Adaptive Datenschutzkonzepte in der Praxis des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV)

Tatjana Mika und Katharina Werhan, Berlin

Das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) hat die Aufgabe, Mikrodaten des statistischen Berichtswesens der Rentenversicherung für die wissenschaftliche Forschung zur Verfügung zu stellen. Dies geschieht auf der rechtlichen Grundlage der Datenschutz-Grundverordnung mittels anonymer Daten für die Forschung. Im Regelfall werden sogenannte Off-Site Scientific Use Files an wissenschaftliche Forschungseinrichtungen versandt. Außerdem wird Wissenschaftlerinnen und Wissenschaftlern die Möglichkeit gegeben, in den Räumlichkeiten der Deutschen Rentenversicherung Bund an Gastwissenschafts-Arbeitsplätzen mit schwächer anonymisierten Daten zu arbeiten, wobei die Ergebnisse der Analysen einzeln vor der Herausgabe auf Anonymität geprüft werden. Zudem wird die universitäre Lehre mit einem speziellen Angebot in Form von Campus Use Files unterstützt. Der Beitrag erörtert die Konzepte der Anonymisierung für die unterschiedlichen Zielgruppen und Forschungszwecke.

1. Einleitung

Mit der Einrichtung des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) hat die gesetzliche Rentenversicherung im Jahr 2004 eine zentrale Empfehlung der Kommission zur Verbesserung der informationellen Infrastruktur eingelöst,[1] nämlich, der Wissenschaft über besondere Verbreitungskanäle Zugang zu prozessproduzierten Mikrodaten der gesetzlichen Rentenversicherung zu verschaffen. Diese Daten sollten die Erforschung der Leistungen der Rentenversicherung ermöglichen und damit die Reformvorschläge zur gesetzlichen Alterssicherung durch empirische Belege rationalisieren. Der Verband Deutscher Rentenversicherungsträger (VDR) erhielt für die Startphase des Forschungsdatenzentrums eine Förderung vom Bundesministerium für Bildung und Forschung (BMBF). Der VDR sammelte zu diesem Zeitpunkt bereits seit Jahrzehnten anonymisierte, über alle Träger der Rentenversicherung harmonisierte Daten. Personenbeziehbare Daten standen dem Dachverband nicht zur Verfügung. Der Bestand an Mikrodaten des zentralen statistischen Berichtswesens wurde daher Grundlage der Forschungsdaten der gesetzlichen Rentenversicherung. Der Vorteil dieser Datengrundlage war und ist, dass sie mehrere Schritte der Qualitätssicherung durchläuft und unabhängig von der Programmierung des Fachverfahrens ist, mit dem die Leistungen der Rentenversicherung von der Sachbearbeitung bewilligt werden. Ein Nachteil ist, dass die Daten nicht unmittelbar mit externen Stichproben zusammengeführt werden können, weil keine direkten Identifikatoren der Versicherten und Rentenbeziehenden in den Daten gespeichert sind. Alle Projekte, bei denen die Daten des statistischen Berichtswesens mit weiteren Datenquellen zusammengebracht werden, erfordern daher zusätzliche Arbeitsschritte, um den Personenbezug kurzfristig wiederherzustellen. Diese Projekte müssen aus organisatorischen Gründen mit allen Trägern der Rentenversicherung koordiniert werden und erfordern daher Gremienbeschlüsse, welche die Arbeitsabläufe strukturieren.

In den Jahren seit der Gründung wurde das Datenangebot des FDZ-RV sukzessive

1 Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001), S. 286 f.

auf alle für die Wissenschaft interessanten Mikrodatensätze des statistischen Berichtswesens der Rentenversicherung ausgeweitet. Zudem kooperiert das FDZ-RV in Projekten mit anderen Infrastruktureinrichtungen der Sozial- und Wirtschaftsforschung, um Daten der Rentenversicherung mit weiteren relevanten Forschungsdaten zu verknüpfen. Mit solchen erweiterten Daten lässt sich die Erreichung gesetzlicher Ziele, wie die Absicherung im Alter, noch besser bewerten, weil beispielsweise die gesamte wirtschaftliche Lage der Personen in die Untersuchung einbezogen werden kann.

Das Kerngeschäft des FDZ-RV ist es allerdings, Scientific Use Files aus den jährlich erhobenen Statistikdaten der gesamten gesetzlichen Rentenversicherung zu veröffentlichen. Die Bereitstellung von Mikrodaten für die Wissenschaft in Form von Scientific Use Files erfordert die Umwandlung von Mikrodaten des statistischen Berichtswesens in anonyme Daten in einem für die Forschung geeigneten Format. Dazu wurden für diese Daten jeweils spezifische Standards entwickelt, die sich schnell nach der Erhebung der Daten umsetzen lassen. Der folgende Beitrag erläutert die Standards der Anonymisierung für Scientific Use Files für den Versand (Off-Site-Zugang) sowie den spezifischen Zugang am Gastwissenschafts-Arbeitsplatz (On-Site-Zugang) und geht auf die Umsetzung dieser Standards ein. Das Ziel dieser Routinen ist es, die Durchführung der wissenschaftlichen Nutzung nicht durch die Erstellung jeweils projektspezifischer Daten zu ermöglichen, sondern nach Möglichkeit auf Standardprodukte zu verweisen, welche nach Abschluss eines kurzen, positiv beschiedenen Antragsverfahrens versandt werden können. Solche schnellen Verfahren entsprechen dem Bedarf der Wissenschaft nach zeitlicher Planbarkeit. Außerdem erleichtert dies auch die Reproduzierbarkeit der Forschung für nachfolgende Untersuchungen. Solche Replikationsstudien[2] sind in der Wissenschaft ein wichtiges Instrument zur Qualitätssicherung. Das FDZ-Verfahren hat sich mit seinen Standardisierungen bewährt, weshalb die Nutzung der prozessproduzierten Forschungsdaten in der Forschung zu Rehabilitation und Alterssicherung deutlich gängiger wurde.[3] Der Vorteil anonymisierter Daten aus Verwaltungsvorgängen der Sozialversicherung ist, dass die Forschungsergebnisse realitätsnah ausfallen. Dies ist besonders günstig, wenn Folgen sozialpolitischer Regelungen nachvollzogen werden sollen, die sich mit Befragungsdaten kaum nachprogrammieren lassen, während sie aber in den administrativen Daten bereits als Merkmale vorliegen. Dies kann beispielweise an Untersuchungen zum Rentenzugang gezeigt werden, bei denen sehr eng am Rentenrecht ermittelt wurde, welche Auswirkungen die Anhebung der Verrentungsalter unter Berücksichtigung der Abschläge in den Altersrenten haben.[4]

Personenbezogene und personenbeziehbare Daten sind individualisierte Daten, das heißt solche, die eine eindeutige Zuordnung der Informationen zu einer bestimmten Person zulassen.[5] Die Nutzung solcher Daten ist in der wissenschaftlichen empirischen Forschung nur sehr selten erforderlich, denn die untersuchten Fälle sollen für typische, also gerade nicht individuelle Konstellationen von Merkmalen stehen. Die Weitergabe personenbeziehbarer Daten an die Forschung ist nur dann zulässig, wenn der Forschungszweck mit anonymisierten Daten nicht erreicht werden kann.[6] Dieser sogenannte Grundsatz der Priorität der Anonymisierung[7] besagt, dass nur in Ausnahmefällen Daten weitergegeben werden dürfen, die einen Rückschluss auf die Person zulassen. Grund für eine solche Weitergabe kann der Wunsch nach Kontaktaufnahme mit den betroffenen Personen sein,[8] um die Personen

2 Vgl. *Erdfelder* und *Ulrich* (2018), S. 3–21.
3 *Brünger* et al. (2025).
4 Vgl. hierzu *Seibold* (2021).
5 „Personenbezogen" bezieht sich auf Daten, die direkt einer bestimmten Person zugeordnet werden können (zum Beispiel da der Name angegeben ist), während „personenbeziehbar" Daten umfasst, die zwar nicht direkt, aber mit zusätzlichen Informationen oder durch Rückschlüsse einer Person zugeordnet werden können.
6 *Schroeder-Printzen* et al. (1990), § 75 Rn. 72.
7 *Rombach* (2004), § 75 Rz. 40.
8 Ebd.

zu ihrer sozialen Absicherung durch die gesetzliche Rentenversicherung zu befragen. Das FDZ-RV institutionalisierte diesen Vorbehalt der Nutzung anonymer Daten für die Rentenversicherung. Soweit möglich, sollen wissenschaftliche Forschungsprojekte mit den Daten des FDZ-RV durchgeführt werden. Gleichzeitig wurde durch die Verwendung der Daten des statistischen Berichtswesens der gesamten gesetzlichen Rentenversicherung die Repräsentativität der Ergebnisse sichergestellt. Die Rentenversicherung hat damit bereits seit 2004 eine konsolidierte Datengrundlage, wie sie durch das FDZ Gesundheit für die gesetzliche Krankversicherung erst seit Oktober 2025 angeboten wird.[9]

In den Ausgangdaten des statistischen Berichtswesens der gesamten Rentenversicherung ist der erste Schritt der Anonymisierung bereits vollzogen. Identifikatoren wie Namen, Adressen oder Sozialversicherungsnummern sind bereits entfernt worden. Bis zum Berichtsjahr 2010 konnte eine Person auch nicht über Statistiken hinweg beobachtet werden, weil die Fallnummer keiner Systematik folgte und bei jedem Statistikabzug der Daten neu vergeben wurde. Seit dem Berichtsjahr 2011 sind Verknüpfungen über Statistiken hinweg mit einem Pseudonym möglich. Nach der Anonymisierung verlieren die Daten den Charakter, Sozialdaten im Sinne des § 67 Abs. 2 Satz 1 SGB X zu sein,[10] weil sie keine Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person mehr enthalten.

Dieser Beitrag bestimmt in Abschnitt 2 zunächst Begrifflichkeiten wie Sozialdaten, Forschungsdaten und Anonymisierung. Abschnitt 3 bietet eine kurze Übersicht und Definition typischer Produkte von Forschungsdatenzentren, bevor dann in Abschnitt 4 die konkreten Maßnahmen zur Wahrung des Datenschutzes bei der Bereitstellung von Forschungsdaten im FDZ-RV beschrieben werden. Abschnitt 5 veranschaulicht diese Maßnahmen anhand von Anwendungsbeispielen.

