

Christina Gathmann, Hans-Martin von Gaudecker, Regina T. Riphahn*

Arbeitsmarktdaten: Gesetzesänderungen gefordert

Exzellente empirische Forschung und evidenzbasierte Politikberatung benötigen hochwertige Daten. Forschende bewerten die Verfügbarkeit und Qualität von Daten in Deutschland als unzureichend, obwohl es in den letzten zwei Jahrzehnten beim Zugang zu Arbeitsmarkt- und Sozialversicherungsdaten Verbesserungen gab. Die Infrastruktur von Forschungsdatenzentren hat den Zugang zu Forschungsdaten im Rahmen enger gesetzlicher Regelungen erleichtert. Die heutigen Datenbedarfe gehen deutlich darüber hinaus: Die Gesetzeslage selbst muss angepasst werden, um leistungsstarke Forschung zu ermöglichen und Politik und Gesellschaft verlässlich auf wissenschaftlicher Basis beraten zu können.

Zusammenführung von Daten

Der Arbeitsmarkt ist von zentraler Bedeutung für die Wettbewerbsfähigkeit und Stabilität einer Volkswirtschaft sowie für den gesellschaftlichen Zusammenhalt. Wie arbeitsmarktpolitische Maßnahmen und Regelungen ausgestaltet sein sollten, ist Gegenstand öffentlicher Diskussionen. Eine wichtige Frage ist, wie sich die gesetzliche Rentenversicherung nachhaltig aufstellen lässt. Um etwa das erwartete Renteneintrittsalter einer Person abzuschätzen, sind Informationen zu Bildungsweg, Beschäftigungskarriere, persönliche Lebenssituation und Gesundheit sowie lokale und qualifikationsspezifische Arbeitsmarktgegebenheiten nötig. Basierend auf den ermittelten Wahrscheinlichkeiten, dass Personen regulär oder vorzeitig in den Ruhestand eintreten, können die Wirkung und Effizienz verschiedener Regelungen und Reformoptionen untersucht werden. Weitergehende Fragen, z. B. ob das Einkommen ausreicht, um nach Eintritt in den Ruhestand ein auskömmliches Leben führen zu können, erfordern darüber hinaus Informationen zu betrieblicher Vorsorge sowie der Einkommens- und Vermögenssituation. All diese Informationen sind prinzipiell vorhanden, werden jedoch in unterschiedlichen Datenquellen erhoben.

Bisher ist eine Verknüpfung solcher Informationen – und damit auch die zur Politikberatung notwendige Analyse – nur sehr eingeschränkt oder überhaupt nicht möglich. Dies liegt an unterschiedlichen Regelungen, etwa bei der amtlichen Statistik und den Sozialversicherungen in Bezug auf die Erhebung und Speicherung von Daten, Beschränkungen bei der Verknüpfbarkeit von Daten unterschiedlicher Träger sowie Restriktionen in der Verwendung (Zweckbindung). Dies ist der Fall, obwohl der europäische Rechtsrahmen die Zusammenführung der Daten grundsätzlich ermöglicht.

Für die Analyse des Renteneintrittsalters etwa wären verknüpfte Daten der Sozialversicherungen mit Daten und Erhebungen der amtlichen Statistik sowie mit weiteren Register- oder Verwaltungsdaten (z. B. den Trägern der betrieblichen Altersvorsorge) erforderlich. Für umfassendere Analysen sind Informationen von Unternehmen, Betrieben und Selbstständigen (z. B. Daten aus der (veröffentlichten) Rechnungslegung oder Steuerstatistik) nötig. Nur eine so verknüpfte Datengrundlage würde es ermöglichen, Handlungsbedarfe differenziert und themenübergreifend zu ermitteln, politische Handlungsempfehlungen abzuleiten und Reformen zu evaluieren.

In Ermanglung einer breiten Datenbasis werden derzeit meist Befragungsdaten für solche Zwecke herangezogen. Diese sollten Register- und anderen Verwaltungsdaten, wie die Daten der amtlichen Statistik, allerdings nur ergänzen, denn letztgenannte weisen erhebliche Vorteile auf, z. B. die hohe Fallzahl, präzise Verläufe, die nicht mit

Prof. Christina Gathmann, Ph.D., ist Leiterin der Arbeitsmarktabteilung am Luxembourg Institute of Socio-Economic Research (LISER) und Professorin an der Universität Luxemburg.

Prof. Dr. Hans-Martin von Gaudecker ist Professor an der Universität Bonn und Team-Leiter am Forschungsinstitut zur Zukunft der Arbeit (IZA) in Bonn.

