

RatSWD Working Paper Series

RatSWD.
Rat für Sozial- und
Wirtschaftsdaten

286

Für einen besseren Datenzugang im Bereich Arbeitsmarkt und Sozialversicherung

Stellungnahme aus dem
Verein für Socialpolitik

Unterarbeitsgruppe Arbeitsmarkt
und Sozialversicherung

Dezember 2023

www.ratswd.de

RatSWD Working Papers

des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers*-Reihe startete Ende 2007. In dieser Online-Publikationsreihe werden konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autorinnen bzw. Autoren dar und nicht die des RatSWD. Die Zuwendungsgeber des RatSWD haben die Publikationen nicht beeinflusst.

Herausgeberin oder Herausgeber der RatSWD Working Papers-Reihe ist die/der Vorsitzende des RatSWD:

seit 2020 Monika Jungbauer-Gans

2014–2020 Regina T. Riphahn

2009–2014 Gert G. Wagner

2007–2008 Heike Solga

Für einen besseren Datenzugang im Bereich Arbeitsmarkt und Sozialversicherung

Stellungnahme aus dem Verein für Socialpolitik

Unterarbeitsgruppe Arbeitsmarkt und Sozialversicherung*

Dezember 2023

<https://doi.org/10.17620/02671.90>

Dieses Paper ist außerdem beim Verein für Socialpolitik erschienen:

https://www.socialpolitik.de/de/vfs_stellungnahmen_daten_arbeitsmarkt

***Mitglieder:** Steffen Altmann, Manfred Antoni, Nikos Askitas, Hanna Brenzel, Matthias Collischon, André Diegmann, Alfred Garloff, Christina Gathmann, Hans-Martin von Gaudecker, Ines Helm, Stefan Mangelsdorf, Tatjana Mika, Dana Müller, Markus Nagler, Michael Oberfichtner, Nico Pestel, Regina T. Riphahn, Judith Saurer, Daniel Schnitzlein, Stefanie Seele, Gesine Stephan, Eduard Storm, Heiko Stüber, Stephan Thomsen

Inhalt

Vorwort	4
1. Zentrale Bedarfe	5
2. Zusammenführung von Daten	6
2.1 Motivation und Problemlage	6
2.2 <i>Beispiele für Datenbedarfe</i>	8
2.3 Konkrete Vorschläge	10
3. Fehlende Daten	12
3.1 Motivation und Problemlage	12
3.2 Konkrete Vorschläge	12
4. Datenschutz	13
4.1 Motivation und Problemlage	13
4.2 Konkrete Vorschläge	14
5. Zugangsbedingungen	18
5.1 Motivation und Problemlage.....	18
5.2 Konkrete Vorschläge	19
Literaturverzeichnis	22
Mikrodatenzugang in Österreich.....	23
Mikrodatenzugang in den Niederlanden.....	26
Mikrodatenzugang in Dänemark.....	29
Mikrodatenzugang in Schweden.....	32

Vorwort

Exzellente empirische Forschung und evidenzbasierte Politikberatung basieren auf der Verfügbarkeit und hohen Qualität von Daten – heute mehr denn je. Eine im Januar 2023 durchgeführte Erhebung unter den Mitgliedern des Vereins für Socialpolitik ergab, dass viele Forschende die Möglichkeiten in Deutschland, Daten für Forschung zu nutzen, für unzureichend halten. Für 80 Prozent der Antwortenden ist der Datenzugang in Deutschland schlechter als in vergleichbaren Ländern.

Zwar gab es in den letzten zwei Jahrzehnten Verbesserungen beim Forschungsdatenzugang, etwa durch den Aufbau einer Infrastruktur, die den Datenzugang für Forschende über Forschungsdatenzentren erleichtert hat. Jedoch blieben diese Verbesserungen in den Grenzen der bestehenden engen gesetzlichen Rahmenbedingungen. Die heutigen Datenbedarfe gehen deutlich über die existierende Infrastruktur und den gesetzten Rahmen hinaus: Die Gesetzeslage selbst muss angepasst werden, um leistungsstarke Forschung zu ermöglichen und Politik und Gesellschaft hochwertig auf wissenschaftlicher Basis beraten zu können. Gleichzeitig sollten Daten Zugangsregeln die Replizierbarkeit wissenschaftlicher Ergebnisse erleichtern.

Die vorliegende Stellungnahme wurde von einer Unterarbeitsgruppe des Vereins für Socialpolitik verfasst und bildet die Position der Fachgesellschaft ab. Der Text stellt die Desiderate im Bereich Arbeitsmarkt und Sozialversicherung in vier inhaltlichen Feldern dar. Dabei liegt der Fokus jeweils auf der Beschreibung der Problemlage und der Ableitung konkreter Vorschläge, die u.a. im geplanten Forschungsdatengesetz der Bundesregierung umgesetzt werden sollten.

Kapitel 2 wendet sich der fehlenden Verknüpfbarkeit von vorhandenen Daten in Deutschland zu. Es beleuchtet detailliert die Ursachen, warum diese Situation für Forschung und evidenzbasierte Politikberatung ein großes Problem ist und wie die Lage verbessert werden kann. Kapitel 3 geht auf die fehlende Bereitstellung von Daten ein. Kapitel 4 diskutiert Restriktionen für den Datenzugang, die sich durch die in Deutschland gewählte Form der Umsetzung des Datenschutzes ergeben. Auch dort, wo bereits Daten verfügbar sind, bestehen Verbesserungsbedarfe. Diese werden in Kapitel 5 thematisiert.

Der Text wird durch Textboxen ergänzt, die einen Blick auf die Situation des Datenzugangs in Nachbarländern eröffnen, die ebenfalls der EU-Datenschutzgrundverordnung unterliegen. Dies verdeutlicht, wie weit Deutschland im internationalen Vergleich zurück liegt. Die resultierenden Kernforderungen und zentralen Bedarfe werden in Kapitel 1 zusammengefasst.

1. Zentrale Bedarfe

Aus der Analyse der bestehenden Situation des Datenzugangs in Deutschland und in vier benachbarten EU-Partnerländern ergeben sich folgende Bedarfe, die im Zuge des geplanten Forschungsdatengesetzes und mit Hilfe von Änderungen anderer Gesetze adressiert werden sollten:¹

1. Eine qualitative Verbesserung des Mikrodatenangebots in Deutschland erfordert die **Verbreitung und pseudonymisierte Nutzbarmachung einheitlicher Identifikationsnummern** für natürliche und juristische Personen einschließlich Unternehmen und Betriebe, damit ohne großen technischen Aufwand Informationen aus verschiedenen Bereichen miteinander verknüpft werden können. [3.2.c]
2. **Forschung als Zweck von Datenverknüpfung sollte rechtskreisübergreifend gesetzlich verankert werden**, um den Zugang zu und die Verknüpfung von Daten verschiedener Anbieter für Forschungszwecke zu ermöglichen. Wichtig hierfür wäre eine Erweiterung von § 13a BStatG, die auch Personen- und Betriebsdaten einschließt, um Datenverknüpfungen innerhalb der amtlichen Statistik sowie mit Daten Dritter zuzulassen. [2.3.a]
3. Es sollten **bundeseinheitliche und rechtskreisübergreifende Datenschutzregeln** etabliert werden. Derzeit müssen für die Nutzung eines einzelnen Datensatzes teils bis zu 16 einzelne Ländergenehmigungen eingeholt werden, die jeweils auf Einzelfallbasis entschieden werden. [4.1, 4.2.c] Forschung sollte **keiner zu engen inhaltlichen Zweckbindung bei der Datennutzung** unterliegen. Technische und organisatorische Maßnahmen sollten bei der Bewertung der Anonymisierung eines Datensatzes mit einbezogen werden. [4.2.a, 3.2.c]
4. Der Gesetzgeber sollte **behördliche Datenbesitzer verpflichten, vorliegende amtliche Daten für wissenschaftliche Zwecke auf Mikroebene zur Verfügung zu stellen**. Einheitliche Identifikationsnummern müssen für Forschungszwecke verfügbar sein, um eindeutige Verknüpfungen individueller Informationen zu ermöglichen. Verwaltungsprozesse sollten konsequent digitalisiert und so standardisiert werden, dass unmittelbar erschließbare Daten entstehen. [3.2.c]
5. Die **Löschung von identifizierenden Hilfsmerkmalen sollte ausgeschlossen** werden. Gesetze sind so anzupassen, dass diese Praxis beendet wird, da andernfalls die Nutzbarkeit vieler Daten für die Forschung unwiederbringlich zerstört wird. [3.2.c]
6. Die finanzielle und personelle **Ressourcenausstattung der Forschungsdatenzentren (FDZ) ist zu verbessern**, damit Daten rasch bereitgestellt werden und die Verarbeitung inklusive Outputkontrolle keinen Flaschenhals mehr bildet. [5.1, 5.2.c]
7. Im Einklang mit Artikel 8 des EU Data Governance Actes muss eine **zentrale Informationsstelle für Forschungsdaten** eingerichtet werden. Diese sollte

¹ Angaben in eckigen Klammern verweisen auf die relevanten Kapitel im Text.

Informationen zu Forschungsdaten bündeln. Zusätzlich kann der zentralen Stelle die Aufgabe eines **Datentreuhänders** zukommen, der Daten aus verschiedenen Quellen zusammenspielt, ohne dass Forschende direkt involviert werden. Für diese Aufgaben bietet sich eine Ausweitung der gesetzlichen Aufgaben des FDZ der amtlichen Statistik an, was mit verbesserter Ressourcenausstattung einher gehen muss. [2.3.b, 4.2.c]

8. Es sollten klare und einheitliche gesetzliche Regelung für den **Remote-Desktop-Datenzugang** für Forschende etabliert werden. [5.2.a]

2. Zusammenführung von Daten

2.1 Motivation und Problemlage

Der Arbeitsmarkt ist von zentraler Bedeutung für Wettbewerbsfähigkeit, Stabilität und gesellschaftlichen Zusammenhalt einer Volkswirtschaft. Wie arbeitsmarktpolitische Maßnahmen und Regelungen ausgestaltet sein sollten, ist dabei regelmäßig Gegenstand öffentlicher Diskussionen.

Eine wichtige Frage ist beispielsweise, wie sich die gesetzliche Rentenversicherung nachhaltig ausgestalten lässt. Die Arbeitsmarktforschung kann hier zentrale Themen empirisch fundieren und Gestaltungshinweise für Wirtschaft, Politik und Administration bereitstellen. Empirisch tragfähige wissenschaftliche Empfehlungen erfordern ein profundes und belastbares Wissen u.a. zur Qualifikation, Gesundheit, Erwerbstätigkeit, Familiensituation und beruflichen und regionalen Mobilität der Erwerbspersonen. Ebenso müssen die Bedarfe von Betrieben und Unternehmen bekannt sein.

Soll zum Beispiel das erwartete Renteneintrittsalter einer Person abgeschätzt werden, sind mindestens folgende Informationen erforderlich: Bildungsweg (insb. Ausbildung, aber auch Weiterbildung), Beschäftigungskarriere (z. B. Zeiten in sozialversicherungspflichtiger Beschäftigung, Zeiten von Erwerbsunterbrechungen), persönliche Lebenssituation (z. B. Familiensituation, Haushaltseinkommen, Pflegeverpflichtungen) und Gesundheit (u. a. chronische Krankheiten, Erwerbsminderungsrisiken, Rehabilitationserfordernisse) sowie lokale und qualifikations-spezifische Arbeitsmarktgegebenheiten. Auf der Basis dieser Informationen aus verschiedenen Quellen lassen sich die Wahrscheinlichkeiten dafür bestimmen, dass Menschen regulär oder vorzeitig in den Ruhestand eintreten. Dies erlaubt es, die Wirkungen und Kosteneffizienz verschiedener Maßnahmen und Interventionen zu untersuchen.

Weitergehende Fragen, wie etwa, ob das Einkommen ausreicht, um nach Eintritt in den Ruhestand ein auskömmliches Leben führen zu können, würden darüber hinaus zumindest Informationen zu betrieblicher Vorsorge sowie der Einkommens- und Vermögenssituation allgemein erfordern.