2. Anonymisierung der Daten des statischen Berichtswesens der Rentenversicherung

2.1 Forschung mit Sozialdaten der Rentenversicherung

Das Sozialgesetzbuch ermöglicht die Forschung mit Daten der Sozialverwaltung. Als übergreifender Rechtsrahmen ist dies im Sozialgesetzbuch 10 (SGB X) geregelt. Speziell gibt § 75 SGB X in Absatz 1 die Regelung:

„Die Übermittlung von Sozialdaten ist zulässig, soweit sie erforderlich ist, für ein bestimmtes Vorhaben

1. Der wissenschaftlichen Forschung im Sozialleistungsbereich […]“

Dies bedeutet, dass auch einzelne Akten und detailreiche Mikrodaten mit personenidentifizierenden Merkmalen für Forschungsprojekte bereitgestellt werden dürfen.[11] Solche Daten sind allerdings unter starken Schutz gestellt, weshalb sie nur nach einer ausführlichen Prüfung der Erforderlichkeit durch den Datenschutz übermittelt werden. Wegen der engen Prüfung der Erforderlichkeit, insbesondere des Umfangs der Merkmale, wird § 75 SGB X vor allem für Auftragsforschung der Sozialleistungsträger verwendet. In diesen Fällen gibt es eine enge Verbindung der gewährten Leistung und der untersuchenden Forschung. Anträge auf Nutzung von Daten auf der Grundlage von § 75 Absatz 1 Nr. 1 SGB X richten sich an die zuständige oberste Verwaltungsbehörde.[12] Die Anträge werden daher bei der Deutschen Rentenversicherung eingereicht, aber von der obersten Ebene der Bundesverwaltung entschieden. Dies ist grundsätzlich das

9 www.forschungsdatenzentrum-gesundheit.de (aufgerufen am 04.10.2025).

10 Vgl. Gemeinsames Rundschreiben zu den datenschutzrechtlichen Vorschriften des SGB I und SGB X in der Fassung des Gesetzes zur Änderung des Bundesdatenschutzgesetzes und anderer Gesetze vom 18. Mai 2001 (BGBl. I, 904 ff.), S. 21.

11 *Bieresborn* (2020), § 75 Rn. 79.

12 Ebd., § 75 Rn. 19.

Bundesministerium für Arbeit und Soziales (BMAS), welches aber die Entscheidung an das Bundesamt für Soziale Sicherung (BAS) delegieren kann. Die Zuständigkeit für Forschungsvorhaben mit Daten der Deutschen Rentenversicherung wurde an das BAS delegiert. Bei der Kombination mehrerer Datenquellen können mehrere oberste Bundesbehörden zuständig sein,[13] was den Prozess der Genehmigung erheblich verlängert.

Die Weitergabe personenbeziehbarer prozessproduzierter Daten aus den Verwaltungsvorgängen der gesetzlichen Rentenversicherung an die Forschung ist demnach grundsätzlich möglich, unterliegt aber strengen Kontrollen der Datenschutzgenehmigung für jedes Merkmal des Einzelvorhabens in Hinblick auf das konkrete Forschungsvorhaben.[14] Für die Nutzung von Sozialdaten, wie sie personenbezogen in den Versicherungskonten bei den Trägern der Rentenversicherung vorliegen, sind demnach entsprechend umfangreiche Anträge auf Datennutzung gemäß § 75 SGB X zu stellen. Um die Anträge den Anforderungen entsprechend schreiben zu können, benötigen die antragstellenden Forschenden eine klare Beschreibung der verfügbaren Daten. Dies ist eine Hürde, denn für elektronische Akten einer Behörde existieren im Regelfall keine Beschreibungen, und wenn, sind sie für Forschende meist nicht zugänglich.

Liegt ein Forschungsvorhaben vor, das tatsächlich nur mit personenbezogenen Daten durchgeführt werden kann, ist außerdem nach Möglichkeit die Einwilligung der Betroffenen einzuholen.[15] Dieser Aufwand ist nur sinnvoll, wenn die Kontaktaufnahme auch mit einer Befragung verbunden wird, aus der sich weitere Daten für das Forschungsprojekt ergeben.[16]

Das Verfahren der Weitergabe von personenbezogenen Daten nach § 75 SGB X ist für die Nutzung der Daten aller Rentenversicherungsträger zudem aufwendig, weil die personenbezogenen Daten, die eine Kontaktierung der Personen zum Einholen der Einwilligung ermöglichen, bei den 16 Trägern der Rentenversicherung liegen.[17]

Das FDZ-RV betreut keine Projekte auf der Grundlage des § 75 SGB X. Durch die Erstellung anonymer Forschungsdaten werden die Verfahren nach § 75 Abs. 1 Nr. 1 SGB X gerade vermieden. Um das Ziel der Bereitstellung anonymisierter Forschungsdaten zu erreichen, müssen aus Sozialdaten der Sozialversicherung in mehreren Verfahrensschritten Forschungsdaten entwickelt werden. Die ersten Schritte der Anonymisierung werden bereits bei der Generierung der Daten für das statistische Berichtswesen der Rentenversicherung durchgeführt, welche im nächsten Abschnitt beschrieben werden.

2.2 Generierung statistischer Daten aus den Konten der Rentenversicherung

Daten, welche in den Konten der Rentenversicherungsträger als Versicherungskonto gespeichert sind, erlauben die genaue Identifikation der Versicherten und Rentenbeziehenden. Das Konto umfasst die Informationen, welche die genaue Zuordnung der Anwartschaften und Leistungen zu einer bestimmten Person möglich machen, darunter der Geburtstag, der Geburtsort, der Name und gegebenenfalls der Geburtsname sowie die aktuellen Adressdaten und weitere Kontaktinformationen. Die Konten der gesetzlichen Rentenversicherung liegen bei den 16 Versicherungsträgern, auf welche die Versicherten und Rentenbeziehenden aufgeteilt sind.

Zentraler Baustein der eindeutigen Identifizierung in den Konten ist die Versicherungsnummer (VSNR). Die Versicherungsnummer wird gemäß § 147 SGB VI von der Rentenversicherung erstellt. Für diese Aufgabe ist die Datenstelle der Rentenversicherung zuständig. Diese Nummer erlaubt auch bei gleichlautendem Namen oder Geburtsdatum

13 Ebd.
14 Ebd., § 75 Rn. 13.
15 § 75 Absatz 1 Satz 2 SGB X.
16 Eine Alternative wäre eine vorab registrierte Zustimmung zur Verwendung von Prozessdaten für die Forschung.
17 Wenn es nur einen Träger der sozialen Sicherung gibt, wie dies bei der Bundesagentur für Arbeit (BA) der Fall ist, vereinfacht dies die Antragstellung erheblich.

eine eindeutige Zuordnung. Aufgrund dieser Eindeutigkeit ist die Versicherungsnummer als solche, auch ohne weitere ergänzende Daten, ein Sozialdatum und zu schützen. Bei der Datenstelle der Rentenversicherung wird auch der Stammsatz vorgehalten, der alle jemals vergebenen VSNR speichert und damit eine erneute Vergabe einer bereits verwendeten Versicherungsnummer verhindert. Durch den Stammsatz ist somit zu jedem Zeitpunkt erkennbar, welcher Träger für ein bestimmtes Konto zuständig ist.

Im Rentenversicherungskonto wird alles gespeichert, was für die Bearbeitung der Leistungsansprüche erforderlich ist, für welche die gesetzliche Rentenversicherung zuständig ist. Die gespeicherten Informationen im Konto ergeben sich zudem aus den Anträgen auf Leistungen sowie gewährten Leistungen auf Rehabilitation und Rente. Darüber hinaus wird eine kleine Anzahl von statistischen Informationen aus den Arbeitgebermeldungen gespeichert. Hierzu zählen die vertraglich vereinbarten Stunden der Erwerbsarbeit in Klassen und die ausgeübte Tätigkeit nach einer von der Bundesagentur für Arbeit entwickelten Klassifikation. Die von den Arbeitgebern gemeldeten versicherungspflichtigen Einkommen werden in den Versicherungskonten gespeichert, sodass für den Fall eines Leistungsanspruchs bereits viele Voraussetzungen der notwendigen Vorversicherungszeiten geprüft werden können. Damit sind die Daten der Rentenversicherung Sozialdaten, die in den meisten Fällen aufgrund des Konzepts der Kontenführung auf aktuellem Stand gehalten sind. Außerdem enthalten sie beispielsweise bei Rentenbeziehenden und Rehabilitandinnen und Rehabilitanden mit Anspruch auf Übergangsgeld Angaben zu Geldzahlungen.

Für das gemeinsame statistische Berichtswesen werden bei der Deutschen Rentenversicherung Bund Auszüge aus den Konten aller Rentenversicherungsträger gesammelt. Diese Auszüge sollen die relevanten Vorgänge der gesamten gesetzlichen Rentenversicherung abbilden. Hierfür haben die Gremien der Rentenversicherung mehrere Datenformate abgesprochen, die zumeist jährlich an die zentrale Stelle gemeldet werden. Die meisten Daten werden auf der Grundlage der Allgemeinen Verwaltungsvorschrift über die Statistik in der Rentenversicherung (RSVwV) erstellt. Die Verordnung regelt, welche Daten an die Aufsichtsbehörden übermittelt werden.

Die Datenabzüge für die Statistik werden gemäß einer in trägerübergreifenden Gremien der Rentenversicherung abgestimmten inhaltlichen Vorgabe erstellt. Die Statistiken sollen eine möglichst präzise Grundlage für die statistische Erfassung der Beiträge zur Rentenversicherung und ihrer Leistungen ermöglichen. Ein Schwerpunkt liegt auf aktuellen Reformen des Rentenrechts. Um eine langfristige Vergleichbarkeit zu ermöglichen, sind dabei möglichst viele Merkmale über die Zeit konstant benannt und definiert. Die Datenabzüge haben jeweils einen inhaltlichen Schwerpunkt, beispielweise den Neuzugang in die Rente in einem Kalenderjahr oder die Inanspruchnahme einer Rehabilitation.