Prof. Regina T. Riphahn, Ph.D., lehrt Statistik und empirische Wirtschaftsforschung an der FAU Erlangen-Nürnberg und saß bis 2020 dem Rat für Sozial- und Wirtschaftsdaten (RatSWD) vor.

© Der/die Autor:in 2023. Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht (creativecommons.org/licenses/by/4.0/deed.de).

Open Access wird durch die ZBW – Leibniz-Informationszentrum Wirtschaft gefördert.

* Dies ist eine Kurzversion der Stellungnahme der Unterarbeitsgruppe Arbeitsmarkt und Sozialversicherung. Die Langfassung findet sich hier: www.socialpolitik.de/vfs_stellungnahmen_daten_arbeitsmarkt. An der Stellungnahme haben folgende Personen mitgewirkt: S. Altmann, M. Antoni, N. Askitas, H. Brenzel, M. Collischon, A. Diegmann, A. Garloff, C. Gathmann, H.-M. von Gaudecker, I. Helm, S. Mangelsdorf, T. Mika, D. Müller, M. Nagler, M. Oberfichtner, N. Pestel, R. Riphahn, J. Saurer, D. Schnitzlein, S. Seele, G. Stephan, E. Storm, H. Stüber, S. Thomsen.

Erinnerungsproblemen behaftet sind, und Informationen, die in Erhebungen nur unzuverlässig erfasst werden (z. B. Angaben zu Gehalt oder Vermögen). Allerdings bleibt das Analysepotenzial einzelner Quellen ohne Verknüpfung untereinander beschränkt, weil z. B. 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. So existieren keine verknüpften Datenquellen im „Übergangssystem“: Schulingangsuntersuchungen werden regelmäßig gelöscht, Schüler:innen „verschwinden“ nach Verlassen der Schule aus schulbezogenen Datenprodukten und tauchen in den Daten der Bundesagentur für Arbeit (BA) nur auf, wenn sie eine Ausbildung oder eine sozialversicherungspflichtige Beschäftigung aufnehmen, Sozialleistungen beziehen oder arbeitslos gemeldet sind. Studierende oder Personen, die selbständig oder verbeamtet arbeiten, können nicht weiter beobachtet werden.

Die Zusammenführung von Daten verschiedener Quellen bietet viele Vorteile: (i) Die Verlässlichkeit von Informationen und die Datenqualität werden 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-Prinzip¹ implementiert und (iii) die Kosten von Datenerhebung und -bereitstellung werden reduziert. Diese Vorteile können jedoch in Deutschland nicht in vollem Umfang genutzt werden.

Die vielfältigen Hemmnisse für eine Verknüpfung von Daten über Träger und Institutionen hinweg sind in der Regel deutschen einzelgesetzlichen Regelungen und nicht dem europäischen Rechtsrahmen geschuldet. Dies zeigt sich darin, dass im Bereich der Sozialgesetzbücher Personendaten zusammengeführt werden dürfen, nicht aber im Bereich des Bundesstatistikgesetzes. Der Abbau von rechtlichen Hürden ist elementar, um die Verknüpfung und wissenschaftliche Nutzung zu gewährleisten. Europäische Nachbarländer sind an diesem Punkt weiter: So wurde z. B. in Österreich 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 sind die skandinavischen Länder, die Niederlande und Frankreich.

¹ Bürger:innen und Unternehmen müssen bestimmte Standardinformationen den Behörden und Verwaltungen nur noch einmal mitteilen.

Zentral ist die Verknüpfbarkeit von Sozialdaten mit Daten der Statistischen Ämter. Ihre Bedeutung lässt sich am Beispiel der Evaluation des gesetzlichen Mindestlohns aufzeigen, dessen Einführung im Jahr 2015 äußerst umstritten war. Umso wichtiger ist es, seine Wirkungen zutreffend und detailliert zu erfassen. Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) bereitet unter anderem individuelle Informationen zu Einkommen der sozialversicherungspflichtigen Beschäftigten auf (Individuelle Erwerbsbiografien, IEB), forscht mit diesen Daten und ermöglicht der Wissenschaft Zugang zu speziell aufbereiteten Datenprodukten. Allerdings sind die Arbeitsstunden nicht erfasst und Personen, die Mindestlöhne erhalten, können nicht zweifelsfrei identifiziert werden. Die Statistischen Ämter bieten über den Mikrozensus individuelle Stundeninformationen an, erheben das Einkommen jedoch in groben Intervallen, was ebenso eine Identifikation von Betroffenen verhindert. Detaillierte Informationen zu Einkommen und Arbeitsstunden sind in der Gehalts- und Lohnstrukturerhebung/Verdienststrukturerhebung (VSE) der amtlichen Statistik vorhanden. Allerdings wird der Datensatz nur alle vier Jahre erhoben. Dies schränkt die Evaluierung der Mindestlohneffekte erheblich ein. Von Seiten der Wissenschaft wie auch der Geschäftsstelle der Mindestlohnkommission wurde daher eine Verknüpfung von VSE und IEB gefordert (Himmelreicher et al., 2017). Eine Prüfung ergab, dass diese Verknüpfung technisch möglich wäre. Aufgrund von Regelungen im Bundesstatistikgesetz (BStatG) ist sie jedoch nicht zulässig.