Bisher ist eine Verknüpfung solcher Informationen nur sehr eingeschränkt oder überhaupt nicht möglich. Maßgebliche Gründe dafür sind unterschiedliche Regelungen etwa bei der amtlichen Statistik und in verschiedenen Bereichen der Sozialversicherungen. Dies gilt für die

Erhebung und Speicherung der Daten, für Beschränkungen der Verknüpfbarkeit von Daten unterschiedlicher Träger (sowohl aus juristischen als auch technischen Gründen) sowie für Einschränkungen in der Verwendung (Zweckbindung). Diese Fragmentierung steht in Verbindung mit dem sogenannten Volkszählungsurteil aus dem Jahr 1983, welches das Recht auf informationelle Selbstbestimmung einführt.

Bisher ist in Deutschland eine umfassende Verknüpfung getrennter Datenerhebungen regelmäßig nicht erlaubt. Stattdessen werden separate Erhebungen zu spezifischen Fragestellungen durchgeführt, obwohl die europäischen Rahmengesetze (DSGVO) gemäß dem Forschungsdatenprivileg (Art. 5, (1), b, DSGVO) und den Verarbeitungsgrundlagen (Art. 6, (1), a-f, DSGVO) die Zusammenführung der Daten grundsätzlich ermöglichen.

Um die Bedarfe am gewählten Beispiel zu illustrieren, wären verknüpfte Daten der Sozialversicherungen (Arbeitslosen-, Renten-, Kranken-, Pflege- und Unfallversicherung) mit Daten und Erhebungen der amtlichen Statistik sowie mit weiteren Register- oder Verwaltungsdaten (zum Beispiel der Familienkassen oder von Trägern der betrieblichen Altersvorsorge) erforderlich. Für umfassende Analysen müssten Informationen auf der Ebene von Unternehmen, Betrieben und Selbstständigen (z. B. Daten aus der (veröffentlichten) Rechnungslegung, Handelsregistern, Steuerstatistik) ergänzt werden. Erst diese Vollständigkeit würde es ermöglichen, Handlungsbedarfe in differenzierter Weise und – noch wichtiger – themenübergreifend zu ermitteln und politische Handlungsempfehlungen abzuleiten.

Der Zugang zu Register- und anderen Verwaltungsdaten (z. B. Sozialdaten und Daten der amtlichen Statistik) ist dabei aus verschiedenen Gründen erforderlich – und häufig auch möglich: Sie weisen erhebliche Vorteile im Vergleich zu Befragungsdaten auf, z.B. die hohe Fallzahl, präzise Verläufe, die nicht mit Erinnerungsproblemen behaftet sind, und Informationen, die in Erhebungen nur unzuverlässig erfasst werden können (z. B. Angaben zu Gehalt, Vermögen oder Gesundheitszustand). Umgekehrt können einige Informationen nur in Befragungen erhoben werden, z. B. subjektives Wohlbefinden und Einstellungen. Allerdings bleibt das Analysepotenzial einzelner Quellen ohne weitere Verknüpfung vorgenannter Daten untereinander beschränkt, weil bspw. in den Daten der Sozialversicherungen in der Regel keine Informationen zum Haushaltskontext enthalten sind. Durch die verschiedenen Zwecke der Erhebungen und die unterschiedlichen Träger und Zuständigkeiten im föderalen System existieren bisher keine personenbezogenen, durchgängigen Lebensverlaufsdaten. Beispielsweise gibt es keine verknüpften Datenquellen im sogenannten „Übergangssystem“: Schülerinnen und Schüler „verschwinden“ nach Verlassen der Schule aus schulbezogenen Datenprodukten. Erst bei Aufnahme einer Ausbildung, einer sozialversicherungspflichtigen Beschäftigung, beim erstem Sozialleistungsbezug, oder einer Arbeitslosmeldung ohne Leistungsbezug tauchen sie in den Datenbanken der Bundesagentur für Arbeit (BA) auf.² Personen, die selbstständig oder verbeamtet arbeiten, können nicht weiter beobachtet werden.

2 Erst durch Verknüpfung verschiedener Quellen könnten Bildungs- und Erwerbsverläufe nachgezeichnet werden. Dies würde vorbeugende Maßnahmen bei Personen mit risikobehafteter Anschlussperspektive sowie ein besseres Monitoring des Übergangssystems erlauben.

Die Zusammenführung von Daten verschiedener Quellen bietet viele Vorteile: (i) Die Verlässlichkeit von Informationen und die Datenqualität wird verbessert, (ii) auf die mehrfache Erhebung gleicher Informationen bei gleichen Erhebungseinheiten – sowohl im Bereich der Register, als auch für ergänzende Befragungen – kann verzichtet werden, was die Antwortlast (*response burden*) für die Befragten verringert und das *once-only* Prinzip³ implementiert und (iii) die Kosten von Datenerhebung und Datenbereitstellung werden reduziert.

In der Gesamtschau der vielfältigen Hemmnisse für eine Verknüpfung der verschiedenen Daten – über Träger und Institutionen hinweg – sind eher deutsche einzelgesetzliche Regelungen verantwortlich als der europäische Rechtsrahmen. Dies zeigt sich beispielsweise darin, dass im Bereich der Sozialgesetzbücher nach den Vorschriften des §75 SGB X oder §282 SGB III Personendaten zusammengeführt werden dürfen (unter bestimmten Bedingungen auch ohne explizite Einwilligung), nicht aber im Bereich des Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz, BStatG).⁴ Auch gibt es bereits Verknüpfungen über Register hinweg, deren Nutzung für die Forschung aber bisher nicht freigegeben ist. Der Abbau von rechtlichen Hürden ist elementar, um die Verknüpfung und wissenschaftliche Nutzung zu gewährleisten.

Europäische Nachbarländer sind an diesem Punkt weiter: So wurde beispielsweise in Österreich mit Hilfe des Forschungsorganisationsgesetzes ein „Austrian Micro Data Center“ errichtet. In diesem Center können Daten, die von Statistik Austria erhoben wurden, über einheitliche Identifikatoren mit Registerdaten für Forschungszwecke zusammengeführt werden. Weitere bekannte Vorreiter in diesem Bereich sind die skandinavischen Länder (siehe Details im Anhang).

2.2 Beispiele für Datenbedarfe

Im Folgenden sind beispielhaft einige Datenbedarfe für Verknüpfungen aufgeführt, die für Forschung und Politikberatung einen erheblichen Mehrwert schaffen könnten.

Ein zentraler Aspekt ist dabei die *Verknüpfbarkeit von Sozialdaten mit Daten der Statistischen Ämter*. Die Bedeutung lässt sich am Beispiel der Evaluation des Mindestlohnes demonstrieren. Der 2015 eingeführte flächendeckende gesetzliche Mindestlohn war eine bedeutende Arbeitsmarktreform und im Vorfeld äußerst umstritten. Umso wichtiger ist es, die Wirkungen des Mindestlohns zutreffend und detailliert zu erfassen. Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg bereitet u.a. individuelle Informationen zu Einkommen der sozialversicherungspflichtigen Beschäftigten auf (in den sogenannten Individuellen Erwerbsbiografien (IEB)), forscht mit diesen Daten und ermöglicht der Wissenschaft Zugang zu diesen Daten in speziell aufbereiteten Datenprodukten. Allerdings werden keine Arbeitsstunden erfasst, so dass Personen, die von Mindestlöhnen betroffen sind, nicht zweifelsfrei

³ Das *once-only* Prinzip fordert, dass Bürger, Bürgerinnen und Unternehmen bestimmte Standardinformationen den Behörden und Verwaltungen nur noch einmal mitteilen müssen

⁴ § 75 SGB X setzt voraus, dass die Daten für das konkrete Forschungsvorhaben auch tatsächlich erforderlich sind. Zudem wird nach Absatz 4 die vorherige Genehmigung der zuständigen obersten Bundes- oder Landesbehörde benötigt. § 75 SGB X regelt dabei Zusammenführungen von Daten nicht explizit, sondern als eine besondere Art der Verarbeitung lediglich implizit.

identifiziert werden können. Die Statistischen Ämter bieten über den Mikrozensus individuelle Stundeninformationen an. Das zugehörige Einkommen ist dort jedoch nur in groben Intervallen enthalten, so dass auch hier betroffene Personen nicht eindeutig zu identifizieren sind. In einer weiteren Datenquelle der amtlichen Statistik sind beide Informationen vorhanden, allerdings wird der Datensatz nur alle vier Jahre erhoben (Gehalts- und Lohnstrukturerhebung GLS / Verdienststrukturerhebung VSE). Auch dies schränkt die Evaluierung der Mindestlohneffekte ein. Von Seiten der Wissenschaft wie auch der Geschäftsstelle der Mindestlohnkommission wurde daher eine Verknüpfung von VSE und IEB gefordert (Himmelreicher et al., 2017). Nach Prüfung zeigte sich, dass diese Verknüpfung aufgrund der Regelungen im BStatG nicht zulässig ist, jedoch technisch möglich wäre.

Anzustreben ist zudem die zeitnahe und möglichst umfassende Verknüpfung der Daten des entstehenden Basisregisters zu deutschen Unternehmen mit den Betriebs- und Personendaten des IAB für die Forschung. Dabei sollte mindestens die Zuordnung von Betrieben zu übergeordneten Unternehmen ermöglicht werden, um innerunternehmerische Mobilität zu identifizieren und analog zum Betriebs-Historik-Panel (BHP) des FDZ des IAB Längsschnittdaten auf Unternehmensebene erzeugen zu können. Sozialversicherungsdaten sollten mit administrativen oder einzelstatistischen Daten zum Unternehmen bzw. Unternehmenserfolg (Umsatz, Wertschöpfung) verknüpfbar sein, um z. B. den Einfluss von Investitionen und Innovationen auf die Beschäftigungs- und Lohnstruktur oder die Wechselwirkungen zwischen der betrieblichen Beschäftigungsentwicklung und Performancemaßen des Unternehmens untersuchen zu können. Eine Möglichkeit, dies umzusetzen, wäre eine Verknüpfung des AFiD-Panels oder ähnlicher Daten mit IAB-Daten (Wagner, 2014). Ein früheres Pilotprojekt hierzu zeigte die Herausforderungen bei der Erstellung kombinierter Firmendaten für Deutschland (KombiFiD, Bender et al., 2007). Teilweise wurden in der Reaktion darauf bereits gesetzliche Grundlagen angepasst (§13a BStatG), teilweise bestehen die Korrekturbedarfe etwa im Hinblick auf Lösungsfristen und die Forderung nach informierter Einwilligung von Unternehmen noch heute.

Zusätzlich gibt es Bedarfe an erst *neu zu erschließenden Daten, die dann mit vorhandenen Daten zu verknüpfen wären* (vgl. Kapitel 3). So sollten die Verwaltungsdaten der Familienkasse der BA für die Wissenschaft erschlossen und (mindestens) mit den administrativen Forschungsdaten des IAB verknüpft werden. Denn mit den bisher häufig genutzten Arbeitsmarktdaten der sozialversicherungspflichtig Beschäftigten kann der Familienkontext nicht abgebildet werden. Damit bleiben Arbeitsangebotsentscheidungen im Haushaltskontext weitgehend unerforscht. Eine systematische Verknüpfung mit den Daten der Familienkasse der BA und/oder den Steuerdaten des Statistischen Bundesamtes könnte diese Lücke weitgehend schließen. Neben verlässlichen Angaben über Geburten und kinderbezogene Transferleistungen ist ein großer Mehrwert dieser Daten, dass sich der Haushaltskontext von Haushalten mit Kindern abbilden lässt (Eltern-Kind-/ intergenerationale Beziehungen, Elternteil-Elternteil-Beziehungen).⁵ Die Erschließung und Verknüpfung z.B. der Verwaltungsdaten der Familienkasse der BA oder der

⁵ So können beispielsweise verheiratete Paare in den administrativen Daten des IAB derzeit nur auf Grund von Adressen und Nachnamen identifiziert werden, siehe Goldschmidt et al. (2017) und Bächmann et al. (2021).