Die Datenmeldungen enthalten weder die VSNR noch den Tag der Geburt noch Name oder Anschrift der Versicherten und Rentenbeziehenden. Obwohl die meisten Datenabzüge alle Personen umfassen, die in einen bestimmten Personenkreis fallen, ist es ohne diese eindeutig identifizierenden Merkmale nicht möglich, einzelne Personen sicher in diesen Daten zu identifizieren. Ebenfalls nicht enthalten sind Informationen zu Bankkonten, welche diese Daten zu einem attraktiven Ziel von Datenangriffen machen könnten.

Die Statistik der Deutschen Rentenversicherung arbeitet seit 2011 mit Verfahren der Pseudonymisierung, bei denen der gleichen Person in allen Statistiken die identische Zeichenkombination zugewiesen wird.[18] Diese Möglichkeit erlaubt die Untersuchung von Fragestellungen, die sich mit den getrennten Statistikmeldungen nicht beantworten ließen.

18 Die Daten der Statistik bis zum Berichtsjahr 2010 sind nicht mit persistenten Identifikatoren ausgestattet. Sie lassen sich auf keinem Weg sicher bestimmten Personen zuordnen.

Die Pseudonymisierung wird in einem mehrschrittigen Verfahren durchgeführt. Der jeweils zuständige Rentenversicherungsträger wandelt mithilfe eines für alle Träger einheitlichen kryptografischen Algorithmus (Schlüssel) die VSNR in das erste Pseudonym um. Die dann einfach pseudonymisierten Daten werden an den Bereich statistisches Berichtswesen der Deutschen Rentenversicherung Bund weitergeleitet, und dort wird unter Verwendung eines weiteren Schlüssels das Pseudonym 1 in ein Pseudonym 2 umgesetzt. Die Programme zur Erstellung der beiden Schlüssel sowie die Schlüssel selbst dürfen von den jeweiligen Schlüsselverwaltern nicht weitergegeben werden. Durch die Pseudonymisierung werden die Daten für alle Personen wirksam anonymisiert, welche keinen Zugang zu dem Verschlüsselungsverfahren haben.[19]
Die Verknüpfung im FDZ-RV wird, wenn eine Verbindung der Daten im Rahmen eines Forschungsprojektes notwendig ist, mit einem weiteren Pseudonym (FDZ_ID) vorgenommen, das auf dem zweiten Pseudonym beruht und dessen Schlüssel im FDZ-RV bewahrt wird.

2.3 Weitergabe anonymisierter Daten an die wissenschaftliche Forschung

Das Ziel der Anonymisierung von Forschungsdaten ist, dass die Daten von den Forschenden keiner bestimmten oder bestimmbaren Personen mehr zugeordnet werden können. Der Begriff der Anonymisierung wurde im Zuge der Angleichung des deutschen Sozialrechts an die Datenschutz-Grundverordnung (DSGVO) aus dem SGB X gestrichen. Weil die Verordnung selbst Anonymisierung nicht definiert, müssen die Erwägungsgründe der DSGVO für die Anwendung mit herangezogen werden, insbesondere Erwägungsgrund 26.[20] Nach der vermittelnden Meinung soll bei der Einschätzung, welche Daten anonym sind, erwogen werden, welche zusätzlichen Informationen der datennutzenden Person objektiv zur Verfügung stehen oder welche diese sich legal verschaffen könnte.[21] Weil Forschende in der Regel keinen Zugang zu personenbezogenen Daten zum Beispiel aus Sozialversicherungskonten haben, gibt diese Interpretation dem FDZ-RV einigen Spielraum.
Bestimmbar ist eine Person nicht nur dann, wenn Informationen im Datensatz unmittelbar auf sie eindeutig hinweisen, etwa durch Namen, Adresse und Geburtsdatum, sondern auch, wenn eine Person aufgrund von Zusatzinformationen eindeutig identifiziert werden kann. Die anspruchsvolle Aufgabe der Anonymisierung besteht darin festzustellen, unter welchen Umständen eine Person aufgrund der im Datensatz enthaltenen Angaben bestimmbar sein könnte, wenn die eindeutig identifizierenden Merkmale Name, Vorname und Geburtsdatum bereits entfernt wurden. Dies ist einerseits denkbar, wenn der Datensatz eine Information enthält, die sich eindeutig nur auf eine Person beziehen kann, wie es bei einer Versicherungsnummer etwa der Fall ist.[22] Eine solche Nummer, die von der Verwaltung mit dem Ziel einer eindeutigen Identifikation der Person in den Datensatz aufgenommen wurde, muss demnach in jedem Fall entfernt werden. Andererseits kann eine Person auch bestimmbar sein, wenn sie in einer kleinen Gruppe die einzige mit einer bestimmten, eindeutig erkennbaren und aus anderen Informationsquellen bekannten Merkmalsausprägung ist. Der Maßstab ist, über welches identifizierende Zusatzwissen die Person verfügt, der die Daten zur Verfügung gestellt werden.[23]

19 Am 04.09.2025 stellte der Europäische Gerichtshof klar, dass Pseudonymisierung ein wirksames Mittel der Anonymisierung sein kann (Rechtssache C-413/23 P, Rn. 16). Dies bemisst sich danach, dass „pseudonymisierte Daten für die Zwecke der Verordnung 2018/1725 [DGSVO] nicht in jedem Fall und für jede Person als personenbezogene Daten betrachtet werden. Denn die Pseudonymsierung kann – je nach Umständen des Einzelfalls – andere Personen als den Verantwortlichen tatsächlich an einer Identifizierung der betroffenen Person hindern, sodass letztere für sie nicht oder nicht mehr identifizierbar ist.“ (ebd., Rn. 86). Der Maßstab für die wirksame Anonymisierung ist dabei, ob „die Identifizierung dieser Person gesetzlich verboten oder praktisch nicht durchführbar ist, z.B. weil sie einen unverhältnismäßigen Aufwand an Zeit, Kosten und Arbeitskraft erfordern würde“ (ebd., Rn. 82).
20 *Bieresborn* (2020), vor §§ 67–85a, Rn. 60.
21 Ebd.
22 Vgl. hierzu *Heese* (2004), S. 313.
23 *Bieresborn* (2020), S. 828, Rn. 60.

Für die Beurteilung der Anonymisierung ist nicht nur der Informationsgehalt des Datensatzes zu beachten, sondern auch die Zielgruppe, die den Datensatz erhalten soll. Eine erfolgreiche Anonymisierung ist erreicht, wenn es der Zielgruppe, die Zugang zu dem Datensatz erhält, nicht möglich sein wird, mit ihren Ressourcen an Zeit, Arbeitskraft und Informationen Personen in dem Datensatz zu bestimmen, also zu re-identifizieren. Je enger die Zielgruppe eingegrenzt und ihre Nutzungsmöglichkeiten eingeschränkt werden können und je stärker die Überwachung durch das FDZ-RV gegeben ist, umso einfacher lässt sich der erforderliche Grad der Anonymisierung gewährleisten. Wenn am Gastwissenschafts-Arbeitsplatz mit beschränkten technischen Mitteln sowie beschränkten zeitlichen und personellen Ressourcen geforscht wird, ist auch mit einem weniger stark anonymisierten Datensatz ausreichender Datenschutz gewährleistet. Unterschieden wird daher zwischen verschiedenen Graden der Anonymisierung. Anonymisierte Daten werden an Vertragspartner in der Wissenschaft ausgeliefert, die sich vertraglich verpflichten, die Vorgaben zum Datenschutz einzuhalten.

Anonymisieren in diesem Sinne bedeutet damit, dass in jedem Einzelfall die personenbezogene Identifikation einen unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft benötigt. Die Anonymisierung bestimmt sich also negativ nach der Chance und dem Aufwand der Re-Identifikation einer Person anhand der Informationen aus dem Datensatz. Die Prüfung einer erfolgreichen Anonymisierung beinhaltet zwei Aspekte: Es wird im ersten Schritt ermittelt, mit welchen Zusatzinformationen die Re-Identifikation einer Person möglich wäre, um dann im zweiten Schritt den Aufwand und Ertrag eines solchen Angriffs zu vergleichen. Wenn die Re-Identifikation einer Person aus dem Datensatz entweder nicht gelingt oder Aufwand an Zeit, Kosten und Arbeitskraft außer Verhältnis zum potenziellen Nutzen der Aufdeckung der personenbezogenen Informationen stehen, dann ist die Anonymisierung des Datensatzes erfolgreich durchgeführt worden. Um eine solche Abwägung durchführen zu können, muss der Personenkreis der Datennutzenden in seinen technischen, zeitlichen und sonstigen relevanten Ressourcen abgeschätzt werden können.

2.4 Ausschluss der Re-Identifikation in anonymisierten Daten

Grundsätzlich sind zwei Strategien der Re-Identifikation zu unterscheiden. Die eine zielt auf die Identifizierung von Personen mit bestimmten Merkmalen in den Daten. So könnte etwa ein Interesse an Bezieherinnen und Beziehern besonders hoher Einkommen bestehen, um Einkommensinformationen über diese Personengruppe an einen kommerziellen Adressenhändler zu veräußern. Solche Versuche werden konsequent verhindert, indem die Identifikatoren wie Namen, Adressen und Versicherungsnummern entfernt werden. Damit kann zwar ermittelt werden, wie viele Personen die Merkmalsausprägung haben, aber sie können nicht aufgrund der Datenanalyse real gefunden werden. Forschungsdaten enthalten nicht nur keine Daten, die zur Identifizierung von Versicherten und Rentenbeziehenden verwendet werden können, sondern auch keine kommerziell verwertbaren Angaben wie Bankverbindungen, weshalb sie für Personen mit kriminellen Absichten nicht von Interesse sind.

Schwieriger zu verhindern ist eine Einzelfallsuche, bei der eine dem Suchenden aus anderen Quellen bekannte Person in den Daten gesucht wird. Es ist jedoch generell für Forschende schwierig, in größerem Umfang in den Besitz personenbezogener Daten zu gelangen, weil die Weitergabe solcher Daten nicht gestattet ist und in vielen Fällen auch organisatorisch verhindert wird.

Unterstellt man Versuche der absichtlichen De-Anonymisierung, dann kommen als Möglichkeiten der Re-Identifikation

Tabelle 1: Beispiel für einen Versuch der Re-Identifikation

	Identifikations-merkmale (z. B. Name, Adresse)	Überschneidungs-merkmale (z. B. Alter)	Nutzmerkmale (z. B. Rentenhöhe)
Mikrodatenfile (z. B. Rentenbestand 2021)	7 137	100 %	vorhanden (Angriffsziel)
Identifikationsfile (Zusatzwissen)	3 385	47,4 %	fehlen

Quelle: eigene Darstellung.

