen. Die konkreten Einschränkungen sind durch die Risikobewertung vorgegeben und im Freigabeprozess festgelegt. Die Freigabe der dieser Art pseudonymisierter Daten erfolgt grundsätzlich nur nach einer Einzelfallprüfung der Daten in Abwägung des individuellen Schadensrisikos für die Befragten.

Faktisch anonymisierte Daten: Daten sind faktisch anonymisiert, wenn einen Personenbezug nur noch mit einem unverhältnismäßig hohen Aufwand wiederhergestellt werden kann (§ 3 Abs. 6 BDSG). Die faktische Anonymisierung zielt darauf ab, durch behutsame Informationsreduktion und Informationsveränderung die Zuordnungsmöglichkeiten von Merkmalsausprägungen zu den Merkmalsträgern zu verringern, um den analytischen Gehalt der Daten weitgehend zu erhalten. Diese Daten besitzen keinen Personenbezug mehr und können im Rahmen von eLabour von sämtlichen Projektpartnern genutzt werden.

Absolut anonymisierte Daten: Die absolute Anonymisierung ist die stärkste Form der Anonymisierung, bei der die Daten durch Verfremden und Löschen bzw. Schwärzen derart verändert werden, dass eine Zuordnung der Informationen zu natürlichen Personen, nach aktuellen technischen Stand nicht möglich ist.

Pseudonymisierte Daten: Pseudonymisieren ist das Ersetzen von Identifikationsmerkmalen durch ein Kennzeichen zu dem Zwecke, die Bestimmung der Betroffenen auszuschließen oder wesentlich zu erschweren (§ 3 Abs. 6a. BDSG). Bei der Pseudonymisierung werden die direkten Merkmalsausprägungen durch Ersatzbegriffe ersetzt, die über einen Schlüssel vergeben werden. Im Nachhinein kann über die entsprechenden Zuordnungsmerkmale (Schlüssel) eine De-Anonymisierung der Betroffenen erfolgen. Da das Zusammenführen von natürlichen Personen und Daten weiterhin grundsätzlichen möglich ist, handelt es sich bei pseudonymisierten Daten auch noch um personenbezogene Daten.

Literatur:

Wissenschaftsrat. 2011. Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf. Gesehen 04.05.2015.

OECD. 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm. Gesehen 04.05.2015.

Kommission Zukunft der Informationsinfrastruktur. 2011. Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder. http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf. Gesehen 04.05.2015.

Deutsche Forschungsgemeinschaft (DFG). 2015. Leitlinien zum Umgang mit Forschungsdaten. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_ forschungsdaten.pdf. Gesehen 02.02.2016.

Deutsche Forschungsgemeinschaft (DFG). 2014. Leitfaden für die Antragstellung – Projektanträge, DFG-Vordruck 54.01 – 04/14. www.dfg.de/formulare/54_01/54_01_de.pdf. Gesehen 04.05.2015.

Deutsche Forschungsgemeinschaft (DFG). 2013. Sicherung guter wissenschaftlicher Praxis. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf. Gesehen 04.05.2015.

Allianz der deutschen Wissenschaftsorganisationen. 2010. Grundsätze zum Umgang mit Forschungsdaten. www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaetze.html. Gesehen 04.05.2015.