Steuer-/Finanzdaten beim Statistischen Bundesamt (StBA) wird aktuell dadurch behindert, dass die Daten dem Steuergeheimnis oder gesonderten Regelungen des BStatG unterliegen.

2.3 Konkrete Vorschläge

(a) Gesetzliche Voraussetzungen schaffen

a1. Nutzungs- und Verknüpfungsmöglichkeiten erweitern und vereinheitlichen

Das Bundesstatistikgesetz sollte im Zuge der Einführung eines Forschungsdatengesetzes forschungsfreundlich reformiert werden. §13a BStatG erlaubt die Zusammenführung verschiedener Datenbestände zu Unternehmen (z.B. der Deutschen Bundesbank, des Statistikregisters oder Daten nach dem Verwaltungsdatenverwendungsgesetz). Allerdings ist die Forschung als Zweck der Verknüpfung bisher nicht explizit aufgeführt. **Wir empfehlen, Forschung als Zweck der Verknüpfung in den Gesetzestext aufzunehmen, da ansonsten Verknüpfungen für die Forschung mit Verweis auf den Gesetzeswortlaut abgelehnt werden könnten.**

Darüber hinaus sollte explizit ermöglicht werden, die Datenbestände verschiedener Träger (z.B. Bundesbank, Statistisches Bundesamt, IAB, Befragungsdaten von Forschungsinstituten, Sozialdaten) zu Unternehmen wie auch zu Personen zu verknüpfen. Derzeit können Daten, die sich auf Personen beziehen, auf Basis der Rechtsgrundlage des §13a BStatG nicht verknüpft werden, und es existiert keine andere Rechtsgrundlage für die Verknüpfung von Personen- und Betriebsdaten im BStatG. **Wir empfehlen eine deutliche Erweiterung des §13a BStatG zur Verknüpfung von Unternehmens-, Betriebs- und Personendaten verschiedener Träger.**

Für die gesetzliche Unfallversicherung fehlt im SGB VII eine Regelung, die vorliegende Einzeldaten für die Forschung zugänglich macht. Im §206 SGB VII ist eine Weitergabe nur für einen engen Zweckbereich ("*Erkennung oder Prävention neuer Berufskrankheiten*") und unter restriktiven Regelungen (die oberste Bundes- oder Landesbehörde muss die Verarbeitung der Daten genehmigen und die Bundesärztekammer und der Bundesbeauftragte für Datenschutz müssen im Einzelfall angehört werden) möglich. **Wir empfehlen, die Regelungen im §204 SGB VII dahingehend anzupassen, dass Einzelfalldaten zur Forschung weitergegeben und verknüpft werden dürfen.**

Als Zugangsvoraussetzung sollten die Erforderlichkeit und das öffentliche Interesse, das durch eine zentrale Stelle festzustellen ist, als ausreichend betrachtet werden. Der Datenschutz spielt hier natürlich eine wichtige Rolle, die wir in Kapitel 4 gesondert beleuchten.

Auch Fragen der Datenverknüpfung von Registerdaten für die Forschung sollten übergreifend in dem neuen Forschungsdatengesetz geregelt werden (vgl. RatSWD 2023). Dabei sollten die Voraussetzungen für eine Verknüpfung transparent sein und nach objektiven Kriterien wie der Erforderlichkeit oder dem öffentlichen Interesse gestaltet werden. Das trägt dazu bei, dass unterschiedliche Stellen gleichartige Fragestellungen gleichartig behandeln.

Die Daten zur Grundsicherung für Arbeitsuchende, die Teil der Forschungsdaten am IAB sind, enthalten Angaben über alle Mitglieder von Bedarfsgemeinschaften. Im Zusammenspiel mit

weiteren Sozialdaten kann mit dieser Datengrundlage der Haushaltskontext für Untersuchungen genutzt werden. Durch die Einführung der Kindergrundsicherung droht, dass Informationen über Kinder und Jugendliche in diesen Daten künftig nicht mehr oder nur noch lückenhaft erfasst werden. **Wir empfehlen, die gesetzlichen Voraussetzungen dafür zu schaffen, dass die Angaben über Kinder und Jugendliche weiter in die Prozessdaten eingehen und dauerhaft ihrer Bedarfsgemeinschaft zugeordnet werden können.**

a2. Identifizierende Informationen vor Löschung schützen

Die Löschung identifizierender Informationen macht eine Verknüpfung mit anderen Datensätzen oder über die Zeit bei Wiederholungsbefragungen unmöglich, selbst wenn sich politische oder rechtliche Rahmenbedingungen und Zielsetzungen verändern. Das Erkenntnispotenzial wird dadurch erheblich gemindert. Verknüpfungsidifikatoren oder Hilfsmerkmale (z.B. des Registerzensus), die für Erzeugung und Verknüpfung von Forschungsdaten künftig erforderlich sein können, sollten gesetzlich vor Löschung geschützt sein. Sie können getrennt gespeichert werden, so dass etwaigem Missbrauch nahezu ebenso effektiv vorgebeugt wird wie unter der heutigen, restriktiven Rechtslage. **Wir empfehlen, identifizierende Informationen vor Löschung zu schützen.**

(b) Datenzugang über Datentreuhänder ermöglichen

Wir empfehlen die Einrichtung eines unabhängigen Datentreuhänders, um die Verknüpfung von Daten aus verschiedenen Quellen zu ermöglichen. Die Unabhängigkeit soll dabei sicherstellen, dass Interessenkonflikte einzelner Datenproduzenten bzw. zwischen Datenproduzenten zu Lasten der Datennutzenden weitgehend reduziert werden. Um die Unabhängigkeit zu gewährleisten, sollte die Finanzierung aus öffentlichen Mitteln erfolgen, wie es in Luxemburg gerade umgesetzt wird.

Die Aufgabe des Datentreuhänders ist die Verwaltung der Zugangsschlüssel, die exklusiv bei diesem verbleiben, und über die die Verknüpfung der verschiedenen, sonst in unterschiedlicher Form anonymisierten Daten erfolgt. Der Datentreuhänder trägt hierbei die Verantwortung für die hinreichende Anonymisierung der verknüpften Daten. Die Nutzerinnen und Nutzer erhalten in erforderlichlichem Umfang anonymisierte Daten zur Befriedigung ihrer spezifischen Informationsbedarfe.

Die besondere Schutzwürdigkeit der Daten sollte dadurch gewährleistet werden, dass ein Zugang zu den Daten nur zum Zwecke der wissenschaftlichen Forschung und deren Reproduzierbarkeit gewährt wird. Der Zugang zu diesen Daten sollte dabei aber auf Grundlage formeller Kriterien gewährt werden, d.h. einerseits anhand der Nachvollziehbarkeit der Projektdarstellung und des Bezugs zur wissenschaftlichen Forschung sowie andererseits der Wahrung der schutzwürdigen Interessen der Datensubjekte.

Der Treuhänder müsste und könnte zugleich die Verantwortung tragen, einem Missbrauch vorzubeugen. Durch entsprechende Prüfung des Zwecks, der beantragenden Person und unter Abwägung von Nutzen-Kosten-Relation würde die Entscheidung zur Verknüpfung mit dem Datentreuhänder einer prüfenden Instanz übertragen – und wäre nicht per se und von vorn

herein unmöglich. Ein funktionierendes Datentreuhändermodell mit exklusivem Zugang für die Wissenschaft könnte zugleich vertrauensbildend für die Akzeptanz in der Bevölkerung sein.

3. Fehlende Daten

3.1 Motivation und Problemlage

Der vorhergehende Abschnitt hat gezeigt, dass es bereits viele sinnvoll nutzbare Datensätze gibt, denen vor allem die Verknüpfung untereinander fehlt. Gleichzeitig bestehen parallel noch große Lücken in der Bereitstellung von Arbeitsmarktdaten.

3.2 Konkrete Vorschläge

(a) Vorhandene Daten bereitstellen

Wir empfehlen die systematische Bereitstellung vorhandener Daten, die bereits in faktisch nutzbarer Form vorliegen, aber bislang nicht generell für wissenschaftliche Untersuchungen zur Verfügung stehen. Teilweise begründet sich dies durch eine fehlende Rechtsgrundlage. Dies gilt beispielsweise für Daten über Kindergeldzahlungen, die bei der BA vorliegen, jedoch über das FDZ der BA im IAB nicht bereitgestellt werden. Weitere Beispiele betreffen die Daten aus dem Bundesanzeiger oder staatliche Zuschüsse und Subventionen zu Löhnen und betrieblicher Weiterbildung. Für die bildungsökonomische Forschung sind administrative Bildungsdaten auf Länderebene nicht verfügbar. So werden Daten aus den Schuleingangsuntersuchungen zwar erhoben und gesammelt, jedoch nicht für wissenschaftliche Untersuchungen bereitgestellt. Dies wäre etwa über das FDZ der statistischen Landesämter möglich.

(b) Ältere Daten verfügbar halten

Für die empirische Wirtschafts- und Sozialwissenschaft ist es essentiell, Längsschnittdaten verwenden zu können. Dabei werden bei vielen Fragestellungen langfristige Verläufe benötigt, z.B. bei der Betrachtung von Erwerbsverläufen in der Arbeitsmarktforschung. Um lange Verläufe aufbereiten und bereitstellen zu können, müssen Datenhaltende in der Lage sein, Daten langfristig zu speichern und Beobachtungseinheiten zeitkonsistent zu verfolgen. Bei Befragungsdaten, z.B. bei Panelstudien, ist dies aufgrund der individuellen Einwilligung der Befragungspersonen i.d.R. ohne Weiteres möglich.

Bei der Erstellung von Forschungsdaten aus Registerdaten können hier jedoch Herausforderungen entstehen, da die operativen Zwecke, für die die Daten ursprünglich erhoben werden, häufig nur eine relativ kurze Speicherdauer erfordern. **Wir empfehlen daher, wo noch nicht gegeben, rechtliche, finanzielle und technische Voraussetzungen zu schaffen, um eine langfristige Speicherung und Nutzung durch die Wissenschaft zu ermöglichen.** Insbesondere sollten aus der bestehenden Datenschutzregulierung nicht implizit kurzfristige Löschfristen abgeleitet werden. Das Forschungsprivileg der DSGVO sollte genutzt werden, um explizit langfristige Speicherdauern für Forschungsdaten festzulegen.

In manchen Fällen werden amtliche Statistiken, wie z.B. das Ausländerzentralregister, nur bis zu einem bestimmten Zeitpunkt in der Vergangenheit (in diesem Fall bis 1995) digital zugänglich gemacht. Ziel sollte es sein, dass alle zur Verfügung stehenden amtlichen Daten zugänglich gemacht werden, sodass Forschungsarbeiten auch länger zurückliegende Daten nutzen können, um Lehren aus der Vergangenheit ziehen zu können.⁶

(c) Verwaltungsdaten nutzbar machen

Verwaltungsprozesse erzeugen aktuell häufig Daten, die für die Forschung nicht nutzbar sind. So ist beispielsweise im Unternehmensbasisdatenregistergesetz (UBRegG) die Forschung bisher nicht berücksichtigt. Lediglich für § 281 SGB III (Statistikzwecke) und § 18i Abs. 6 SGB IV (Zwecke der Betriebsdatenführung) können die Informationen seitens der BA verwendet werden. Damit ist die Nutzung des Basisregisters für Unternehmen bisher stark auf Verwaltungs- und Statistikzwecke eingeschränkt.