„Matchen“, also das Zusammenführen zweier Datenquellen aufgrund übereinstimmender Merkmale infrage. Beispielsweise würde auf Basis von Überschneidungsmerkmalen der anonymisierte Datensatz mit einem personifizierten zusammengefügt. Die Überschneidungsmerkmale müssen in zusammengefügten Datensätzen in der identischen Form vorliegen.

Tabelle 1 verdeutlicht, dass der Versuch der gezielten Verknüpfung, der eine bestimmte Person beispielsweise im Datensatz Rentenbestand 2021 identifizieren will, zunächst einen Identifikationsdatenfile benötigt, der Name und Adresse enthält, und daher später eine Zuordnung der Informationen zu einer bestimmten Person ermöglicht. Die Datensicherheit steigt damit allgemein, wenn personenbezogene Daten von keiner Institution, die sie hält, in Umlauf gebracht werden. Das Zusatzwissen muss darüber hinaus Merkmale enthalten, die sich sowohl in dem Identifikationsfile als auch in dem angegriffenen, anonymisierten Datensatz befinden. Anhand dieser Überschneidungsmerkmale werden dann die Datensätze zusammengeführt, wodurch eine eindeutige Identifikation der Personen in dem angegriffenen Datensatz gelungen wäre. Ziel des Angriffs sind die zusätzlich gewonnenen Informationen, die in dem Identifikationsfile nicht enthalten sind, die Nutzmerkmale beispielsweise des Rentenbestandsdatensatzes.

Das Verfahren kann nur dann gelingen, wenn die Überschneidungsmerkmale eine nur einmal auftretende Kombination bilden, weshalb sich jedem Fall in dem Identifikationsfile nur ein Fall aus dem angegriffenen Mikrofile zuordnen lässt. Gibt es sehr viele übereinstimmende Fälle mit der gleichen Merkmalskombination, etwa weil die Merkmale in groben Kategorien abgebildet werden, zum Beispiel Alter in Lebensjahren, oder weil jede Merkmalskategorie sehr häufig besetzt ist, wie etwa beim Merkmal Geschlecht die Kategorien Mann und Frau, dann scheitert eine eindeutige Zuordnung.[24] Daher ist neben der Anzahl der Überschneidungsmerkmale vor allem die Anzahl der Merkmalsausprägungen entscheidend. Mit höherem Informationsgehalt des Merkmals, beispielsweise dem Geburtstag, steigt die Wahrscheinlichkeit einer eindeutigen Kombination in den Überschneidungsmerkmalen.[25] Eingeschränkt wird die Möglichkeit der Identifikation von Einzelpersonen außerdem, wenn zwar Überschneidungsmerkmale vorliegen, diese jedoch durch Vergröberung im anonymisierten Daten-

24 Vgl. hierzu *Müller* et al. (1991), S. 47 f.
25 Ebd., S. 53.

bestand nicht mehr deckungsgleich sind beziehungsweise weniger detaillierte Informationen bieten als der Identifikationsfile. Das theoretische Re-Identifikationsrisiko ist im Wesentlichen durch drei Eigenschaften der Datenbasis beeinflusst:[26]

1. Informationsgehalt der Überschneidungsmerkmale:
   Eine Re-Identifikation setzt die Einzigartigkeit von Ausprägungskombinationen voraus. Je höher die Anzahl der Überschneidungsmerkmale, der Differenzierungsgrad der Ausprägungen sowie ihrer Verteilungen, desto deutlicher sind die Datensätze im Merkmalsraum voneinander abgegrenzt und desto eher werden „Eins-zu-Eins-Zuordnung“ möglich sein.
2. Stichprobeneigenschaft der Daten:
   Eine Re-Identifikation ist nur dann möglich, wenn eine Person im Mikrodatenfile mit den notwendigen Zusatzinformationen erfasst ist. Die Stichprobe stellt daher eine prinzipielle Schranke für die Erfolgswahrscheinlichkeiten von De-Anonymisierungen dar. Ist eine Ausprägung in der Grundgesamtheit mehrmals besetzt und stehen zur Untersuchung nur Stichproben zur Verfügung, ist eine Re-Identifikation auch bei einer „Eins-zu-Eins-Zuordnung“ nicht mehr mit Sicherheit möglich, sondern nur noch mit einer gewissen Wahrscheinlichkeit, da Verwechslungen mit statistischen Doppelgängern nicht ausgeschlossen werden können.
3. Kompatibilität zwischen Mikrodatenfile und Zusatzwissen:
   Die De-Anonymisierung setzt voraus, dass die gesuchten Datensätze im Mikrodatenfile und Zusatzwissen identisch abgebildet sind.

Bevor in Abschnitt 4 die konkreten Maßnahmen des FDZ-RV bei der Generierung anonymisierter Forschungsdaten erläutert werden, wird zunächst eine kurze Übersicht über typische Produkte von Forschungsdatenzentren und eine Begriffsbestimmung dieser Produkte eingeschoben.

3. Typische Produkte von Forschungsdatenzentren

Eine Reihe von Forschungsdatenzentren bieten spezialisierte Daten für die wissenschaftliche Forschung an.[27] Auch die Sozialversicherungen richteten Forschungsdatenzentren ein, zum Teil, weil sie einen gesetzgeberischen Auftrag zur Forschung und Forschungsförderung haben,[28] zum Teil, weil sie direkt gesetzlich verpflichtet wurden.[29] In der Genese dieses Datenangebots über die letzten zwei Jahrzehnte benannten die Forschungsdatenzentren ihre Daten stets aus ihrer eigenen institutionellen Logik heraus. Tabelle 2 veranschaulicht ein Konzept der einheitlicheren Bezeichnung, das eine übergreifende Arbeitsgruppe entwickelte und welchem nach Möglichkeit gefolgt werden sollte.[30]

Die Bezeichnungen sollen den wissenschaftlich Forschenden schon im Namen erkennbar machen, ob die Datenprodukte an die Arbeitsplätze an Universitäten und unabhängigen Forschungseinrichtungen versandt werden (Off-Site-Version des Scientific Use File), im Verfahren des Fernzugriffs ausgewertet werden können (Remote-Version des Scientific Use File) oder ob sie nur in den Räumen des Forschungsdatenzentrums genutzt werden dürfen (On-Site-Version des Scientific Use File). Dies ist wichtig, da die wissenschaftliche Forschung bereits bei der Planung eines Projekts wissen muss, ob die Durchführung mit Dienstreisen und längeren Zeiten der speziellen Datenprüfung vor Ort im Forschungsdatenzentrum einhergeht. Viele Forschende können ohne entsprechende

26 Vgl. hierzu *Wirth* (1992).

27 Die Forschungsdatenzentren werden vom Rat für Sozial- und Wirtschaftsdaten akkreditiert und sind im Ausschuss für Forschungsdateninfrastruktur organisiert. https://www.konsortswd.de/ueber-uns/fdi-ausschuss (aufgerufen am 06.10.2025).

28 Vgl. hierzu *Buschmann-Steinhage* (2015).

29 § 282 Abs. 7 SGB III verpflichtet die Bundesagentur für Arbeit für Forschungsprojekte auf dem Gebiet der Arbeitsmarkt- und Berufsforschung Daten bereitzustellen. § 303 a bis f SGB V legt die Aufgaben des FDZ Gesundheit fest, welches die Daten der gesetzlichen Krankenversicherungen aufbereitet und zur Verfügung stellt.

30 *Fuß* und *Schlücker* (2025).

Tabelle 2: Empfehlung zur einheitlichen Bezeichnung für Datenprodukte im Angebot der Forschungsdatenzentren

	Public Use File (PUF)	**Campus Use File (CUF)**	**Scientific Use File (SUF)**
Nutzungszweck	offen	Lehre	wissenschaftliche Analyse
Nutzungsvoraussetzung	keine	keine	vertragspflichtig
Zugangsweg			
Off-Site	PUF	CUF	SUF Off-Site-Version
Remote Access*			SUF Remote-Version
On-Site			SUF On-Site-Version
Detailgrad der Daten	inhaltliche Aussagen bei eher niedrigem Detailgrad möglich	keine inhaltlichen Aussagen möglich, primärer Fokus auf Datenstruktur	inhaltliche Aussagen mit hohem (Off-Site) bis sehr hohem (On-Site) Detailgrad möglich

Anmerkung: *Für eine ausführliche Definition verschiedener Remote Access Möglichkeiten siehe Ellert (2025) in dieser Ausgabe.

Quelle: eigene Darstellung.

Projektmittel Datenzugänge vor Ort im Forschungsdatenzentrum aus finanziellen und organisatorischen nicht nutzen.

Darüber hinaus bieten Forschungsdatenzentren mit Campus Use Files ergänzend auch Lehrmaterial für den universitären Unterricht, damit bereits der Nachwuchs den sachgerechten Umgang mit empirischen Daten an realistischem Material lernen kann. Schließlich gibt es bei einigen Forschungsdatenzentren auch Daten für die Öffentlichkeit, sogenannte Public Use Files. Dies ist allerdings nur ein Nebenprodukt der eigentlichen Zuständigkeit für Forschungsdatenzentren, weil öffentliche Daten in der wissenschaftlichen Forschung nur sehr eingeschränkt genutzt werden.

Das FDZ-RV richtet sich seit 2022 nach dem in Tabelle 2 beschriebenen Bezeichnungsvorschlag. So können nun Forschende durch die Bezeichnung und die ausführliche Dokumentation der Daten auf der Website des FDZ-RV bereits vor der Antragstellung wissen, wie aufwendig sich der Datenzugang für sie gestaltet.