Weitere Daten sind prinzipiell vorhanden, aufgrund von gesetzlichen Beschränkungen jedoch nicht für die Forschung zugänglich. Auch diese sollten mit vorhandenen Daten verknüpft werden. Ein Beispiel sind Daten zum Haushaltskontext, die bei der Familienkasse der BA und in den Steuerdaten des Statistischen Bundesamtes vorhanden sind, aber nicht mit Arbeitsmarktdaten der sozialversicherungspflichtig Beschäftigten verknüpft werden können. Nur so aber könnten Arbeitsangebotsentscheidungen im Haushaltskontext, Wechselwirkungen zwischen Elternteilen und intergenerationale Einflüsse erforscht werden.

Konkrete Vorschläge

(a) *Gesetzliche Voraussetzungen schaffen*: Das BStatG sollte im Zuge der Einführung eines Forschungsdatengesetzes forschungsfreundlich reformiert werden. Wir empfehlen, Forschung als Zweck der Verknüpfung in den Gesetzestext aufzunehmen, da sonst Verknüpfungen für die Wissenschaft mit Verweis auf den Gesetzeswortlaut abgelehnt werden können. Wir empfehlen auch eine deutliche Erweiterung des §13a BstatG, um die Datenbestände verschiedener Träger zu Unternehmen, Betrieben und Personen (z. B. von Bundesbank, Statistischem Bundesamt, IAB, Befragungsdaten von Forschungsinstituten, Sozialdaten)

verknüpfen zu können. Das neue Forschungsdatengesetz sollte auch die Datenverknüpfung von Registerdaten für die Forschung übergreifend regeln (RatSWD, 2023). Die Voraussetzungen für eine Verknüpfung sollten transparent und nach objektiven Kriterien wie der Erforderlichkeit oder dem öffentlichen Interesse gestaltet sein, damit unterschiedliche Träger das gleiche Anliegen gleichartig behandeln. Durch die geplante Kindergrundsicherung würde es ohne Möglichkeit der Verknüpfung von Datensätzen zu einer Verschlechterung der Datenlage kommen, da die meisten Kinder nicht mehr in den über das IAB nutzbaren Daten enthalten wären. Selbst die Zahlen im aktuellen Gesetzentwurf sind nahezu im Blindflug entstanden, was die Unsicherheit über die Kosten erhöhte und keinen passgenauen Zuschnitt von Leistungshöhen zuließ. Wir empfehlen, die Verknüpfbarkeit von SGB-II-Daten, Daten der Familienkassen und Steuerdaten explizit zu ermöglichen.

(b) Datenzugang über Datentreuhänder ermöglichen: Wir empfehlen die Einrichtung eines unabhängigen Datentreuhänders, der die Verknüpfung von Daten aus verschiedenen Quellen ermöglicht. Die Unabhängigkeit stellt dabei sicher, dass Interessenkonflikte einzelner Datenproduzenten bzw. zwischen Datenproduzenten zulasten der Nutzenden reduziert werden. Um die Unabhängigkeit zu gewährleisten, sollte die Finanzierung aus öffentlichen Mitteln erfolgen, wie es in Luxemburg gerade umgesetzt wird. Die Aufgabe des Treuhänders ist die Verwaltung der Zugangsschlüssel, die exklusiv bei diesem verbleiben und eine Verknüpfung der verschiedenen anonymisierten Daten ermöglichen. Der Datentreuhänder verantwortet die hinreichende Anonymisierung der verknüpften Daten. Die Nutzenden erhalten in erforderlichem Umfang anonymisierte Daten zur Befriedigung ihrer spezifischen Informationsbedarfe. Der besonderen Schutzwürdigkeit der Daten sollte dadurch Rechnung getragen werden, dass ein Zugang zu den Daten nur zum Zwecke der wissenschaftlichen Forschung und deren Reproduktion gewährt wird. Der Zugang zu diesen Daten sollte dabei aber auf Grundlage formeller Kriterien gewährt werden, d.h. anhand der Nachvollziehbarkeit der Projektdarstellung und des Bezugs zur wissenschaftlichen Forschung sowie der Wahrung der schutzwürdigen Interessen der Datensubjekte. Der Treuhänder sollte zugleich die Verantwortung tragen, Missbrauch vorzubeugen. Durch entsprechende Prüfung des Zwecks, der beantragenden Person und unter Abwägung von Nutzen-Kosten-Relation wird die Entscheidung zur Verknüpfung damit einer prüfenden Instanz übertragen.