Wir empfehlen Gesetze so anzupassen, dass diese Verwaltungsdaten für die Forschung genutzt werden können. Darüber hinaus sollten Verwaltungen damit beauftragt werden, die Datenbereitstellung mitzudenken und umzusetzen. Es ist nötig, Verwaltungsprozesse so aufzusetzen, dass erschließbare Daten erzeugt werden. Dies wäre der Fall, wenn solche Prozesse konsequent digitalisiert und im Zuge dessen auch standardisiert würden. Ein Beispiel sind die aktuell nicht oder nur schwer nutzbaren Daten zu Anzeigen von Massenentlassungen bei der BA, die dezentral und nicht einheitlich erfasst werden. Um eine Harmonisierung zu ermöglichen, muss die eindeutige und dauerhafte Identifikation von Personen und Unternehmen über Datenquellen hinweg, die Aufbewahrung und Speicherung sowie die Zulässigkeit von Zusammenführungen von Daten einheitlich behandelt werden, soweit es von der Sache her möglich ist. **Das Forschungsdatengesetz sollte dazu beitragen, Regelungen über die Rechtskreise hinweg zu harmonisieren.**

4. Datenschutz

4.1 Motivation und Problemlage

Der Datenschutz ist neben der europäischen Datenschutzgrundverordnung (DSGVO) durch unterschiedliche nationale gesetzliche Vorgaben geregelt. Diese Regelungen sind nicht primär nach Sachgründen, wie der Art der Erhebung (freiwillige oder auskunftspflichtige Befragung versus auskunftspflichtige Angaben im Rahmen eines Antrags- oder Meldeverfahrens) oder der Materie (Schutzgrad der Daten), sondern nach Rechtskreisen geregelt. Relevant für den Datenzugang im Bereich der Arbeitsmarkt- und Sozialforschung sind insbesondere die Regelungen im Bundesstatistikgesetz und in den Sozialgesetzbüchern, wobei die Regelungen teilweise in Spezialgesetzen zu finden sind. Hierdurch entstehen oftmals eng verstandene Zweckbindungen, welche die Datennutzung auf Aufgabenbereiche einzelner Verwaltungen

⁶ Darüber hinaus werden diese Daten an unterschiedlichen Standorten in unterschiedlichen Formaten gelagert. Eine weitere wichtige Verbesserung wäre es, die amtlichen Daten zentral zugänglich zu machen.

beschränken und ihre Auswertung für Forschungsvorhaben und Politikberatung stark begrenzen oder unmöglich machen.

Bundesländerübergreifende Vorhaben werden derzeit durch oftmals voneinander abweichende Landesdatenschutzregelungen oder -auslegungen beeinflusst. Derzeit müssen teilweise Datenschutzbeauftragte aller Bundesländer einzeln zustimmen, wenn spezifische Informationen für die Durchführung eines Forschungsprojekts erforderlich sind.

4.2 Konkrete Vorschläge

Verbesserungen im Datenzugang für die Wissenschaft, die sich auf den Datenschutz konzentrieren, lassen sich für drei Schwerpunkte identifizieren, die im Folgenden erläutert werden.

(a) Forschungsprivileg der DSGVO ausschöpfen

Die Erhebung von Daten für die amtliche Statistik und für Sozialversicherungsträger unterliegen einer Zweckbindung, die gesetzlich geregelt ist. Das bedeutet, die Daten dürfen nur für diese Zwecke erhoben und verwendet werden. Eine restriktive Auslegung der Zweckbindung kann Forschungsprojekte verhindern.

Ein Beispiel: Sozialdaten, die im Rahmen des Meldeverfahrens der Sozialversicherung an die Deutsche Rentenversicherung und Bundesagentur für Arbeit gehen, dürfen für wissenschaftliche Untersuchungen verarbeitet werden. Im Bereich des Sozialgesetzbuches III bedeutet dies, dass die Daten anonymisiert für die Arbeitsmarkt- und Berufsforschung (§282 SGB III Abs. 7) für wissenschaftliche Einrichtungen zur Verfügung gestellt bzw. übermittelt werden dürfen, wenn diese Daten für das Forschungsvorhaben erforderlich sind. Was alles zur Arbeitsmarkt- und Berufsforschung zählt, wird beispielhaft im §282 SGB III aufgeführt, ist aber nicht abschließend geregelt und betrifft insbesondere die Forschung des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Die Auslegung der Zweckbindung für den Datenzugang für Dritte erfolgt restriktiv und umschließt vorrangig Forschungsprojekte, die sich mit den Auswirkungen auf den Arbeitsmarkt beschäftigen. Forschungsprojekte, die sich mit den Auswirkungen des Arbeitsmarktes auf gesellschaftliche Phänomene wie z. B. soziale Teilhabe, Lebenslagen oder soziale Ungleichheit beschäftigen, können diese Daten nicht oder nur über Umwege (z.B. Forschung zu SGB II-Leistungen) nutzen. In diesem Fall ist die Zweckbindung der Forschung bei der Datennutzung enger als die Aufgabe der BA selbst, die durchaus für die Lebenslage der Arbeitslosen zuständig ist. Generell ist die Einschränkung auf einzelne Sparten der sozialen Sicherung problematisch, weil hierdurch Schnittstellenprobleme des Sozialstaats und prekäre Lebenslagen, die in den Aufgabenbereich mehrerer Träger der sozialen Sicherung fallen, nicht untersucht werden können. Damit unterbleiben Impulse für eine Verbesserung der Allokation und Effizienz von öffentlichen Mitteln. So kann etwa die bekämpfte Armut im Längsschnitt nur unzureichend untersucht werden, während die Dunkelziffer der Nichtinanspruchnahme von Sozialleistungen mangels guter Daten von Niedrigeinkommensbeziehenden auf Schätzungen angewiesen ist.

Das Forschungsprivileg in der DSGVO (Artikel 5, (1), b) besagt, dass Daten, die für andere festgelegte, eindeutige und legitime Zwecke erhoben wurden, abweichend vom ursprünglichen Erhebungszweck für wissenschaftliche und statistische Zwecke verarbeitet werden dürfen, sofern eine Rechtsgrundlage (nach Artikel 6, (1) a-f, DSGVO) dafür vorliegt. Um die Regelung nicht unintendiert zu eng zu fassen, sollte das Forschungsdatengesetz Möglichkeiten (etwa nach Artikel 6, (2) und (3), DSGVO) ausschöpfen, um die Verarbeitungsgrundlage möglichst weit zu fassen. Dies gilt insbesondere, da die DSGVO zwar (gemäß Artikel 6, (1) e, DSGVO) die Verarbeitung von Daten erlaubt, die für die Wahrnehmung einer Aufgabe erforderliche ist, die im öffentlichen Interesse liegt oder in Ausübung öffentlicher Gewalt erfolgt. Sie stellt es den Mitgliedstaaten aber frei, hier eigene Regelungen zu treffen (Artikel 6, (2) und (3), DSGVO). Manchmal wird zur Begründung für die Nichtbereitstellung von Daten eine fehlende juristische Ausurteilung oder eine fehlende gesetzliche Grundlage angeführt. **Das Forschungsdatengesetz sollte Klarstellungen formulieren, und definieren, welche zentrale Stelle das öffentliche Interesse feststellt.** Damit können im öffentlichen Interesse liegende Forschungsprojekte ermöglicht werden. Durch breiter gefasste Zweckbindung im Einklang mit dem Forschungsprivileg der DSGVO kann Forschung zur Weiterentwicklung des Sozialstaates ermöglicht werden.

Voraussetzung für eine umfassendere Möglichkeit der Nutzung und Zusammenführung von Personen-, Betriebs- und Unternehmensdaten erscheint daher weniger eine Anpassung der DSGVO, als eine Anpassung bzw. Reform des BStatG und anderer einzelgesetzlicher Regelungen entsprechend dieser Ziele. Insbesondere sollten im Rahmen der Erarbeitung eines Forschungsdatengesetzes Fragen der Datenverknüpfung von Registerdaten und anderen administrativen Daten (wie beispielsweise Daten, die nach Statistikgesetzen erhoben wurden) für die Forschung so geregelt werden, dass Verknüpfungen möglich sind, falls sie für bestimmte Forschungsprojekte erforderlich sind und ein öffentliches Interesse vorliegt.⁷ Die Beurteilung sollte dabei nicht allein von den Rechtsauffassungen beteiligter Stellen abhängen, sondern das öffentliche Interesse wäre vorab zu definieren. **Wir empfehlen, im Registermodernisierungsgesetz eine Forschungsklausel einzuführen (privilegierter Datenzugang für wissenschaftliche Zwecke).** Dies könnte bspw. durch Festschreibung der Verwendbarkeit der SteuerID und UnternehmensID für die Datenverknüpfung zu Forschungszwecken geschehen (vgl. RatSWD 2023).

(b) Flexibilität in der Anonymisierung von Daten nutzen

Bislang entscheidet das Schutzniveau der Daten, welcher Datenzugang möglich ist. Je mehr Informationen in den Daten enthalten sind, desto restriktiver ist der Datenzugang. Technische und organisatorische Maßnahmen helfen, die Geheimhaltung von Einzelangaben zu gewährleisten und fließen so bei der Definition von faktischer Anonymität mit ein. Die faktische Anonymität von Daten setzt sich aus drei Vorkehrungen zur Wahrung der Geheimhaltung von Einzelangaben zusammen: 1) technischen Vorkehrungen, 2) organisatorischen/vertraglichen

⁷ Im Sinne der Erleichterung der evidenzbasierten Politikberatung ist es erwägenswert, etwa in der Bundeshaushaltsordnung eine Evaluationspflicht einzufügen, die dann die Feststellung eines öffentlichen Interesses begründen kann.

Vorkehrungen, sowie 3) dem Informationsgehalt der Daten. Hohe technische und organisatorische Vorkehrungen können dabei ein geringes Schutzniveau der Einzelangaben, bspw. lediglich eine formale Anonymisierung, ausgleichen (Engelter & Sommer 2016). Der Datenzugang über Remote-Desktop-Verbindung erlaubt beispielsweise die Analyse von Daten mit einem geringeren Schutzniveau als dies bei Datenzugangswegen möglich ist, bei denen die Daten „unkontrolliert“ an Datennutzende übermittelt werden. Ein Remote-Desktop-Zugang kann das geringere Schutzniveau der Daten durch die erhöhten organisatorischen und technischen Maßnahmen ausgleichen.

Wir empfehlen daher gesetzlich zu regeln, dass die bei der Datenverarbeitung vorgenommenen technischen und organisatorischen Maßnahmen in die Bewertung einbezogen werden, ob Daten als anonym gelten. Hinsichtlich der technischen und organisatorischen Maßnahmen sollte ein einheitlicher Standard für alle öffentliche Datenproduzenten eingeführt und regelmäßig an aktuelle technische Entwicklungen angepasst werden. Bei der Festlegung der Standards sollte der Datenschutz ebenso wie die empirische Forschung gehört werden.

(c) Datenschutzvorgaben föderal harmonisiert umsetzen

Die Bereitstellung und der Zugang zu Daten für Forschungszwecke werden durch organisatorische Hürden verzögert oder sogar unmöglich gemacht. Diese organisatorischen Hürden sind meist dem Föderalismus geschuldet. So müssen z.B. neue Projekte, die amtliche Statistiken nutzen möchten, zunächst von allen statistischen Landesämtern geprüft und genehmigt werden. Dieses Verfahren, das nicht nur Ressourcen der Landesämter bindet, verzögert den Datenzugang zusätzlich.

Wünschenswert wäre eine einheitliche Regelung des Datenzugangs im Bereich der amtlichen Statistik. Nach §16 BStatG Absatz 6 Nr. 2 ist der Zugriff auf formal anonymisiertes Datenmaterial im Rahmen von Gastaufenthalten in einem Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder möglich. Dies trifft aber nicht auf alle amtliche Mikrodaten zu. Amtliche Statistiken, wie z.B. Schuldaten, werden von Bundesländern uneinheitlich zur Verfügung gestellt, zum Teil als Individualdaten und zum Teil als aggregierte Daten. Dies verhindert eine flächendeckende Analyse. Zudem wird im Bereich Bildung eine mögliche Identifizierung von Bundesländern mitunter verweigert, um Vergleiche zwischen Bundesländern und ihren Schulsystemen zu verhindern. Ein weiteres Beispiel betrifft Wirtschaftsstatistiken. Hierfür ist der Zugang zu formal anonymem Datenmaterial an Gastwissenschaftsarbeitsplätzen nicht für alle Bundesländer gleichermaßen gegeben. Durch die uneinheitliche Datenbereitstellung entsteht Mehraufwand bei der Nutzenden wie auch bei den Statistischen Ämtern.