Der Schwerpunkt der Arbeit des FDZ-RV liegt auf der Bereitstellung von Scientific Use Files, also anonymisierter Daten, deren Nutzung ausschließlich anerkannten wissenschaftlichen Institutionen vorbehalten ist. Dabei konzipiert das FDZ-RV seine Datenprodukte vom Ausgangspunkt des Datenschutzes aus. Der Grad der Anonymisierung bestimmt, wie das Datenangebot klassifiziert wird. Je nach Detailgrad der Daten werden die Scientific Use Files nach Vertragsschluss als Off-Site-Version zum Download oder als On-Site-Version für die Arbeit am Gastwissenschafts-Arbeitsplatz zur Verfügung gestellt. Das FDZ-RV erstellt keine gesonderten Remote-Versionen der Daten, die On-Site-Versionen werden sowohl für die

Datenfernverarbeitung (Remote Access) als auch für die Arbeit am Gastwissenschafts-Arbeitsplatz genutzt.
Abschnitt 4 beschreibt die Maßnahmen, die zur Sicherstellung des Datenschutzes im FDZ-RV unternommen werden, im Detail. Im Anschluss bietet Abschnitt 5 einige Beispiele dafür, wie diese Maßnahmen konkret umgesetzt wurden.

4. Maßnahmen zur Sicherstellung des Datenschutzes im FDZ-RV

Die DSGVO regelt grundsätzlich, dass erhobene Daten nur in dem Umfang genutzt werden dürfen, der mit dem ursprünglichen Erhebungszweck zu vereinbaren ist. Allerdings privilegiert die DSGVO in Artikel 5 Absatz 1 b) die Nutzung von Daten für die Forschung, weshalb in diesem Kontext die Nutzung auch ohne unmittelbare Zweckbindung an den ursprünglichen Erhebungszweck rechtmäßig ist. Die Formulierung lautet: „… eine Weiterverwendung für im öffentlichen Interesse liegende Archivzwecke, für wissenschaftliche oder historische Forschungszwecke oder für statistische Zwecke gilt gemäß Artikel 89 Absatz 1 nicht als unvereinbar mit den ursprünglichen Zwecken“. Dies bedeutet, dass es grundsätzlich rechtmäßig sein kann, die Daten aus Verwaltungsprozessen der Rentenversicherung für wissenschaftliche Forschungsvorhaben zu übermitteln. Die Archivierung, die wissenschaftliche und historische Forschung und Statistik werden in Artikel 5 Absatz 1 e) zudem als legitime Zwecke der dauerhaften Speicherung von Identifikatoren genannt. Für die Forschung sind demnach fallweise auch gespeicherte personenbeziehbare Daten zugänglich zu machen. Auch diese Nutzung steht allerdings unter dem Vorbehalt der Datenminimierung gemäß Artikel 5 Absatz 1 c), die Identifikatoren müssen für die Beantwortung der Forschungsfrage also unbedingt erforderlich sein.[31]
Das FDZ-RV geht also auf der Grundlage der DSGVO von der grundsätzlichen Legitimität der Datennutzung durch die wissenschaftliche Forschung aus, bemüht sich aber um Anonymität und Datenminimierung, soweit die Maßnahmen die Durchführung der Forschung nicht behindern.[32]
Datenschutz wird dabei grundsätzlich durch eine Vielzahl an Maßnahmen erreicht. Es ist üblich, dass datenseitige, technische und organisatorische Maßnahmen koordiniert angewandt werden, um das Ergebnis der Anonymität zu erreichen.[33] In der Literatur wurden hierfür die Begriffe „Portfolioansatz“ und „Five-Safes-Modell“ geprägt.[34] Gemeint ist, dass ein Portfolio aus fünf Maßnahmenarten angewandt wird: Safe People (geschulte Forschende), Safe Projects (geprüfte Projekte), Safe Settings (kontrolliertes technisches Umfeld), Safe Outputs (kontrollierte Ergebnisse) und Safe Data (anonymisierte Daten).[35] Je weitreichender und wirkungsvoller organisatorische und technische Maßnahmen sind, desto weniger restriktiv können die Vorkehrungen bei den Daten selbst sein und umgekehrt.[36]
Im Folgenden wird weiter ausgeführt, wie die datenseitigen, organisatorischen und technischen Maßnahmen im FDZ-RV konkret angewandt werden, um den Schutz der Forschungsdaten zu gewährleisten.

4.1 Organisatorische Maßnahmen

Gleich drei der „Five Safes“ fallen in die Kategorie der organisatorischen Maßnahmen. Im FDZ-RV wird bereits im Antragsprozess sichergestellt, „dass die Daten nicht in unbefugte Hände geraten“[37]. Die Forschungsdaten der Rentenversicherung werden grundsätzlich auf Projektbasis zur Verfügung gestellt. Dabei wird geprüft, ob die Projekte inhaltlich in den Bereich der

31 Ein Beispiel hierfür ist Forschung mit Geburtstagen und Sterbedaten, zum Beispiel um die Sterblichkeit an bestimmten Tagen mit besonderen Wetterbedingungen zu erforschen.
32 Persönliche Identifikatoren werden im FDZ-RV grundsätzlich nicht gehalten.
33 Vgl. hierzu *Schwartmann* et al. (2024), Artikel 5 Rn. 53.
34 Vgl. hierzu *Desai* et al. (2016).
35 Vgl. hierzu *Schiller* et al. (2017).
36 Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001), S. 266 ff.
37 Ebd., S. 266.

Alterssicherungsforschung fallen und ob die beantragten Daten für die Beantwortung der Forschungsfragen notwendig und geeignet sind (Safe Projects). Je nach notwendiger Detailtiefe der Daten zur Beantwortung der Forschungsfrage gibt es zudem verschiedene weitere Nutzungsvoraussetzungen. Sollen Scientific Use Files genutzt werden, ist eine weitere Bedingung für die Antragsberechtigung, dass die Projektleitung eine promovierte Person ist, die zudem an einer wissenschaftlichen Einrichtung angestellt ist. Somit wird sichergestellt, dass die Datennutzenden den Prinzipien der guten wissenschaftlichen Praxis folgen. Der für die Nutzung von Scientific Use Files verpflichtend abzuschließende Datennutzungsvertrag enthält zudem eine Geheimhaltungsverpflichtung, die von allen Datennutzenden unterzeichnet werden muss („Safe People"). Zusätzlich enthält der Vertrag einige technische und organisatorische Voraussetzungen, zu deren Einhaltung sich die Datennutzenden verpflichten. Dies gilt insbesondere bei der Nutzung von Off-Site-Versionen von Scientific Use Files, die den Nutzenden zum Download zur Verfügung gestellt werden und auf den Rechnern der jeweiligen wissenschaftlichen Einrichtung genutzt werden können („Safe Settings").

4.2 Datenseitige Maßnahmen

Allgemein lassen sich vier Anonymisierungstechniken unterscheiden, die wiederum in Kombination angewandt werden: Pseudonymisierung (Entfernen direkter Identifikatoren), Informationsreduktion (Merkmalsauswahl und -vergröberung), Stichprobenziehung und Verhindern von Matching. Im Folgenden wird erläutert, wie diese Techniken bei der Konzeption und Bereitstellung von Off-Site Scientific Use Files im FDZ-RV angewandt werden.[38] Bei der Bereitstellung von On-Site-Versionen von Scientific Use Files und bei Campus Use Files werden die gleichen Maßnahmen angewandt – jedoch in Kombination mit technischen und organisatorischen Maßnahmen in abgewandelter Form. Details hierzu werden in den Abschnitten 4.2.5 und 4.2.6 erläutert.

4.2.1 Pseudonymisierung

Wie in Abschnitt 2 dargelegt, sind die den Forschungsdaten des FDZ-RV zugrundeliegenden Daten des statistischen Berichtswesens bereits von allen Identifikatoren getrennt, die eine unmittelbare Zuordnung zu den Versicherten, Rehabilitanden und Rentenbeziehenden der Deutschen Rentenversicherung ermöglichen würden. Im FDZ-RV liegen zu keinem Zeitpunkt der Datenverarbeitung Daten aus den Verfahren der Rentenversicherung mit direkten Identifikatoren (wie Name oder Sozialversicherungsnummer) vor. Wenn solche Identifikatoren vorliegen, dann resultieren sie aus Befragungen, in denen die Angaben zur Person freiwillig übermittelt wurden und werden getrennt von den Informationen aus den Forschungsdaten gehalten.[39]

4.2.2 Informationsreduktion

Eine Re-Identifikation durch die gezielte Suche von bekannten Personen wird durch eine Informationsreduktion verhindert. Hierbei werden verschiedene nachvollziehbare und dokumentierte Einzelmaßnahmen angewandt.
Zunächst wird sowohl in Scientific Use Files als auch in Campus Use Files nur eine Auswahl der verfügbaren Merkmale bereitgestellt. Diese Vergröberung, etwa durch das Entfernen der Monatsangabe für den Zeitpunkt der Geburt, sorgt für eine größere Zahl von Treffern bei der gezielten Suche nach Einzelpersonen, sodass die gesuchte

38 In den Anfängen des FDZ-RV orientierte man sich an Empfehlungen zur Weitergabe, die in einem gemeinsamen Forschungsprojekt des Statistischen Bundesamtes und Zentrums für Umfragen, Methoden und Analysen (ZUMA) Mannheim für den Mikrozensus entwickelt wurden: *Köhler* (1999); *Müller* et al. (1991); *Wirth* (1992); *Stegmann* et al. (2005).