(c) Vorhandene Daten bereitstellen: Wir empfehlen die systematische Bereitstellung vorhandener Daten, die bereits in nutzbarer Form vorliegen, aber bislang kaum für wissenschaftliche Untersuchungen zur Verfügung stehen. Teilweise begründet sich dies durch eine fehlende Rechts-

grundlage. Dies gilt z.B. für Daten über Kindergeldzahlungen, die bei der BA vorliegen, jedoch nicht bereitgestellt werden. Weitere Beispiele betreffen etwa die Daten aus dem Bundesanzeiger oder staatliche Zuschüsse und Subventionen zu Löhnen und betrieblicher Weiterbildung. Administrative Bildungsdaten auf Länderebene sind nicht verfügbar, wären aber für die Analyse bildungsspezifischer Arbeitsmarktchancen entscheidend.

(d) Ältere Daten verfügbar halten: Längsschnittdaten sind bei vielen Fragestellungen in den Wirtschafts- und Sozialwissenschaften essenziell, z.B. bei der Betrachtung von Erwerbsverläufen in der Arbeitsmarktforschung. Um langfristige Verläufe bereitstellen zu können, müssen Datenanbieter Daten langfristig speichern und Beobachtungseinheiten zeitkonsistent verfolgen. Bei Befragungsdaten, z. B. bei Panelstudien, ist dies aufgrund der individuellen Einwilligung der Befragungspersonen in der Regel ohne Weiteres möglich. Bei Registerdaten können 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 bestehenden Regelungen nicht implizit kurzfristige Löschrufen abgeleitet werden. Im Gegenteil sollte das Forschungsprivileg der europäischen Datenschutzgrundverordnung (DSGVO) dahingehend genutzt werden, explizit langfristige Speicherdauern für Forschungsdaten festzulegen. In manchen Fällen werden amtliche Statistiken nur ab einem bestimmten Zeitpunkt digital zugänglich gemacht (z. B. Ausländerzentralregister ab 1995). 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.

Datenschutz

Der Datenschutz ist neben der DSGVO durch unterschiedliche nationale gesetzliche Vorgaben geregelt. Dies geschieht nicht primär nach Sachgründen, wie der Art der Erhebung (freiwillige Befragung, auskunftspflichtige Angaben im Rahmen eines Antrags- oder Meldeverfahrens etc.) oder der Materie (Schutzgrad der Daten), sondern nach Rechtskreisen. Relevant für den Datenzugang der Arbeitsmarkt- und Sozialforschung sind insbesondere die Regelungen im Bundesstatistikgesetz und in den Sozialgesetzbüchern, wobei häufig Spezialgesetze zum Tragen kommen. Hierdurch entstehen oftmals eng verstandene Zweckbindungen, welche die Datennutzung auf Aufgabenbereiche einzelner Verwaltungen beschränken und ihre Auswertung für Forschungsvorhaben und Politikberatung oft unmöglich machen. Bun-

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

(a) *Forschungsprivileg konstituieren*: Das Forschungsdatengesetz sollte Klarstellungen formulieren und definieren, welche zentrale Stelle das öffentliche Interesse feststellt. Eine breiter gefasste Zweckbindung im Einklang mit dem Forschungsprivileg der DSGVO würde Forschung zur Weiterentwicklung des Sozialstaates ermöglichen. Wir empfehlen, im Registermodernisierungsgesetz eine Forschungsklausel einzuführen (privilegierter Datenzugang für wissenschaftliche Zwecke). Dies könnte beispielsweise durch Festschreibung der Verwendbarkeit der Steuer-ID und Unternehmens-ID für die Datenverknüpfung zu Forschungszwecken geschehen (RatSWD, 2023).

(b) *Datenzugang vereinheitlichen*: Im Bereich der amtlichen Statistik wäre es wichtig, den Datenzugang einheitlich zu regeln, sowie Harmonisierungen über verschiedene Datenquellen sowie über alle Bundesländer hinweg vorzunehmen. Der Zugang zu formal anonymisierten Wirtschaftsstatistiken sollte für alle Bundesländer gleichermaßen möglich sein. Die gegenwärtig uneinheitliche Datenbereitstellung produziert Mehraufwand sowohl für Nutzende als auch genehmigende und bereitstellende Stellen.

