

6

ZEITGEMÄSSE DATENINFRASTRUKTUR FÜR FUNDIERTERE ENTSCHEIDUNGEN

- I. Motivation: Bessere Daten für bessere Entscheidungen**
- II. Ausgangslage: Verbesserungsbedarf bei Datenerhebung, -verknüpfung und -zugang**
 - 1. Lückenhafte Erhebung von Daten
 - 2. Zeitverzug bei der Datenbereitstellung
 - 3. Verfügbarkeit existierender Daten eingeschränkt
 - 4. Datenverknüpfung bestenfalls auf Einzelfallbasis
 - 5. Datenzugang nicht benutzerfreundlich
- III. Handlungsoptionen: Eine Dateninfrastruktur für das 21. Jahrhundert**
 - 1. Rahmenbedingungen des Statistischen Verbunds verbessern
 - 2. Datenlücken schließen
 - 3. Geringerer Zeitverzug und höhere Anpassungsfähigkeit
 - 4. Verfügbarkeit bereits erhobener Daten verbessern
 - 5. Bessere Verknüpfbarkeit über verschiedene Datenanbieter hinweg
 - 6. Einfacher und benutzerfreundlicher Datenzugang

Anhang

Literatur

WICHTIGSTE BOTSCHAFTEN

- Die deutsche Forschungsdateninfrastruktur ist im internationalen Vergleich rückständig. Daten fehlen, Datenverknüpfungen sind kaum möglich, und der Zugang ist wenig benutzerfreundlich.
- Die Statistikgesetzgebung sollte regeln, welche Informationsbedarfe die amtliche Statistik erfüllen soll, und ihr Freiheit bei den konkreten Erhebungen lassen, um sie flexibler und reaktionsschneller zu machen. Die Ressourcen der amtlichen Statistik müssen erweitert werden.
- Ein Forschungsdatengesetz sollte zügig auf den Weg gebracht werden. Der Forschung sollte darin ein hohes Gewicht in der Güterabwägung zum Datenschutz eingeräumt werden.

DAS WICHTIGSTE IN KÜRZE

Datenbasierte Analysen sind essenziell für fundierte Entscheidungen in der Politik und öffentlichen Verwaltung. Evidenzbasierte empirische Forschung und Politikberatung sind auf zuverlässige Daten angewiesen, um Politik, öffentliche Verwaltung und Öffentlichkeit über Wirkungszusammenhänge aufzuklären, die Folgen geplanter Maßnahmen abzuschätzen sowie die Effektivität und Effizienz implementierter Maßnahmen zu evaluieren. Dafür ist eine umfangreiche, einfach zugängliche Dateninfrastruktur erforderlich. Sie muss anpassungsfähig sein, um kurzfristig auf neu entstehende Informationsbedarfe zu reagieren. Investitionen, die die amtliche Statistik dazu ertüchtigen, eine solche Dateninfrastruktur bereitzustellen, sind Investitionen in besser fundierte, zielgenauere und kosteneffizientere Politikentscheidungen.

Die Dateninfrastruktur in Deutschland ist im internationalen Vergleich rückständig. Es bestehen Defizite im Umfang der verfügbaren Daten sowie deren zeitnahe Verfügbarkeit. Manche Daten werden gar nicht erhoben. Zudem sind bestehende **administrative Daten** aus der Verwaltung für die Wissenschaft und Politikberatung **nicht zugänglich**. Gleichzeitig können **existierende Daten** verschiedener Datenproduzenten nur in **sehr eingeschränkter Form** zu neuen Datenprodukten **verknüpft werden**, die neue Einsichten ermöglichen. Selbst in den Fällen, wo Daten prinzipiell für die Forschung verfügbar sind, ist der **Zugang** zu ihnen in Deutschland, insbesondere im Bereich der amtlichen Statistik, wenig benutzerfreundlich und **technisch veraltet**.

Um die Forschungsdateninfrastruktur zu verbessern, sind in allen genannten Handlungsbereichen dringend Fortschritte zu erzielen. In einem ersten Schritt sollte das Bundesstatistikgesetz als gesetzliche Grundlage des Verbunds aus Statistischem Bundesamt und den Statistischen Landesämtern (Statistischer Verbund) mit dem Ziel einer **echten Outputorientierung der Statistikgesetzgebung** weiterentwickelt werden. Dazu sollte der Gesetzgeber in Zukunft festlegen, welche Informationsbedarfe bestehen und welche Dienstleistungen der Statistische Verbund erbringen soll. Er sollte ihm jedoch gesetzlich die Freiheit lassen, die dafür notwendigen Erhebungen selbst zu definieren. Darüber hinaus sollten zusätzliche Daten beispielsweise in den Bereichen Bildung und Vermögen erhoben sowie eine hochfrequente Haushaltsbefragung etabliert werden.

Eine **umfangreichere Nutzung administrativer Prozessdaten** könnte das Angebot der amtlichen Statistik kostengünstig erweitern. Ferner sollte ein **Forschungsdatengesetz den Datenzugang verbessern, die Verknüpfungsmöglichkeiten** zwischen Daten verschiedener öffentlicher Datenproduzenten für die Forschung **ausweiten** und der **Forschung bei der Güterabwägung zum Datenschutz ein hohes Gewicht einräumen**. Um den Zugang zu Daten benutzerfreundlicher zu gestalten, sollte der Fernzugriff auf Daten als Standardzugangsweg für Forschungsdaten etabliert werden und das öffentlich zugängliche Datenangebot des Statistischen Verbunds auf den neuesten technischen Stand gebracht werden.