39 Ein Beispiel für ein solches Projekt wird in Abschnitt 5.3 beschrieben.

Einzelperson eben nicht eindeutig identifiziert werden kann. Bei Off-Site Scientific Use Files erfolgt diese Auswahl mit dem Ziel, einerseits für eine breite Zahl an Forschungsfragen relevante Informationen zur Verfügung zu stellen und gleichzeitig die Informationen so weit zu reduzieren, dass das Re-Identifikationsrisiko minimiert ist. Einige Merkmale werden auch nicht weitergegeben, da sie nur eine kleine Anzahl von Personen betreffen und kein Forschungsinteresse vermutet wird.[40]

Weiterhin werden die ausgewählten Merkmale vergröbert, also zum Beispiel zusammenfasst. Dabei muss für Off-Site Scientific Use Files systematisch geprüft werden, ob das Erkennen einer Person aufgrund ihrer einzigartigen Merkmalskombination allgemein bekannter Variablen wie Geburtsjahr, Geschlecht oder Wohnort unmöglich ist. Ist diese Voraussetzung nicht erfüllt, ist eine sachgerechte Vergröberung/Zusammenfassung vorzunehmen, bis eine Re-Identifizierung von Einzelpersonen in der Grundgesamtheit ausgeschlossen ist. Beispielsweise wird als unmittelbare Regionalangabe in der Regel das Bundesland übermittelt, je nach Grundgesamtheit ist aber auch eine Differenzierung nach Raumordnungsregionen (96 Kategorien) möglich, oder es muss eine stärkere Einschränkung in Ost und West vorgenommen werden. Auch Angaben zu Staatsangehörigkeiten oder das Geburtsland zählen zu den oft bekannten Merkmalen einer Person und werden in Off-Site Scientific Use Files nur als Gruppen von Nationalitäten oder gar als dichotomes Merkmal „deutsch“/ „nicht deutsch“ weitergegeben. Weiterhin werden Angaben zu Entgeltpunkten und Geldbeträgen aus Zahlungen nur in gerundeter Form weitergegeben. Zusätzlich wird hier durch Zusammenfassung der Werte im obersten Prozent der Verteilung eine Begrenzung vorgenommen. Jahresangaben von Geburtsdaten werden an den Rändern der Verteilung gruppiert, um untypisch junge oder alte Personen der Grundgesamtheiten Erwerbstätige oder Rentnerinnen und Rentner zu anonymisieren. So werden im Rentenzugang die Personen mit erster Rentenzahlung bis zum Alter 29 und ab dem Alter 80 jeweils zusammengefasst. Diese Anonymisierung betrifft besonders junge Erwerbsgeminderte und besonders spät zugehende Altersrentnerinnen und -rentner. Die Geburtsmonate werden in der Regel bei Off-Site Scientific Use Files gar nicht weitergegeben.[41] Diese Altersgrenzen werden über viele Jahre konstant gelassen. Anders wird im Scientific Use File „Aktiv Versicherte“ verfahren. Hier werden für das Berichtsjahr 2022 die Geburtsjahre bis 1952 zusammengefasst sowie die 2007 und später Geborenen. Damit sind Personen verstärkt anonymisiert, die besonders lange Zeiträume nach der Regelaltersgrenze noch erwerbstätig sind und besonders junge, welche bereits vor oder direkt nach Ende der Schulplicht erwerbstätig sind. Diese Angaben zu den Geburtsjahrgängen werden gleitend bei der Aktualisierung der Datenprodukte angepasst.

4.2.3 Stichprobenziehung

Die Ziehung einer Zufallsstichprobe ist ein besonders effektiver Schutzmechanismus. Rückschlüsse auf die Daten einzelner Personen zu ziehen wird verhindert, da die Kenntnis darüber entzogen wird, ob der gesuchte Merkmalsträger in der Stichprobe enthalten ist oder nicht. Für Off-Site-Versionen von Scientific Use Files und Campus Use Files werden grundsätzlich nur Stichproben weitergegeben. On-Site werden auch Vollerhebungen bereitgestellt, dann aber mit deutlich reduziertem Merkmalskranz.

40 Ein Beispiel für solche Merkmale sind spezielle Berufsgruppen aus bestimmten Herkunftsregionen, welche nach dem „Fremdrentenrecht“ (FRG) eine Anerkennung ihrer Versicherungszeiten aus dem Ausland in der Deutschen Rentenversicherung erhielten. Diese Untergruppen sind teilweise sehr klein, und die Versicherungszeiten im Ausland sind selten Gegenstand der Forschung.

41 Eine Ausnahme ist die Zeitreihe Rentenwegfall 2001–2022. Details hierzu werden in Abschnitt 5.2 erläutert.

4.2.4 Verhindern von Matching

Wie in Abschnitt 2.4 beschrieben gilt es darüber hinaus zu verhindern, dass die Kombination von zwei Datensätzen die Re-Identifizierung von betroffenen Personen ermöglicht. Neben der zuvor beschriebenen Vergröberung von Informationen (beispielsweise des Geburtsdatums) ist eine weitere Methode hierzu, die systemfreie Anordnung der Beobachtungen im zu übermittelnden Datenbestand und die Vergabe von Identifikationsnummern, die über Berichtsjahre oder Datenbestände hinweg nicht die gleiche Person identifizieren. Dieses Verfahren wird für alle Off-Site Scientific Use Files angewandt, sodass Forschende verschiedene Datenprodukte des FDZ-RV nicht miteinander verknüpfen können.
Für die Beantwortung einiger Forschungsfragen ist das Zusammenführen verschiedener Datenquellen jedoch unerlässlich. Dies kann Daten innerhalb einer datenhaltenden Stelle, aber auch zwischen verschiedenen datenhaltenden Stellen bedeuten. Für letzteres fehlt oftmals die rechtliche Grundlage oder die Genehmigungsverfahren sind sehr aufwendig. Um Projekte auf Basis der bestehenden oder künftiger rechtlicher Grundlagen durchführen zu können, müssen Forschungsdatenzentren rechtlich und technisch in die Lage versetzt werden, einen Identifikator vorzuhalten, der solche Verknüpfungen ermöglicht. Die verknüpften Daten werden aus Datenschutzgründen in der Regel aber die Server der datenhaltenden Stellen nicht verlassen dürfen und damit nur On-Site, sprich am Gastwissenschafts-Arbeitsplatz oder im kontrollierten Fernrechnen, nutzbar sein.

4.2.5 Besonderheiten bei der Bereitstellung von On-Site-Versionen von Scientific Use Files

Für On-Site-Versionen der Scientific Use Files werden die beschriebenen Techniken ebenfalls angewendet, jedoch können wegen der zusätzlichen technischen Schutzmechanismen die Anzahl der Merkmale, deren Detailtiefe oder die Stichprobe größer sein. Die am Gastwissenschafts-Arbeitsplatz oder im kontrollierten Fernrechnen bereitgestellten Daten werden eng mit den Forschenden abgestimmt. Auch hier gilt der Grundsatz der Datenminimierung. Für die Forschungsfrage wichtige Untersuchungsmerkmale werden möglichst differenziert weitergegeben; im Gegenzug wird der weitergegebene Merkmalskatalog im Vergleich zum Off-Site Scientific Use File eingegrenzt und die weniger wichtigen Variablen werden deutlich vergröbert.
Da bei der Nutzung am Gastwissenschafts-Arbeitsplatz und im kontrollierten Fernrechnen durch organisatorische (Abschnitt 4.1) und technische Maßnahmen (Abschnitt 4.3) überprüft wird, welche zusätzlichen Informationen den Daten zugespielt werden[42] und welche Informationen das IT-System der Deutschen Rentenversicherung Bund verlassen, können On-Site-Versionen der Daten – wenn das Forschungsprojekt dies erfordert – als Vollerhebung und in pseudonymisierter Form zur Verfügung gestellt werden. Somit können On-Site-Informationen aus verschiedenen internen Datenquellen und Berichtsjahren miteinander verknüpft werden. Abschnitt 5.2 zeigt die Unterschiede zwischen einer Off-Site- und On-Site-Version eines Scientific Use Files am Beispiel des Datenprodukts „Zeitreihe Versichertenrentenwegfall wegen Todes 2001–2022“.

4.2.6 Besonderheiten bei der Bereitstellung von Campus Use Files

Anders als bei den Scientific Use Files, die dem eingeschränkten Kreis der Wissenschaft und Forschung zur Verfügung gestellt werden, sollen die Campus Use Files für die universitäre Lehre zur Verfügung gestellt werden. Hier gilt nicht mehr das Wissenschaftsprivileg der Anonymisierung, die

42 In der Regel handelt es sich hierbei um aggregierte Daten wie beispielsweise Arbeitslosenquoten auf Ebene der Landkreise oder Belastungsskalen auf Ebene von Berufen.

Daten müssen vollständig anonym sein. Zur Erreichung dieser sehr starken Anonymisierung werden für Campus Use Files weitergehende Prüfkriterien festgelegt:

1. Es wird ein sehr reduzierter Merkmalskranz angeboten, der keinerlei wissenschaftliche Ansprüche erfüllt.
2. Die Stichprobengröße ist in der Regel noch kleiner als bei Off-Site Scientific Use Files.
3. Jede Ausprägung eines kategorisierten Merkmals soll im Campus Use File mindestens fünffach besetzt sein. Dies gilt ebenso bei ganzzahlig gerundeten Entgeltpunktmerkmalen. Bei quasi-stetigen Monatsvariablen müssen die zusammengefassten Kategorien an den Rändern der Verteilung diese Voraussetzung erfüllen.
4. Kombinationen sensibler soziodemografischer Merkmale und solche, über die sehr einfach Zusatzinformationen zu erhalten sind (zum Beispiel Geschlecht + deutsch/ausländisch + Familienstand + Altersgruppe + Anzahl der Kinder), die im anonymisierten Datenbestand Einzelfälle bilden, sollen mindestens dreimal im Originaldatensatz besetzt sein.
5. Sensible Merkmale werden grundsätzlich klassifiziert (zum Beispiel Einkommen) oder entfernt (zum Beispiel Detailmerkmale zum Reha-Geschehen).

4.3 Technische Maßnahmen

Wie zuvor beschrieben ist aus Datenschutzgründen eine hohe Detailtiefe in Off-Site Scientific Use Files nicht möglich, da diese für eine möglichst breite Anzahl an Forschungsprojekten geeignet sein sollen und somit einen möglichst umfassenden Merkmalskatalog beinhalten müssen. Das hat zur Folge, dass es zu einer deutlichen Reduzierung der Tiefe der Merkmalsausprägungen kommen muss. Für manche Forschungsfragen reichen die Detailtiefe oder der Merkmalskatalog der standardisierten Off-Site-Versionen von Scientific Use Files nicht aus.[43] Zudem kann es sein, dass für Forschungsprojekte Daten aus verschiedenen Berichtsjahren oder internen Datenquellen miteinander verknüpft werden sollen.

Ergibt sich aus einem Forschungsvorhaben ein Bedarf an Daten, die in Umfang und/oder Detailtiefe über einen standardisierten Off-Site Scientific Use File hinausgehen, ist die Nutzung der Daten via Remote Access möglich.[44] Dies bedeutet, dass die Daten auf den sicheren Servern der Deutschen Rentenversicherung verbleiben und entweder im kontrollierten Fernrechnen (Remote Execution) oder am Gastwissenschafts-Arbeitsplatz (Remote Desktop) zugänglich gemacht werden (Safe Settings).