Wir empfehlen, dass alle Statistischen Landesämter die Normen für Gastaufenthalte gleich interpretieren und eine einheitliche Regelung für alle Statistischen Ämter festgelegt wird. Mittelfristig könnte eine zentrale Stelle vorgesehen werden, die dies übernimmt. Alternativ kann, analog zum Onlinezugangsgesetz (OZG), das "Einer für Alle-Prinzip" im Bundesdatenschutzgesetz eingeführt werden, damit die Genehmigung eines – ggf. im

Rotationsverfahren festzulegenden und daher nicht notwendigerweise immer desselben – Landesdatenschutzbeauftragten für die Datenverarbeitung durch die übrigen Entscheidungsbehörden anerkannt wird. So würden Antragsverfahren beschleunigt, Aufwände seitens der Datenschutzbeauftragten reduziert und Anreize bei den genehmigenden Behörden geschaffen, einheitliche und transparente Antrags-/Entscheidungsprozesse zu schaffen. Eine zentralisiertere Verwaltung von Datenzugängen und -bereitstellungen würde nicht nur zu einer schnelleren Bereitstellung, sondern auch zu einer Einsparung von Ressourcen führen, die dann für andere Aufgaben beim Datenzugang wie etwa die Digitalisierung von Prozessen eingesetzt werden können

(d) Austausch zwischen Forschenden und Datenschutzbeauftragten verbessern

Ein wichtiger Aspekt bei der Umsetzung von verbesserten Möglichkeiten der Datenverknüpfung ist, dass die beteiligten Seiten (z.B. Datenschutzbeauftragte und Forschende) ihr Verständnis konkreter Datenverarbeitungsprozesse und Rahmenbedingungen abgleichen.

So werden in der Wirtschaftsforschung häufig umfangreiche Informationen über Individuen benötigt, um kausale Zusammenhänge zu bestimmen. Als Beispiel wurden oben die Informationen zum Arbeitszeitumfang von Personen genannt, die vom Mindestlohn betroffen sein könnten. Diese Information mag die Wahrscheinlichkeit einer möglichen Deanonymisierung einer Person aus dem Datensatz erhöhen. Jedoch ist das Ziel der Wissenschaft nicht Deanonymisierung, sondern die Abschätzung der kausalen Wirkung des Mindestlohns auf bestimmte Aggregate wie Beschäftigung oder Einkommen. Damit ist es zwar wichtig, dass einer bestimmten Person bzw. einem Pseudonym ein bestimmtes Merkmal zugeordnet werden kann, aber darüber hinaus ist nicht von Interesse, welche Person genau dies betrifft.⁸ Es ist von fundamentaler Bedeutung, dass dieser Umstand seitens der Datenschützer:innen gewürdigt wird. Umgekehrt herrscht auf Seiten der Wissenschaft teilweise Unsicherheit über Zuständigkeiten und maßgebliche rechtliche Grundlagen, wie Zugangsmöglichkeiten, geeignete Treuhänder und Rollendefinitionen, die abgebaut werden sollte.

Empfohlen wird die Schaffung einer Plattform zum institutionalisierten Austausch zwischen Datenschutzbeauftragten und Forschenden. Dies kann Freiräume schaffen, die dem Gebot der Wissenschaftsfreiheit nach Artikel 5 Absatz 3 des Grundgesetzes Rechnung tragen. Zusätzlich können interdisziplinäre Weiterbildungsangebote für beide Seiten hilfreich sein.

⁸ Um Missbrauch zu verhindern, sind natürlich (weiterhin) Regelungen für eine lückenlose Dokumentation der Datenverarbeitung sowie Sanktionen bei eventuellen Verstößen gegen datenschutzrechtliche Vorgaben erforderlich.

5. Zugangsbedingungen

5.1 Motivation und Problemlage

Die Bereitstellung von Daten durch die FDZs erfolgt in der Regel entweder als sogenannte Off-Site Scientific Use Files (SUF), im Gastaufenthalt vor Ort am FDZ oder per kontrollierter Datenfernverarbeitung (Remote-Job-Submission, auch als Remote-Execution bezeichnet). Nur sehr wenige FDZ in Deutschland erlauben bisher den Echtzeit-Zugang zu Forschungsdaten per Remote-Desktop-Verbindung.

Bei der Bereitstellung als Off-Site SUF werden Originaldaten zumeist stark vergrößert oder anderweitig verändert, um die Anonymität der darin abgebildeten Merkmalsträger zu gewährleisten. Mittels dieses Zugangs können Forschende die Daten am eigenen Arbeitsplatz speichern und verarbeiten. Dieser Zugangsweg weist durch den direkten Zugriff Vorteile für Forschende auf, da die Datennutzung nicht durch Öffnungszeiten des anbietenden FDZ begrenzt wird und Auswertungsergebnisse ohne zwischenzeitliche Datenschutzprüfung durch FDZ-Personal weiterverarbeitet werden können. Allerdings sind technische Vorkehrungen zur faktischen Anonymisierung der übermittelten Daten nicht mehr vorhanden, insofern müssen vertragliche Regelungen und eine Reduktion des Informationsgehalts diese gewährleisten. In der Regel bedeutet der letzte Punkt eine erhebliche Reduktion des Analysepotentials der Daten. Weiterhin verursachen Erstellung, Anonymisierung und Bereitstellung eines Off-Site SUF erhebliche Zusatzaufwände für die Mitarbeitenden des jeweiligen FDZ. Dadurch bleibt dem FDZ-Personal weniger Zeit für das sonstige Tagesgeschäft sowie für die Aktualisierung bestehender und die Erschließung neuer (verknüpfter) Datensätze. Zudem erlauben die Datennutzungsverträge vieler FDZs die Verarbeitung ihrer Off-Site SUFs ausschließlich in den Räumlichkeiten der Institutionen der Forschenden. Der Remote-Zugriff auf einen in der IT-Infrastruktur der Institution gespeicherten Off-Site SUF vom Homeoffice der Nutzenden aus ist oft vertraglich ausgeschlossen, obwohl sich eine gesetzliche Einschränkung dieser Art weder im BStatG noch im SGB findet.

Bei der Bereitstellung per Gastaufenthalt oder Remote-Job-Submission verbleiben die Daten physisch in der IT-Infrastruktur des Datenanbieters. Der Vorteil dieser Arten des Datenzugangs liegt in dem höheren Analysepotential von weniger stark vergrößerten Daten.

Bei der Remote-Job-Submission übermitteln Forschende ihre Auswertungsprogramme an ein FDZ, bspw. über ein Browser Interface, d.h. sie müssen nicht zu einem FDZ anreisen. Dort werden die Programme auf den Daten ausgeführt, ohne dass Forschende die Daten selbst sehen. Je nach Grad der Automatisierung dieses Prozesses kann hierdurch eine Entlastung für das FDZ entstehen. Allerdings müssen Forschende ihre Syntax ohne Ansicht der Daten entwickeln, was ein aufwendiger und fehleranfälliger Prozess sein kann.

Während eines Gastaufenthalts können Forschende vor Ort am FDZ mit den bereitgestellten Daten arbeiten, was gegenüber der Remote-Job-Submission eine weit höhere Geschwindigkeit bei der Entwicklung von Auswertungsprogrammen bedeutet, da Fehler unmittelbar korrigiert werden können. Dieser Zugang führt jedoch zu unterschiedlichen Zugangsvoraussetzungen,

da Forschende in der Nähe eines FDZ einfacheren Datenzugang genießen als solche, die weiter weg von einem FDZ wohnen. Darüber hinaus führen Kapazitätsbeschränkungen zu starken Überlastungen einzelner FDZs, so dass Termine teils Monate im Voraus gebucht werden müssen. Gastaufenthalte erfordern eine Betreuung seitens der FDZ-Mitarbeitenden vor Ort, was erhebliche personelle und technische Ressourcen in Anspruch nimmt.

Zurzeit sind jedoch insbesondere die FDZ der amtlichen Statistik unterbesetzt. Dies führt dazu, dass Forschungsprojekte teils nur mit Verzögerung begonnen oder abgeschlossen werden können, was sich wiederum negativ auf die Forschung, ihre Replizierbarkeit und auf die evidenzbasierte Politikberatung auswirkt. Teilweise werden diese personellen Engpässe weiter durch technische Probleme verstärkt, beispielsweise veraltete IT-Ausstattung (insbesondere Software, aber auch Hardware).

Neben den Zugangswegen sind auch die Datenaktualität und schnellere Datenverfügbarkeit für eine evidenzbasierte Forschung zu aktuellen Themen der wirtschaftspolitischen Debatte zentral. Auch hier zeigen sich wieder Konsequenzen der Überlastung entsprechender Stellen. Oft können die Anbieter wichtiger Mikrodatensätze – insbesondere die FDZ der amtlichen Statistik – die jüngsten Daten nur mit erheblicher Verzögerung für die Wissenschaft zur Verfügung stellen.

Wenn beispielsweise die Auswirkungen der Corona-Pandemie auf die Erwerbstätigkeit oder die Effekte der zuletzt hohen Inflation auf die Verteilung des Lebensstandards in der Bevölkerung analysiert werden sollen, benötigt die Wissenschaft möglichst zeitnah Zugang zu Mikrodaten. Auch die Evaluation politischer Entscheidungen – etwa während der Phase hoher Energiepreise – erfordert hochwertige und rasche Datenbereitstellung. Dies ist in Deutschland nicht gegeben: Anfang Juni 2023 waren als aktuellste Datensätze der Mikrozensus als Mikrodatensatz für die Wissenschaft bis 2020, die Lohn- und Einkommensteuerstatistik als Scientific Use File bis 2017 verfügbar, und als weniger anonymisierte Mikrodaten im Gastaufenthalt sogar nur bis einschließlich 2016 verfügbar.⁹

5.2 Konkrete Vorschläge

(a) Remote-Desktop-Datenzugang rechtlich absichern und ausbauen

Auch beim Datenzugang per Remote-Desktop verbleiben die Daten physisch in der IT-Infrastruktur des Datenanbieters, was eine hohe Datensicherheit bietet. Die Forschenden verarbeiten die Daten per direktem Fernzugriff auf den Rechenserver des FDZ. Sie erhalten keine Kopie der Daten, können diese aber in Echtzeit sehen und bearbeiten. Damit können sie produktiv Auswertungsprogramme entwickeln sowie deren Ergebnisse direkt einsehen und weiterverarbeiten (RatSWD 2019). In den vier in diesem Text vorgestellten europäischen Ländern ist der Datenzugang auf diesem Weg möglich. Die Vorteile für Forschende sind erheblich: Daten müssen i.d.R. deutlich weniger stark vergrößert werden und behalten damit ihr volles Analysepotential, die Kosten für Gastaufenthalte entfallen, und Auswertungen

⁹ Für Forschende ist der Zugang zu maschinenlesbaren und aktuellen Daten ebenso bedeutsam wie Informationen zu Veröffentlichungskalendern.

können passend zu den eigenen Arbeitsabläufen und -zeiten vorangetrieben werden. Auch ein FDZ kann von der Einführung eines Remote-Desktop-Zugangs zu den eigenen Daten profitieren, da die Aufwände für die Erstellung von Off-Site SUFs und für die Betreuung von Gastaufenthalten entfallen. Zudem ist zu erwarten, dass die Aufwände für Datenschutzprüfungen von zu exportierendem Output durch das FDZ-Personal sinken. Kurzum, ein FDZ kann mit einem Datenzugang per Remote-Desktop seine Daten mit hohem Analysepotential und gleichzeitig sehr hohem Datensicherheitsniveau anbieten, und das bei reduzierten Transaktions- und Bürokratiekosten für alle Beteiligten.

Die Hindernisse für den Aufbau eines Remote-Desktop-Zugangs zu Forschungsdaten sind finanzieller oder rechtlicher Natur. Ein solcher Zugang stellt höhere Anforderungen an die Leistungsfähigkeit der Server, weil mehr gleichzeitige Zugriffe erfolgen, als wenn nur Gastaufenthalte oder Remote-Job-Submission (hier kann die Startzeit von Berechnungen zentral gesteuert werden) möglich sind. Zudem können hohe Kosten für Software anfallen, die den Remote-Zugriff leistungsfähig und mit hoher Datensicherheit gewährleistet.