Zugangsbedingungen

Die Bereitstellung von Daten durch die Forschungsdatenzentren (FDZ) erfolgt in der Regel entweder als herunterladbare Scientific-Use-Files, im Gastaufenthalt vor Ort am FDZ oder per kontrollierter Datenfernverarbeitung (Remote-Job-Submission). Nur sehr wenige FDZ in Deutschland erlauben den Echtzeitzugang zu Forschungsdaten per Remote-Desktop-Verbindung. Für eine evidenzbasierte Forschung zu aktuellen wirtschaftspolitischen Debatten sind Datenaktualität und schnelle Datenverfügbarkeit zentral. Wenn beispielsweise die Auswirkungen der Coronapandemie auf die Erwerbstätigkeit oder die Effekte der hohen Inflation auf die Verteilung des Lebensstandards in der Bevölkerung „in Echtzeit“ analysiert werden sollen, benötigt die Wissenschaft sehr zeitnah Zugang zu Mikrodaten. Dies ist in Deutschland nicht gegeben: Anfang Juni 2023 waren

der Mikrozensus als Mikrodatensatz für die Wissenschaft bis 2020, die Lohn- und Einkommensteuerstatistik als Scientific-Use-File bis 2017, und die weniger anonymisierten Mikrodaten im Gastaufenthalt nur bis 2016 verfügbar.

(a) *Remote-Desktop-Datenzugang ausbauen*: Beim Zugang per Remote-Desktop verbleiben die Daten physisch in der IT-Infrastruktur des Anbieters, was eine hohe Datensicherheit bietet. Die Forschenden verarbeiten die Daten per direktem Fernzugriff auf die Rechnerserver des FDZ. Sie erhalten keine Kopie der Daten, können diese aber in Echtzeit sehen und bearbeiten. Damit können sie produktiv Auswertungsprogramme entwickeln, deren Ergebnisse einsehen und weiterverarbeiten (RatSWD, 2019). Analysepotenzial und Datensicherheit bleiben hoch, Transaktions- und Bürokratiekosten im Vergleich zum Gastaufenthalt gering. Die Hindernisse für den Aufbau eines flächendeckenden Remote-Desktop-Zugangs sind finanzieller und rechtlicher Natur. Erforderlich sind leistungsfähige Server und Softwarelizenzen. Rechtliche Herausforderungen entstehen dadurch, dass der Remote-Desktop-Datenzugang nicht explizit geregelt ist und dadurch bei (Genehmigungs-)Behörden (Ministerien, BfDI etc.) Uneinigkeit besteht. Wir empfehlen, dass das Forschungsdatengesetz bzw. eine Reform der Datenschutzregelungen klare Vorgaben macht, unter welchen Bedingungen Forschenden der Zugriff per Remote-Desktop auf Daten möglich ist. Die DSGVO eröffnet hier einen weiten Spielraum, der aktuell jedoch selten genutzt wird.

(b) *Finanzielle und personelle Ausstattung von Forschungsdatenzentren verbessern*: Die finanzielle Ausstattung der FDZ muss so gestaltet sein, dass flexibel veränderlichen Bedarfen begegnet werden kann (durch ausreichend Personal sowie zeitgemäße Technik). Zusätzliche personelle Ressourcen können Friktionen bei etablierten Datenzugängen und zeitliche Verzögerungen bei der Bereitstellung von Mikrodaten für die Wissenschaft reduzieren.

Literatur

- Himmelreicher, R., P. vom Berge, B. Fitzenberger, R. Günther und D. Müller (2017), Überlegungen zur Verknüpfung von Daten der Integrierten Erwerbsbiographien (IEB) und der Verdienststrukturerhebung (VSE), *RatSWD Working Paper Serie*, 262.
- RatSWD – Rat für Sozial- und Wirtschaftsdate (2019), Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger, *RatSWD Output*, 5(6), <https://doi.org/10.17620/02671.42> (1. November 2023).
- RatSWD – Rat für Sozial- und Wirtschaftsdate (2023), Nutzung von Registerdaten für Zwecke der Forschung sichern, 22. Juni.

Title: *Labour Market Data: Legislative Changes Needed*

Abstract: *In recent years, access to German research data on labour markets and social insurances has improved substantially. Nevertheless, reforms are urgently needed to catch up with other countries. Among the most important measures, the German government should legalise the merging of information from heterogeneous data sources for research, make more register available for research, harmonise the handling of data protection, and make remote desktop data accessible more broadly. Research data availability is essential for a functioning and informed democracy that aims to use its resources wisely and to learn from experience.*