Die Auswertungen werden bei beiden Varianten auf den Servern der Deutschen Rentenversicherung durchgeführt. Die Forschenden erhalten lediglich Zugriff auf einen Ordner, in dem für ihre Forschungsfrage zugeschnittene On-Site-Versionen der Scientific Use Files abgelegt sind (Safe Settings). Dort können sie ebenfalls ihre Forschungsergebnisse abspeichern. Diese werden nach einer Datenschutzprüfung an die Forschenden gesendet (Safe Output). Das FDZ-RV bietet mit Stand 2025 sechs Gastwissenschafts-Arbeitsplätze in Berlin und einen Arbeitsplatz in Würzburg an. Durch vertragliche Vereinbarungen konnte an drei weiteren Orten ein Zugang zu Daten des FDZ-RV in den Räumen der dort angesiedelten Forschungsdatenzentren gesichert werden. Die Orte sind Mannheim, Köln und München.

5. Anwendungsbeispiele

5.1 Scientific Use File „Zeitreihe Versichertenrentenzugang – Erwerbsminderung und Diagnosen“ [45]

Der Scientific Use File „Zeitreihe Versichertenrentenzugang – Erwerbsminderung und Diagnosen“ enthält gemäß den in

43 Vgl. hierzu *Himmelreicher* (2004).
44 Für eine ausführliche Definition verschiedener Remote Access Möglichkeiten siehe *Ellert* (2025) in dieser Ausgabe.
45 Vgl. hierzu *Werhan* (2023).

Abschnitt 4.2 beschriebenen Methoden keine direkten Identifikatoren und ist systemfrei sortiert. Er enthält eine 20-Prozent-Zufallsstichprobe der Erwerbsminderungszugänge im jeweiligen Berichtsjahr. Somit kann selbst bei anscheinender Re-Identifikation einer Person nie davon ausgegangen werden, dass es sich tatsächlich um diese handelt, da das Vorhandensein statistischer Zwillinge in der Grundgesamtheit nicht ausgeschlossen ist.

Die Regionalangabe mit der tiefsten Merkmalsausprägung sind die Raumordnungsregionen des Bundesinstituts für Bau-, Stadt- und Raumordnung (BBSR)[46]. Diese 96 Regionen sind eine in der Forschung gängige Einheit und dienen der funktionalen Gliederung Deutschlands zum Zweck der Raumordnung. Damit bieten sie einen höheren Detailgrad als das Bundesland und gewährleisten gleichzeitig die Einhaltung der Vorgaben über die Mindestanzahl von Fällen.

Die Variable zur Staatsangehörigkeit wurde dichotomisiert und gibt lediglich Auskunft darüber, ob eine Person die deutsche Staatsbürgerschaft oder eine andere Staatsbürgerschaft besitzt.

Angaben zu Entgeltpunkten sind gemäß des Anonymisierungskonzepts ganzzahlig gerundet und am oberen Rand der Verteilung durch Zusammenfassung begrenzt. Außerdem sind Geldwerte ebenfalls ganzzahlig gerundet, und als oberster Wert wird der Mittelwert des obersten Prozent im jeweiligen Berichtsjahr angegeben (arithmetisches Mittel der Werte im 99. Perzentil).

5.2 Scientific Use File „Zeitreihe Versichertenrentenwegfall wegen Todes“[47]

Da ein großes Interesse der Mortalitätsforschung darin besteht zu analysieren, wie lange Personen welche Art von Rente in welcher Höhe bezogen und ob Personen zum Regelrentenalter in die Rente eingetreten sind oder nicht, wurden für diesen Datensatz – abweichend von den Off-Site Scientific Use Files „Versichertenrentenzugang“ und „Versichertenrentenbestand“ – auch der Geburts- und Sterbemonat der Versicherten aufgenommen. Da das Geburts- und das Todesdatum zumeist öffentlich verfügbare Daten sind, wurde ein zweistufiges Anonymisierungskonzept erstellt:

1. Eine Off-Site-Version des Scientific Use Files, die Forschenden nach Vertragsunterzeichnung zur Nutzung am institutseigenen Rechner zur Verfügung gestellt wird und ein hohes Maß an Anonymisierung voraussetzt.
2. Eine standardisierte On-Site-Version des Scientific Use Files, die mehr Detailtiefe bietet, jedoch nur in kontrollierter Umgebung (Gastwissenschafts-Arbeitsplatz oder Remote Execution) und mit kontrollierter Ergebnisfreigabe zugänglich ist.

Aufgrund der schwachen Besetzung von ausländischen Rentenbeziehenden in ostdeutschen Bundesländern konnte bei multiplen Kombinationen das in Abschnitt 4.2.2 beschriebene Mindestfallzahl-Kriterium nicht eingehalten werden. Daher wurde entschieden, in der Off-Site-Version nur Personen mit deutscher Staatsangehörigkeit auszuweisen und Forschungsfragen, die Migrationsaspekte einbeziehen wollen, auf die On-Site-Version zu verweisen. Für die Off-Site-Version wurde zudem eine Zufallsstichprobe von 20 Prozent gezogen. Auch die Merkmalsauswahl und Detailtiefe wurden je nach On-Site- beziehungsweise Off-Site-Version angepasst. Als Regionalmerkmal wird beispielsweise in der Off-Site-Version eine Variable mit lediglich vier Ausprägungen „Nord (SH, NI, HB, HH)“, „Ost (MV, BB, BE, SN, ST, TH)“, „Süd (BY, BW)“ und „West (NW, HE, RP, SL)“ sowie eine Variable zur Stadt-Land-Unterscheidung (4 Kategorien) ausgewiesen.

46 https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/regionen/Raumordnungsregionen/raumordnungsregionen.html (aufgerufen am 06.10.2025).

47 Vgl. hierzu Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) (2024b).

In der Off-Site-Version wird somit eine Re-Identifikation anhand des Wohnortes verhindert. Da für die Mortalitätsforschung regionale Besonderheiten relevant sind, wären detailliertere Merkmale wie das Bundesland oder sogar Raumordnungsregionen (96 Kategorien) erforderlich. Da diese in Kombination mit den weiteren Merkmalen des Datensatzes aber die in Abschnitt 4.2.2 beschriebenen Fallzahlkriterien nicht erfüllen, werden sie lediglich in der On-Site-Version zur Verfügung gestellt. Hier verhindern die zusätzlichen organisatorischen und technischen Maßnahmen eine Re-Identifikation von Einzelpersonen.

5.3 Scientific Use File „SHARE-RV“

Bei dem Scientific Use File SHARE-RV (SHARE-ERIC 2024) handelt es sich um das Produkt eines von drei sogenannten Record-Linkage-Projekten des FDZ-RV.[48] Als Record-Linkage wird die Verknüpfung zweier Datensätze anhand eines eindeutigen Indikators, der in beiden Datensätzen enthalten ist, bezeichnet. Im Beispielsfall SHARE-RV werden Befragungsdaten des Survey of Health, Ageing and Retirement in Europe (SHARE) mit Daten des FDZ-RV verknüpfbar gemacht. Hierzu werden deutsche Staatsbürgerinnen und -bürger und ihre Partnerinnen und Partner, die freiwillig an den SHARE-Befragungen teilnehmen, während des Interviews um Zustimmung gebeten, ergänzende Daten aus der Rentenversicherung in die Auswertung einbeziehen zu dürfen. Die Befragten können sich freiwillig für oder gegen diese Ergänzung entscheiden. Im Einwilligungsfall senden die Befragten die unterschriebene Einwilligung sowie Daten zur Ermittlung ihrer Sozialversicherungsnummer an das Forschungsdatenzentrum. In mehreren Schritten, die jeweils von verschiedenen Stellen in der Rentenversicherung ausgeführt werden, werden die für das SHARE-Projekt benötigten Daten pseudonymisiert (SHARE-ID zu SHARE-Projekt-ID). Durch frühzeitiges Löschen der Zuordnung von Versicherungsnummer zu SHARE-Projekt-ID bei den Stellen in der Rentenversicherung wird eine einfache Identifizierung über die Versicherungsnummer unmöglich gemacht.

Auch bei der Aufbereitung der Daten werden die in Abschnitt 4.2 beschriebenen Methoden angewandt; mit dem Ergebnis, dass es sich bei den zur Verknüpfung bereitgestellten Daten in Form der Scientific Use Files um anonymisierte Daten handelt, die unter den vertraglich festgelegten Voraussetzungen an wissenschaftlich Forschende versandt werden können.

Die Kooperationen zwischen wissenschaftlichen Datenproduzenten, welche Befragungen durchführen, mit dem FDZ-RV und weiteren Anbietern prozessproduzierter Daten nehmen stetig zu.[49]

6. Zusammenfassung und Ausblick

Forschungsdatenzentren sind eine bedeutende Weiterentwicklung der informellen Infrastruktur für die Sozialwissenschaften. Die Veröffentlichung von anonymen Mikrodaten aus Verwaltungsprozessen, also Daten, bei denen eine Person als Einheit über verschiedene Merkmale hinweg beobachtbar ist, ist inzwischen die zentrale Grundlage vieler empirischer sozialwissenschaftlicher Forschungsprojekte. Diese Daten aus Verwaltungsprozessen, aufbereitet als Off-Site Scientific Use Files oder schwächer anonymisierte Daten für Gastwissenschaftlerinnen und Gastwissenschaftler, erfordern die kreative und konsequente Anwendung datenschutzrechtlicher Anonymisierungskonzepte. Das Ziel ist die Ermöglichung eines Forschungsergebnisses, das die Strukturen

48 In weiteren Projekten werden die Daten des FDZ-RV mit Befragungsdaten des Sozio-oekonomischen Panels (SOEP) und mit Prozessdaten des Forschungsdatenzentrums der Bundesagentur für Arbeit im Institut für Arbeitsmarkt und Berufsforschung (IAB) verknüpft.