Rechtliche Herausforderungen entstehen teilweise dadurch, dass der Remote-Desktop-Datenzugang nicht explizit rechtlich geregelt ist. Dadurch kann bei (Genehmigungs-)Behörden (Ministerien, BfDI etc.) Uneinigkeit hinsichtlich der Zulässigkeit der Datennutzung per Remote-Desktop, bzw. bezüglich der Frage entstehen, unter welchen Rahmenbedingungen eine solche Nutzung zulässig ist. **Wir empfehlen, dass das Forschungsdatengesetz bzw. eine Reform der Datenschutzregelungen durch klare Vorgaben regelt (z.B. über zu gewährleistende technische und organisatorische Maßnahmen), unter welchen Bedingungen Zugriff per Remote-Desktop für Forschende auf Daten wie sie im Gastaufenthalt zur Verfügung stehen (formal anonymisierte Daten nach BStatG §16 Abs. 6 (2)) zulässig ist.** Hierbei ist entscheidend, dass gesetzlich explizit geregelt wird, dass nicht allein der Informationsgehalt der Daten in die Gesamtbewertung einbezogen wird, sondern ebenso die dem Datenschutz dienenden technischen und organisatorischen Maßnahmen. Die DSGVO eröffnet hier einen weiten Spielraum, der aktuell jedoch selten ausgenutzt wird und bisweilen zu sehr restriktiven Entscheidungen führt.

(b) Datenzugang auch vom Homeoffice aus zulassen

Angesichts stark gestiegener und ggf. noch weiter zunehmender Nutzung von Homeoffice auch bei Forschenden in den Wirtschafts- und Sozialwissenschaften, stellt es ein Hindernis dar, wenn bei der Nutzung von Off-Site SUFs oder beim Remote-Desktop-Datenzugang die Nutzung ausschließlich aus den Räumlichkeiten der Institutionen der Forschenden zulässig ist, wie z.B. beim FDZ BA-IAB, dem FDZ des Statistischen Bundesamtes oder dem FDZ der Länder. Wo rechtlich bereits zulässig, sollten FDZ ihren Datennutzenden daher zugestehen, die bereitgestellten Daten auch vom Homeoffice über Remote-Desktop-Regelungen nutzen zu dürfen. Die Daten sollten dabei zwar zur Wahrung der Datensicherheit physisch in den IT-Infrastrukturen der Institutionen der Forschenden bzw. des FDZ verbleiben. Ein Remote-Zugriff auf die Daten ohne physische Übermittlung sollte, sofern eine sichere Übertragung gewährleistet ist, nicht vertraglich verwehrt werden.

(c) Finanzielle und personelle Ausstattung von Forschungsdatenzentren verbessern

Aus allen beschriebenen Herausforderungen und Lösungsvorschlägen wird ersichtlich, dass eine auskömmliche und ggf. an sich ändernde Bedarfe anzupassende finanzielle Ausstattung der FDZ unerlässlich ist. Nur so ist es möglich, FDZ mit ausreichend Personal sowie zusätzlicher oder zeitgemäßer Technik auszustatten. Damit kann, auf ggf. angepasster rechtlicher Grundlage, ein leistungsfähiger und sicherer Remote-Desktop-Datenzugang durch weitere FDZ implementiert werden. Zudem würden zusätzliche personelle Ressourcen ermöglichen, die FDZ bei ihren Serviceangeboten zu unterstützen und die oben beschriebenen Friktionen bei etablierten Datenzugängen und die zeitlichen Verzögerungen bei der Bereitstellung von Daten zu reduzieren. Zu guter Letzt würde dies Datenanbieter befähigen, Mikrodaten nach ihrer Erhebung zügiger für die Wissenschaft aufzubereiten und zur Verfügung zu stellen.

Literaturverzeichnis

- Bächmann, A.-C., Frodermann, C., Lochner, B., Oberfichtner, M. & Trenkle, S. (2021). Identifying couples in administrative data for the years 2001–2014, FDZ-Methodenreport 03/2021 (en), Nürnberg. DOI:[10.5164/IAB.FDZM.2103.en.v1](https://doi.org/10.5164/IAB.FDZM.2103.en.v1)
- Bender, S., Wagner, J. & Zwick, M. (2007). KombiFiD - Kombinierte Firmendaten für Deutschland. Konzeption der Machbarkeitsstudie für eine Zusammenführung von Unternehmensdaten der Statistischen Ämter, des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit und weiterer Datenproduzenten. FDZ-Methodenreport 05/2007, Nürnberg.
- Engelter, M., & Sommer, K. (2016). Die Novellierung des Bundesstatistikgesetzes 2016, *Wirtschaft und Statistik* 2016(6), 11-19. https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2016/06/novellierung-bundesstatistikgesetz-2016-062016.pdf?__blob=publicationFile
- Goldschmidt, D., Klosterhuber, W. & Schmieder, J.F. (2017). Identifying couples in administrative data, *Journal for Labour Market Research* 50, 29-43. DOI [10.1007/s12651-017-0218-4](https://doi.org/10.1007/s12651-017-0218-4)
- Himmelreicher, R., vom Berge, P., Fitzenberger, B., Günther, R. & Müller, D. (2017). *Überlegungen zur Verknüpfung von Daten der Integrierten Erwerbsbiographien (IEB) und der Verdienststrukturerhebung (VSE)*, (RatSWD Working Paper 262/2017). Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://www.konsortswd.de/publikation/wp262-2017/>
- RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2019). *Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger* (RatSWD Output Series, 6. Berufungsperiode Nr. 5). Berlin, <https://doi.org/10.17620/02671.42>
- RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2023), *Positionspapier des RatSWD: Nutzung von Registerdaten für Zwecke der Forschung sichern*, 22.06.2023, Berlin. <https://www.konsortswd.de/wp-content/uploads/RatSWD-Positionspapier-Nutzung-von-Registerdaten.pdf>
- Wagner, J. (2014). Daten des IAB-Betriebspanels und Firmenpaneldaten aus Erhebungen der Amtlichen Statistik – substitutive oder komplementäre Inputs für die Empirische Wirtschaftsforschung? *Journal for Labour Market Research* 47, 63-70. DOI [10.1007/s12651-013-0129-y](https://doi.org/10.1007/s12651-013-0129-y)

Mikrodatenzugang in Österreich

Die Gründung des Austrian Micro Data Centers (AMDC) am 1. Juli 2022 war ein Schlüsselereignis für den Forschungsdatenzugang in Österreich. Seither können Forschende österreichischer Forschungseinrichtungen, die die Zugangsvoraussetzungen gemäß Forschungsorganisationsgesetz erfüllen, über das AMDC einen Antrag auf Zugang zu Mikro- und Registerdaten stellen.

Das Datenangebot umfasst grundsätzlich alle für die Produktion statistischer Ergebnisse von Statistik Austria unmittelbar verwendeten Mikrodaten. Dazu kommen Daten aus der Verwaltung des Bundes, die für die wissenschaftliche Nutzung im AMDC freigegeben sind.¹⁰ Die Freigabe erfolgt durch eine Verordnung auf Basis des Forschungsorganisationsgesetzes. Welche Registerdaten freigegeben werden, obliegt den jeweiligen Ministerien. Meta-Informationen zu verfügbaren Mikrodaten sind in einem Katalog hinterlegt.

Wissenschaftliche Einrichtungen können auch selbst Daten einbringen, die gemäß den Bestimmungen des Bundesstatistikgesetzes mit den angebotenen Mikrodaten verknüpft werden können.

Grundsätzlich bestehen drei Zugangsarten, um Mikrodaten für wissenschaftliche Zwecke zu verwenden:

1. Faktisch anonyme Daten können von Statistik Austria an wissenschaftlich arbeitende Personen in Form von Scientific Use Files abgegeben werden.
2. Fachlich geeignete Personen können auf Basis einer entsprechenden vertraglichen Vereinbarung einen Vor-Ort-Zugang zu Mikrodaten im Safe Center von Statistik Austria erhalten.
3. Dafür berechtigten wissenschaftlichen Einrichtungen kann auf Basis einer entsprechenden vertraglichen Vereinbarung ein gesicherter Onlinezugang (Remote-Desktop-Umgebung) eingeräumt werden. Der Zugang ist für Personen, die in einem Dienstverhältnis zu einer akkreditierten wissenschaftlichen Einrichtung stehen, im Rahmen eines konkreten Forschungsvorhabens und für einen vereinbarten Zeitraum möglich. Die Akkreditierung wissenschaftlicher Einrichtungen gilt für bis zu fünf Jahre und kann anschließend erneuert werden.

Der Mikrodatenzugang für Forschungsvorhaben ist wie folgt strukturiert:

1. Forschende mit einem Dienstverhältnis zu beim AMDC akkreditierten Forschungseinrichtungen stellen Anträge beim AMDC, um Datenzugang für ihre Forschungsvorhaben zu erhalten.

¹⁰ Hierzu gehören bundesgesetzlich vorgesehene Register wie bspw. Registerzählung, Erwerbsstatistik, Geburtsregister, Steuerstatistiken, Wanderungsstatistik, Schul- und Bildungsregister, Hochschulstatistik, Todesursachenstatistik, Unternehmensregister, die über bereichsspezifische Personenkenneichen (bPK) verknüpft werden können.

2. Für Forschungsvorhaben müssen alle beteiligten Forschenden Vereinbarungen zu technisch-organisatorischen Maßnahmen unterschreiben. Für die Prüfung der Einhaltung dieser Maßnahmen ist die Forschungseinrichtung verantwortlich.
3. Wird das Forschungsvorhaben von Statistik Austria akzeptiert, erstellt diese ein Angebot.
4. Auf der Grundlage des von der Forschungseinrichtung und Statistik Austria akzeptierten Angebots wird ein Vertrag erstellt und von beiden Seiten unterzeichnet. Die Datenbereitstellung geht mit erheblichen finanziellen Belastungen für die Forschenden einher.
5. Jedes Projektmitglied erhält einen Remote-Zugang. Sollten an einem Forschungsvorhaben mehrere Institutionen beteiligt sein, müssen alle Institutionen akkreditiert sein; die Lead-Institution beantragt den Datenzugang.

Der Gründung des AMDC gingen mehrere Gesetzesinitiativen voraus. Die Bestimmungen zur Registerforschung wurden mit dem Datenschutz-Anpassungsgesetz 2018 eingeführt. Außerdem wurden das Bundesstatistikgesetz 2000 und das Forschungsorganisationsgesetz novelliert. Dadurch wurde der Zugang zu Mikrodaten der Statistik Austria sowie zu den Registerdaten der Bundesministerien unter kontrollierten Bedingungen gewährt und die Möglichkeit einer Verknüpfung dieser Daten über das AMDC geschaffen.

Die Einrichtung und der Betrieb des AMDC wurde als Aufgabe der Statistik Austria im Bundesstatistikgesetz 2000 verankert. In den §31-§31b wurde der Zugang der Wissenschaft zu Statistikdaten, die Mitwirkung von Statistik Austria als Auftragsverarbeiter für die im Forschungsorganisationsgesetz genannten registerführenden Stellen, sowie die Verknüpfung von Statistik und Registerforschungsdaten geregelt. Damit wird die zuvor dezentral organisierte Registerforschung nun zentral über die Statistik Austria abgewickelt ("one-stop-shop") und der Zugang zu den Statistik- und Registerforschungsdaten über einen Fernzugriff für wissenschaftliche Einrichtungen erlaubt.

Forschungsförderprogrammen zur Nutzung der neu verfügbaren, verknüpfbaren Registerdaten. Wissenschaft und Forschung kommt innerhalb der EU-Datenschutz-Grundverordnung (DSGVO) eine besondere Rolle zu. Die DSGVO hat sich zum Ziel gesetzt „die Grundrechte und Grundfreiheiten natürlicher Personen und insbesondere ihr Recht auf Schutz ihrer personenbezogenen Daten zu schützen und den freien Verkehr personenbezogener Daten innerhalb der Union zu gewährleisten“. Dabei soll „[d]ie Verarbeitung personenbezogener Daten [...] im Dienste der Menschheit stehen“ und „[z]ur Erleichterung der wissenschaftlichen Forschung“ beitragen. Es wurden Öffnungsklauseln (z. B. Art. 89) in der DSGVO vorgesehen, um Datenverarbeitung zu Forschungszwecken zu erlauben.