49 Vgl. hierzu *Börsch-Supan* et al. (2020); *Lüthen* et al. (2022),; *Mika* und *Czaplicki* (2010); *Schmitz* et al. (2019); SHARE-ERIC (2024).

der Realität richtig wiedergibt, dabei jedoch keine Rückschlüsse auf Einzelpersonen zulässt. Dies bringt auch umfangreiche Datenprüfungen mit sich, die eine gute Qualität der Daten sicherstellen und die Anonymisierungskonzepte regelmäßig prüfen. Das Ergebnis sind standardisierte Datenprodukte des Forschungsdatenzentrums, zu denen in einem schnellen Verfahren Zugang gegeben wird. Hierdurch wird auch die Verwaltung entlastet, weil umfangreiche Anträge auf Nutzung von Sozialdaten gemäß § 75 SGB X vermieden werden. Dieses Angebot vorkonfigurierter Daten stößt auf große Nachfrage bei den Forschenden, denen verlässliche und schnelle Datenzugänge wichtiger sind als eine Detailtiefe, welche die Prozesse der Verwaltung vollständig abbildet.

Forschungsdatenzentren sind inzwischen aus der empirischen Sozialforschung nicht mehr wegzudenken. In einer nächsten Ausbaustufe sollen weitere Datenquellen erschlossen, vor allem aber die vorhandenen Daten verknüpft werden. Für diese nächste Ausbaustufe der deutschen Dateninfrastruktur müssen einige Gesetzesgrundlagen geändert werden, denn der Umgang mit den prozessproduzierten Forschungsdaten ist in speziellen Gesetzen wie dem Steuerstatistikgesetz, dem Gesundheitsdatennutzungsgesetz sowie verschiedenen Untergesetzen des Sozialgesetzbuches geregelt. Während die Zusammenarbeit der Forschungsdatenzentren in den letzten zwanzig Jahren in konstruktiver Kooperation funktioniert hat, weshalb viele einzelne Projekte umgesetzt werden konnten, müssen für eine gelungene Reform der Gesetzesgrundlagen die zuständigen Ministerien zusammenarbeiten.

Literatur

Bieresborn, Dirk (2020): In: Schütze, Bernd (Hrsg.): SGB X, Sozialverwaltungsverfahren und Sozialdatenschutz, Kommentar. 9. Auflage, München, C. H. Beck.

Börsch-Supan, Axel/Czaplicki, Christin/Friedel, Sabine/Herold, Imke/Korbmacher, Julie und *Mika, Tatjana* (2020): SHARE-RV: Linked Data to Study Aging in Germany. Jahrbücher für Nationalökonomie und Statistik 240 (1), S. 121–132. DOI: 10.1515/jbnst-2018-0034.

Brünger, Martin/Brzoska, Patrick/du Prel, Jean-Baptist/Ellert, Sebastian/Exner, Anne-Kathrin/Knoop, Tobias/Leinberger, Sarah/March, Stefanie/Mika, Tatjana/Reims, Nancy/Rohrbacher, Max/Schuler, Michael/Wahidie, Diana und *Hetzel, Christian* (2025): Nutzung von Routinedaten in der Rehabilitationsforschung – Teil 1: Ein Überblick über Art, Zugang, Qualität und Datenschutz. Rehabilitation 64 (3), S. 167–175. DOI: 10.1055/a-2575-9422.

Buschmann-Steinhage, Rolf (2015): Forschung im Bereich der gesetzlichen Rentenversicherung. In: Mülheims, Laurenz/Hummel, Karin/Peters-Lange, Susanne/Toepler, Edwin und Schuhmann, Iris (Hrsg.): Handbuch Sozialversicherungswissenschaft. Wiesbaden, Springer VS, S. 295–308.

Desai, Tanvi/Ritchie, Felix und *Welpton, Richard* (2016): Five Safes: designing data access for research. Economics Working Paper Series. https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf (aufgerufen am 06.10.2025).

Dietz, Martin/Müller, Anne/Schludi, Martin und *Walwei, Ulrich* (2015): Multidisziplinäre Arbeitsmarktforschung am Institut für Arbeitsmarkt- und Berufsforschung. In: Mülheims, Laurenz/Hummel, Karin/Peters-Lange, Susanne/Toepler, Edwin und Schuhmann, Iris (Hrsg.), Handbuch Sozialversicherungswissenschaft. Wiesbaden: Springer VS: S. 308–321.

Ellert, Sebastian (2025): Remote Access in verschiedenen Forschungsdatenzentren – Stand der Technik. Deutsche Rentenversicherung 80 (3), S. 229-240.

Erdfelder, Edgar und *Ulrich, Rolf* (2018): Zur Methodologie von Replikationsstudien. Psychologische Rundschau 69 (1), S. 3–21. DOI:10.1026/0033-3042/a000387.

Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) (2024a): Codeplan Scientific Use File „Zeitreihe Versichertenrentenzugang – Erwerbsminderung und Diagnosen“. https://fdz-rv.de/fileadmin/Dokumente_Datenprodukte/RTZN_EM/CPL.SUF.RTZN.2001-2020.EM.1-0.deu.pdf (aufgerufen am 06.10.2025).

Forschungsdatenzentrum der Rentenversicherung (FDZ-RV) (2024b): Codeplan Scientific Use File „Zeitreihe Versichertenrentenwegfälle wegen Todes 2001–2022“. https://fdz-rv.de/fileadmin/Dokumente_Datenprodukte/RTWF_TD/CPL.SUF.RTWF.2001-2022.TD.1-0.deu.pdf (aufgerufen am 06.10.2025).

Fuß, Daniel und *Schlücker, Friederike* (2025): Empfehlung des FDI Ausschusses für eine einheitliche Bezeichnung von Datenprodukten der Forschungsdatenzentren. KonsortSWD Working Paper 2025 (11). DOI: 10.5281/zenodo.14899704.

Heese, Claudia (2004): Aspekte des Datenschutzes im Forschungsdatenzentrum der Rentenversicherung. DRV-Schriften Band 55: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau. S. 41–48.

Himmelreicher, Ralf K. (2004): Befunde einer schriftlichen Befragung unter potenziellen Nutzerinnen und Nutzern des Forschungsdatenzentrums der gesetzlichen Rentenversicherung. DRV-Schriften Band 55: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau. S. 49–60.

Köhler, Sabine (1999): Anonymisierung von Mikrodaten in der Bundesstatistik und ihre Nutzung – Ein Überblick. Schriftenreihe „Forum der Bundesstatistik“ Band 31, S. 133–150.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) (2001): Wege zur einer besseren informationellen Infrastruktur: Baden-Baden: Nomos.

Lüthen, Holger/Schröder, Carsten/Grabka, Markus M./Goebel, Jan/Mika, Tatjana/Brüggmann, Daniel/Ellert, Sebastian und *Penz, Hannah* (2022): SOEP-RV: Linking German Socio-Economic Panel Data to Pension Records. Jahrbücher für Nationalökonomie und Statistik 242 (2), S. 291–307. DOI: 10.1515/jbnst-2021-0020.

Mika, Tatjana und *Czaplicki, Christin* (2010): SHARE-RV: Eine Datengrundlage für Analysen zu Alterssicherung, Gesundheit und Familie auf der Basis des Survey of Health, Ageing and Retirement in Europe und der Daten der Deutschen Rentenversicherung. RVaktuell 12/2010, S. 396–400.

Müller, Walter/Blien, Uwe/Knoche, Peter und *Wirth, Heike* (1991): Die faktische Anonymität von Mikrodaten. Schriftenreihe „Forum der Bundesstatistik“ Band 19.

Rombach, Wolfgang (2004): In: Hauck, Karl/Noftz, Wolfgang/Oppermann, Dagmar und Becker, Peter (Hrsg.): Sozialgesetzbuch SGB X: Kommentar.

Schiller, David H./Eberle, Johanna/Fuß, Daniel/Goebel, Jan/Heining, Jörg/Mika, Tatjana/Müller, Dana/Röder, Frank/Stegmann, Michael und *Stephan, Karsten* (2017): Standards des sicheren Datenzugangs in den Sozial- und Wirtschaftswissenschaften: Überblick über verschiedene Remote-Access-Verfahren. RatSWD Working Paper Nr. 261. DOI: 10.17620/02671.15.

Schmitz, Marc/du Prel, Jean-Baptist und *Hasselhorn, Hans Martin* (2019): lidA – Leben in der Arbeit: Kohortenstudie zu Arbeit, Alter, Gesundheit und Erwerbsteilhabe. lidA-Factsheet 2019 04. https://arbeit.uni-wuppertal.de/fileadmin/arbeit/Factsheets/FS_2019_04__lidA_Studienbeschreibung.pdf (aufgerufen am 06.10.2025).

Schroeder-Printzen, Günther/Engelmann, Klaus/Schmalz, Ursula/Wiesner, Siegfried und *von Wulffen, Matthias* (1990): Sozialgesetzbuch Verwaltungsverfahren – SGB X: Kommentar. München, C. H. Beck‘sche Verlagsbuchhandlung.

Schwartmann, Rolf/Jaspers, Andres/Thüsing, Gregor und *Kugelmann, Dieter*

(2024): DS-GVO/BDSG, Datenschutzgrundverordnung Bundesdatenschutzgesetz: C. F. Müller.

Seibold, Arthur (2021): Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities. American Economic Review 111 (4), S. 1126–1165. DOI: 10.1257/aer.20191136.

SHARE-ERIC (2024): SHARE-RV: Linking German SHARE Survey Data with Administrative Records. https://share-eric.eu/data/data-set-details/share-rv (aufgerufen am 21.08.2025).

Stegmann, Michael/Luckert, Hilmar und *Mika, Tatjana* (2005): Die Bereitstellung prozessproduzierter Daten der GRV im Forschungsdatenzentrum der Rentenversicherung (FDZ-RV). Grundsätze zur faktischen Anonymisierung von Mikrodaten und zu Gastwissenschaftler-Arbeitsplätzen. Deutsche Rentenversicherung 60 (2–3), S. 203–215.

Werhan, Katharina (2023): Data on the Population of Reduced Earnings Capacity-Pensioners in Germany, 2001–2020. Jahrbücher für Nationalökonomie und Statistik 243 (5), S. 585–595. DOI: 10.1515/jbnst-2022-0042.

Wirth, Heike (1992): Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes. ZUMA Nachrichten 16 (30): S. 7–65. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-209679 (aufgerufen am 06.10.2025).

Anschrift der Verfasserinnen:

Tatjana Mika
Legiendamm 12 F
10179 Berlin

Katharina Werhan
Deutsche Rentenversicherung Bund
Abteilung Forschung und Entwicklung
Forschungsdatenzentrum der
Rentenversicherung (FDZ-RV)
Ruhrstraße 2
10709 Berlin