Um diese Öffnungsklauseln bestmöglich zu nutzen, wurden spezielle Datenschutz- und Datenverarbeitungsbestimmungen vor allem im Forschungsorganisationsgesetz geschaffen. Diese bieten praxisnahe Regelungen für die im öffentlichen Interesse liegenden Archivzwecke,

die wissenschaftlichen oder historischen Forschungszwecke oder die statistischen Zwecke, insbesondere für pseudonymisierte Daten und Regelungen zur Registerforschung.

Zusätzlich zum gesetzlich verankerten AMDC wird mit dem Austrian Socio-Economic Panel ein neues langlaufendes Haushaltspanel bei Statistik Austria aufgebaut. Das ASEP basiert auf der Kombination von Survey- und Registerdaten und bietet damit in Kombination mit dem AMDC neue Möglichkeiten für die empirische Forschung.

Mikrodatenzugang in den Niederlanden

Der Zugang und der Gebrauch von amtlichen Mikrodaten hat in den letzten 10-15 Jahren in den Niederlanden deutlich zugenommen und ist aus der empirischen Forschung, insbesondere in den Sozial- und Wirtschaftswissenschaften, nicht mehr wegzudenken. Datenanbieter ist das Statistische Amt der Niederlande (*Centraal Bureau voor de Statistiek*, CBS). Das CBS ermöglicht Forschenden bereits seit 1994 einen bedingten und kostenpflichtigen Zugang zu Mikrodaten. Die Forschung mit diesen Daten kann seitdem in den CBS-Niederlassungen in Den Haag oder Heerlen durchgeführt werden. Seit 2006 ist der Zugriff per Remote-Desktop möglich.¹¹

Beim Mikrodatenangebot des CBS handelt es sich um verknüpfbare (Register-)Daten auf der Ebene von Personen, Unternehmen und Adressen, mit denen eigenständige statistische Analysen durchgeführt werden können. Diese Daten werden formal anonymisiert zur Verfügung gestellt. Rechtliche Grundlage dafür ist Artikel 41 des CBS-Gesetzes. Die Verknüpfung verschiedener Mikrodatensätze ist technisch gut möglich, da Personen und Unternehmen amtliche Identitätsnummern haben, die in einer Reihe von administrativen Prozessen genutzt werden. Personen, die in den Niederlanden mit Wohnsitz gemeldet sind oder beschäftigt sind, erhalten beispielsweise eine Bürgerservicenummer, die unter anderem (aber nicht nur) Sozialversicherungsnummer und Steuer-ID in einem ist. Grundsätzlich ist der Zugang zu den Mikrodaten beschränkt auf europäische Universitäten, wissenschaftliche Organisationen, Forschungsinstitute und statistische Einrichtungen, die den Zugang beim CBS beantragen.

Der Zugang zu Mikrodaten erfolgt in fünf Schritten:¹²

1. Beantragung

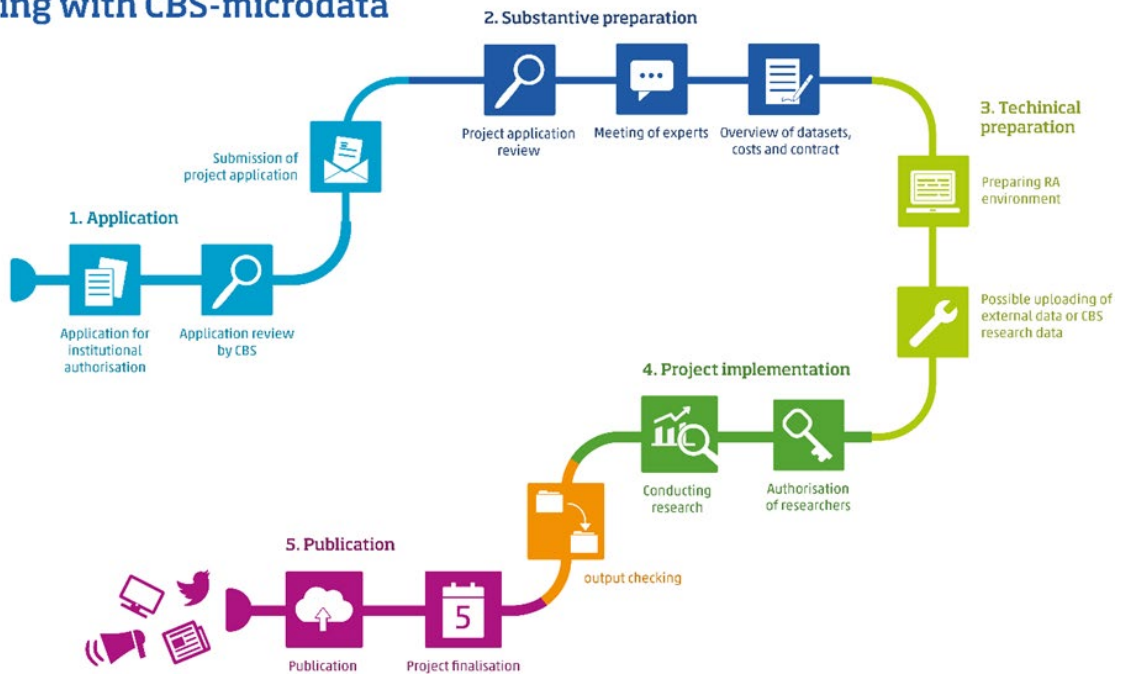
Zunächst muss beim CBS ein Antrag auf institutionelle Zulassung eingereicht werden. Vom CBS wird u. a. geprüft, ob die Haupttätigkeit der Einrichtung in der statistischen oder wissenschaftlichen Forschung besteht und ob die Forschungsergebnisse öffentlich zugänglich gemacht werden. Sobald die Einrichtung eine Genehmigung erhalten hat, können Projektanträge per Formularvordruck elektronisch eingereicht werden, indem eine Beschreibung des geplanten Forschungsprojekts an den CBS Microdata Service gerichtet wird. Darin sollte bereits eine Auflistung der benötigten Datensätze aus dem Mikrodatenkatalog¹³ mitgeliefert werden. Dieser umfasst verschiedenste Themenbereich, u. a. zu Bevölkerung, Arbeitsmarkt und soziale Sicherheit und Bildung.

¹¹ https://privacy-web.nl/wp-content/uploads/po_assets/578514.pdf

¹² <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>

¹³ <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/microdata-catalogue>

Working with CBS-microdata



2. Substanzielle Vorbereitung

Der Projektantrag wird vom CBS anhand einer Reihe von Kriterien geprüft, von denen die wichtigsten die Zulässigkeit und Durchführbarkeit der Forschung sind. Anschließend werden die Antragsteller zu einem Treffen mit den CBS-Experten auf dem jeweiligen Gebiet der Daten, die benötigt werden, eingeladen. In gemeinsamer Absprache wird eine Reihe von Datensätzen festgelegt, auf die für die Forschungsarbeit zugegriffen werden kann. Die nachträgliche Beantragung zusätzlicher Datensätze ist formlos und elektronisch möglich. Alle Daten sind pseudonymisiert und über bedeutungslose Schlüssel miteinander verknüpfbar. Schließlich erstellt das CBS eine Kostenübersicht für das Projekt, einen Vertrag, sowie Vertraulichkeitserklärungen. Die Kosten für die Datenbereitstellung sind nicht unerheblich und können sehr schnell mehrere Tausend Euro jährlich umfassen.¹⁴

3. Technische Vorbereitung

Im Anschluss an die administrativen Vorbereitungen richten CBS-Mitarbeiter für das jeweilige Forschungsprojekt ein Projektkonto in der sicheren Remote-Desktop-Umgebung ein, einschließlich der passenden Zugriffsrechte für Forschende. Optional ist auch das Hinzufügen von externen (verknüpfbaren) Daten möglich. Auch kann CBS einen maßgeschneiderten Datensatz zusammenstellen.

4. Projektdurchführung

Zunächst erfolgt eine Autorisierung der Forschende, die mit den Mikrodaten arbeiten. Neue Forschende erhalten einen Hardware-Token und Informationen über die Arbeit in der Remote-

¹⁴ <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research/services-and-costs>

Desktop-Umgebung. Anschließend müssen sie einen Awareness-Test zu den Regeln des Datenschutzes machen.

Dann kann mit der Analyse der Daten in der eigenen Arbeitsumgebung begonnen werden. Während des Projekts können verschiedene Anpassungen vorgenommen werden, z. B. das Hinzufügen weiterer Forschender oder Datensätze. Für die Analysen steht eine Reihe von Standardsoftwarepaketen zur Verfügung, z. B. Python, R, Stata oder SPSS. Verschiedene andere Pakete sind auf Anfrage erhältlich.

Vor der Herausgabe von Ergebnissen aus der Remote-Desktop-Umgebung erfolgt eine Output-Kontrolle. Gemäß dem niederländischen Statistikgesetz dürfen keine Informationen das CBS verlassen, die auf Personen, Institutionen oder Unternehmen zurückgeführt werden können. Ergebnisse müssen den Ausgaberrichtlinien entsprechen, eine inhaltliche Prüfung der Qualität der Analysen erfolgt nicht.

5. Veröffentlichung

Der letzte Schritt besteht im Abschluss des Projekts. Sobald das Projekt abgeschlossen wurde, wird das Konto geschlossen und in einem Projektarchiv fünf Jahre lang aufbewahrt. Die Ergebnisse von Forschungsarbeiten mit CBS-Mikrodaten müssen Interessenten vollständig, unverzüglich und kostenlos zur Verfügung gestellt werden.

Vor 15 Jahren war der Zugang zu administrativen Daten in den Niederlanden nur unwesentlich besser als in Deutschland. Die Möglichkeiten des Datenzugangs, wie sie schon damals in skandinavischen Ländern bestanden, schienen für beide Länder unerreichbar. Mittlerweile ist das Angebot in den Niederlanden eher mit dem der skandinavischen Länder zu vergleichen. Ein verbleibender wesentlicher Unterschied ist eine (noch) geringere internationale Nutzerbasis. Strukturelle Startvorteile im Vergleich zu Deutschland waren regelmäßig verwendete einheitliche Identifikationsnummern für natürliche und juristische Personen sowie eine deutlich schwächer ausgeprägte föderale Struktur. Entscheidendes Merkmal für den Erfolg scheint aus unserer Sicht jedoch der Wille gewesen zu sein, Mikrodaten der Forschung zur Verfügung stellen zu wollen und ein echtes Servicezentrum dafür zu schaffen. Die Preise sind nicht vernachlässigbar, allerdings werden Anfragen schnell und kompetent beantwortet und technische Lösungen gesucht und gefunden. Ein so gestalteter zentraler Zugangspunkt hat aus Nutzersicht große Vorteile.

Mikrodatenzugang in Dänemark

In Dänemark wird Forschenden der Zugang zu amtlichen Mikro- und Registerdaten über den Forschungsservice des Statistischen Amtes (Statistics Denmark) ermöglicht. Forschende von Hochschulen und autorisierten Forschungseinrichtungen können hierfür auf eine umfangreiche Sammlung von Registerdaten zugreifen, die ein großes Potenzial für wissenschaftliche Analysen bieten. Die Verknüpfung unterschiedlicher Datensätze wird hierbei über die Identifikationsnummern von Personen (CPR-Nummer) bzw. Unternehmen (CVR-Nummer) ermöglicht. Dies erlaubt eine umfassende Verknüpfung amtlicher Daten zu Forschungszwecken (beispielsweise Melde- und Steuerdaten, Daten der elektronischen Gesundheitsakte, Sozial- und Rentenversicherung). Darüber hinaus bietet der Forschungsservice von Statistics Denmark die Möglichkeit, Individualdaten aus externen Datenquellen (z.B. Daten aus Haushalts- und Unternehmensbefragungen, Feld- oder Laborexperimenten) über individuelle Identifikationsnummern mit amtlichen Registerdaten zu verknüpfen.

Die umfassende Verfügbarkeit von Mikrodaten, die hohe Datenqualität und die Verknüpfbarkeit von Individualdaten aus verschiedenen Datenquellen hat Dänemark zu einem Vorreiter in der Arbeitsmarktforschung sowie der empirischen Wirtschaftsforschung und evidenzbasierten Politikevaluation gemacht. Trotz der umfangreichen Datenverfügbarkeit wird die Privatsphäre der Einzelpersonen in Dänemark streng geschützt. Vor dem Zugriff durch Forschende werden die Daten entweder anonymisiert oder pseudonymisiert. Darüber hinaus erfolgt der Datenzugriff in Rahmen eines mehrstufigen Zulassungs- und Prüfverfahrens, das die Einhaltung der geltenden Registergesetzgebung und den Datenschutzgrundsätzen von Statistics Denmark sicherstellt.

Zugang zu Mikrodaten über den Forschungsservice von Statistics Denmark

1. Institutionelle Zulassung: Um als Forschende im Rahmen des Forschungsdatenservice auf Mikrodaten zugreifen zu können, muss zunächst die Forschungseinrichtung der Forschenden für den Forschungsservice zugelassen werden. Hierfür muss die Forschungseinrichtung eine Vereinbarung mit Statistics Denmark über den Zugang zu pseudonymisierten Mikrodaten abschließen. Für eine institutionelle Zulassung kommen in Dänemark ansässige, öffentlich finanzierte Forschungseinrichtungen (Universitäten, Forschungsinstitute, Ministerien, Behörden usw.), allgemeinnützige Stiftungen, Nichtregierungsorganisationen, sowie (Beratungs-)Unternehmen in Frage, wobei letztere keinen Zugriff zu Mikrodaten mit Unternehmensdaten erhalten.

Um für eine Zulassung in Frage zu kommen, muss die Forschungseinrichtung einen hohen und anhaltenden Fokus auf Datensicherheit haben. In der Forschungseinrichtung muss außerdem eine Forschungsdatenmanagerin bestimmt werden, die dafür verantwortlich ist, die Einhaltung der Regeln in der Zulassungsvereinbarung kontinuierlich zu überwachen, einschließlich der Gewährleistung, dass alle Personen mit Datenzugang mit den Regeln zum Zugriff auf Mikrodaten vertraut sind. Schließlich muss die Forschungseinrichtung sicherstellen, dass

mehrere Personen über Fähigkeiten im Umgang mit Registerdaten und über solide Kenntnisse der einschlägigen Datenschutzregeln verfügen, sodass auch im Fall von Personalwechseln das entsprechende institutionelle Wissen in der Forschungseinrichtung erhalten bleibt.

Ausländische Forschungseinrichtungen können nicht für den Forschungsdatenservice zugelassen werden. Im Ausland ansässige Forschende können aber unter bestimmten Voraussetzungen – insbesondere im Rahmen von Kooperationen mit Wissenschaftler:innen in Dänemark – auf Mikrodaten zugreifen.

2. Projektantrag: Wenn die institutionelle Zulassung für die Forschungseinrichtung eines Forschenden vorliegt, können Forschende im Rahmen von Projektanträgen den Zugriff auf Registerdaten beantragen. Um ein neues Projekt zu erstellen, muss zunächst ein Projektantrag im Datenportal von Statistics Denmark gestellt werden.

Im Rahmen des Datenantrags stellen die Forschenden dar, welches Analyseziel sie verfolgen und welche Registerdaten sie für diese Analyse benötigen. Falls die Verknüpfung mehrerer Datensätze beantragt wird, muss diese außerdem begründet werden. Die Registerdaten von Statistics Denmark können für Forschungszwecke außerdem mit externen Daten verknüpft werden. Dies kann Daten aus selbst durchgeführten Umfragen umfassen oder Auszüge aus Registern, die sich außerhalb von Statistics Denmark befinden (in der Arbeitsmarktforschung beispielsweise Daten aus der Agentur für Arbeitsmarkt und Beschäftigung). Bevor externe Mikrodaten an den Forschungsservice von Statistics Denmark übermittelt werden können, muss ihre Verwendung in einem bestimmten Projekt genehmigt werden. Die externen Mikrodaten müssen hierbei Gegenstand eines genehmigten Projektvorschlags sein, bei dem überprüft wird, ob die externen Daten ausschließlich Informationen enthalten, die für das spezifische Projekt verwendet werden sollen.

Der Projektantrag wird üblicherweise in einem mehrstufigen Verfahren geprüft. Zunächst wird innerhalb der Forschungseinrichtung der Antragsteller durch die verantwortliche Forschungsdatenmanagerin und ggf. weitere Parteien (z.B. die zentrale Datenschutzbeauftragte, Ethikkommission) überprüft, ob der Antrag mit den internen Regelungen sowie der institutionellen Zulassungsvereinbarung konform ist. Anschließend wird der Projektantrag von Statistics Denmark im Hinblick auf die Einhaltung der technischen und rechtlichen Anforderungen überprüft.

3. Projektdurchführung: Nach der Genehmigung eines Projektantrags wird ein Projektordner erstellt, auf den ausschließlich die beteiligten Forschenden zugreifen können. Der eigentliche Zugriff auf die Daten erfolgt über eine sichere Remote-Desktop-Umgebung. Damit analysieren die beteiligten Forschenden die Mikrodaten direkt auf den Servern von Statistics Denmark. Um Analyseergebnisse zu extrahieren, können Forschende die Ergebnisdateien an ihre persönliche Email-Adresse bei ihrer Forschungseinrichtung senden. Der Versand ist hierbei auf Wissenschaftler:innen beschränkt, die bei der in Dänemark ansässigen Forschungsinstitution angestellt sind. Wissenschaftliche Assistenten und im Ausland ansässige Koautoren können Zugriff auf Mikrodaten erhalten (s.o.), aber keine Dateien exportieren.

Beim Export von Analyseergebnissen existiert ein striktes Verbot, Mikrodaten oder jegliche Analyseergebnisse, mithilfe derer potentiell individuelle Wirtschaftssubjekte identifiziert werden könnten, zu exportieren. Statistics Denmark überwacht kontinuierlich mithilfe von Algorithmen, ob zum Export anstehende Dateien Informationen enthalten, die diese Regeln verletzen. Falls eine Verletzung der Regeln festgestellt wird, wird der Zugang der beteiligten Forschungseinrichtung temporär gesperrt, um die Verletzung der Datenschutzregeln zu überprüfen. Im Fall von wiederholten Verletzungen durch Forschende derselben Forschungseinrichtung kann die institutionelle Zulassung der Forschungseinrichtung dauerhaft entzogen werden.

Weitere Informationen zum Forschungsdatenservice von Statistics Denmark:

<https://www.dst.dk/en/TilSalg/Forskningservice>

Mikrodatenzugang in Schweden

In Schweden werden die meisten amtlichen Daten zentral über *Statistics Sweden* (SCB) verwaltet. Das Datenangebot umfasst grundsätzlich alle verfügbaren Mikro- und Registerdaten (zum Beispiel Meldedaten, Steuerdaten, Arbeitsmarktdaten, Unternehmensdaten). Manche Daten, die in anderen Registern verwaltet werden, können zugespielt werden. Gesundheitsdaten sind zum Beispiel über Socialstyrelsen (Nationale Behörde für Gesundheit und Soziales) verfügbar. Die bereitgestellten Daten sind pseudonymisiert. Grundsätzlich können alle Daten zu Forschungszwecken miteinander verknüpft werden, entweder über die *personnummer* (persönliche Identifikationsnummer, die jeder erhält der in Schweden gemeldet ist), über die *organisationsnummer* (Firmennummer), oder über die *arbetsställennummer* (Betriebsnummer). Informationen zu Schülern und Lehrern können über den *skolenhetskod* (Schulidentifikationsnummer) verknüpft werden.

Beantragung

Forschende können Datenzugang beantragen. Bei Forschungsvorhaben, die den Umgang mit sensiblen personenbezogenen Daten beinhalten, muss bereits vor dem Antrag bei SCB ein Ethik-Review bei einem regionalem Ethik-Board eingeholt werden (*Etikprövningsmyndigheten*). Jeder Antrag bei SCB wird dann einer spezifischen Person von SCB zugewiesen, die für die Bearbeitung des Antrags verantwortlich ist und als Ansprechperson im gesamten Prozess dient. Diese Person übernimmt auch die Geheimhaltungsprüfung/Datenschutzprüfung. Normalerweise dauert dieser Prozess zwischen 1 und 6 Monaten. Wenn zusätzliche Daten von anderen Behörden hinzugespielt werden sollen, müssen auch die anderen Behörden eine Geheimhaltungsprüfung vornehmen. SCB übernimmt dann die Verknüpfung und alle Daten werden über SCB bereitgestellt. SCB wendet das Prinzip der vollständigen Kostendeckung an, das bedeutet, dass die tatsächlichen Kosten für die Bearbeitung und Bereitstellung des Antrags in Rechnung gestellt werden. Die Kosten je Antrag liegen üblicherweise zwischen 3.000 und 6.000 EUR.¹⁵

Geheimhaltung

Der Zugang zu amtlichen Mikrodaten werden über das *Offentlighets- und Sekretesslag 16* (Gesetz über öffentlichen Zugang zu Informationen und Geheimhaltung) geregelt. Kapitel 24 regelt die "Geheimhaltung zum Schutz der Privatsphäre in der Forschung und Statistik".

Laut §8 von Kapitel 24 gilt, dass Daten, die für Forschungs- oder Statistikzwecke benötigt werden und keine direkten Rückschlüsse auf Einzelpersonen durch Namen, andere Identifikationsmerkmale oder ähnliche Informationen zulassen, weitergegeben werden dürfen,

¹⁵ Weitere Informationen (in Englisch): <https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/> Verfügbare Mikrodaten (in Schwedisch): <https://www.scb.se/vara-tjanster/bestall-data-och-statistik/bestalla-mikrodata/>

¹⁶ https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/offentlighets--och-sekretesslag-2009400_sfs-2009-400

wenn klar ist, dass die Offenlegung der Daten keinen Schaden oder Nachteil für die betroffene Person oder nahestehende Personen verursacht.

Laut § 3 des Forschungsgesetzes gilt, dass wenn eine Behörde in ihrer Forschungstätigkeit von einer anderen Behörde eine geheimhaltungsbedürftige Information erhält, die Geheimhaltungsbestimmung auch bei der empfangenden Behörde anwendbar ist.

Datenzugang

Seit 2005 können Forschende die schwedischen Mikrodaten über MONA (Microdata Online Access) benutzen. Dies ist eine Remote-Desktop-Umgebung, die Zugang zu den Servern von Statistics Sweden ermöglicht. Der Zugang erfolgt über Zwei-Faktor-Authentifizierung, konkret ein Passwort sowie eine Smartphone-App. Der Remote-Zugang kann überall in der EU benutzt werden. Internetnutzung ist auf dem Server nicht möglich.

Resultate werden in einen dafür vorgesehenen Ordner verschoben und dann automatisch innerhalb von wenigen Minuten den Forschenden Online bereitgestellt. Es können alle gängigen Outputformate verwendet werden, zum Beispiel Text-Files, PDF für Grafiken, Tabellen in Word Format. Zusätzlich ist auch der Upload von aggregierten Daten möglich.

Impressum

Kontakt:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Geschäftsstelle

Am Friedrichshain 22

10407 Berlin

office@ratswd.de

<https://www.ratswd.de>

Die Geschäftsstelle des RatSWD wird als Teil von KonsortSWD im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.

Berlin, Dezember 2023



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

DOI: [10.17620/02671.90](https://doi.org/10.17620/02671.90)

Zitationsvorschlag:

Altmann, S., Antoni, M., Askitas, N., Brenzel, H., Collischon, M., et al. (2023). *Für einen besseren Datenzugang im Bereich Arbeitsmarkt und Sozialversicherung. Stellungnahme aus dem Verein für Socialpolitik.*(RatSWD Working Paper 286/2023). Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.90>