I. MOTIVATION: BESSERE DATEN FÜR BESSERE ENTSCHEIDUNGEN

530. **Datenbasierte Analysen sind eine unverzichtbare Grundlage für gute Entscheidungen in Politik und öffentlicher Verwaltung.** Ohne eine verlässliche Datenbasis ist eine evidenzbasierte Ausgestaltung von Politikmaßnahmen und Rahmenbedingungen nicht möglich. Fehlt eine solche Basis, kann dies dazu führen, dass Maßnahmen beschlossen werden, die die beabsichtigten Ziele gar nicht (ineffektive Maßnahmen) oder nur unter zu hohen Kosten (ineffiziente Maßnahmen) erreichen oder die mit unbeabsichtigten Nebenwirkungen einhergehen.

Daten sind nicht nur **ex ante als Grundlage für eine evidenzbasierte Ausgestaltung** von Politikmaßnahmen notwendig, sondern auch **ex post für die evidenzbasierte Evaluation**. Evaluationsergebnisse können dazu genutzt werden, Maßnahmen sowohl nachzubessern als auch künftig zielgenauer und kosteneffizienter auszugestalten. Dagegen können fehlende Daten Informationsvorteile für Insider und Lobbyorganisationen ermöglichen, die die Politik und öffentliche Verwaltung in einer Situation der Unsicherheit aufgrund fehlender oder unzureichender Daten leichter beeinflussen können.

Daten können vor allem in aggregierter Form direkt von den politischen und regulatorischen Entscheidungsträgern genutzt werden. Sehr häufig werden Entscheidungsträger jedoch durch eine Auswertung und Interpretation von Daten durch die unabhängige Forschung und die evidenzbasierte Politikberatung unterstützt. Die **Bereitstellung der notwendigen Forschungsdateninfrastruktur** ist somit eine **Investition in besser fundierte, zielgenauere und kosteneffizientere Politikentscheidungen**.

531. Die **Dateninfrastruktur in Deutschland ist im Vergleich zu anderen Staaten rückständig**. Dies wurde **während der Corona-Pandemie und der Energiekrise besonders deutlich**. ↘ ZIFFER 540 Während der Corona-Pandemie fehlten beispielsweise **Informationen über** die potenzielle oder tatsächliche Betroffenheit von **Unternehmen am aktuellen Rand** (JG 2021 Kästen 11 und 15). Mit solchen Informationen hätte man die Corona-Hilfen zielgenauer ausgestalten, nahezu in Echtzeit evaluieren und bei Bedarf nachbessern können. In Frankreich wurde hingegen mit der Umsetzung von Unternehmenshilfen direkt deren Monitoring und Evaluation beschlossen. Dazu wurden Daten über die Auszahlung der Hilfen oder die Nutzung von Kurzarbeit erhoben, mit Unternehmensmikrodaten des nationalen Statistikamts verknüpft und von einer unabhängigen Kommission ausgewertet. In Deutschland wird erst jetzt, mehr als drei Jahre nach Beginn der Pandemie, die Dateninfrastruktur zur Evaluation der Unternehmenshilfen geschaffen (Profit, 2023).

Darüber hinaus war in Deutschland keine hinreichend große, kurzfristig anpassbare, **hochfrequente Haushaltsbefragung** verfügbar, mit der man die individuelle Betroffenheit von Haushalten durch die Pandemie hätte messen können.

In den USA hat hingegen das US Census Bureau im Jahr 2020 in kurzer Zeit den Household Pulse Survey und den Small Business Pulse Survey aufgebaut, um die wirtschaftlichen Auswirkungen der Pandemie auf Haushalte und kleine Unternehmen zu messen. Zudem wurde der Current Population Survey (CPS) kurzfristig angepasst, um die Auswirkungen der Pandemie zu bewerten, beispielsweise auf Remote Work oder Corona-bedingte Unternehmensschließungen und Arbeitsplatzverluste (U.S. Bureau of Labor Statistics, 2020). In Norwegen wurden bereits zu Beginn der Corona-Pandemie innerhalb weniger Tage Registerdaten in Echtzeit genutzt, um die Auswirkungen auf die wirtschaftliche Situation von Haushalten zu messen (Alstadsæter et al., 2020). Sie dienen dann zur Information der Öffentlichkeit und als Entscheidungsgrundlage für die Regierung.

In der **Energiekrise** haben fehlende oder **nicht hinreichend detaillierte Daten über die Energieverbräuche und -kosten** der privaten Haushalte am aktuellen Rand eine Einschätzung der finanziellen Auswirkungen der Energiekrise erschwert. Auch in diesem Fall hätte eine **hochfrequente Haushaltsbefragung** Informationen über die unterschiedliche Betroffenheit der verschiedenen Einkommensgruppen und Haushaltstypen liefern und die Ausgestaltung der Unterstützungsmaßnahmen verbessern können. Darüber hinaus fehlten Daten über die Angebotsseite des Energiemarkts. So werden bis heute, anders als beispielsweise in Teilen der USA (Palmer et al., 2022), keine Daten darüber erhoben, welche Kraftwerkstypen wann auf dem Strommarkt den Preis setzen. Infolgedessen war keine direkte Abschätzung möglich, in welchem Umfang die verschiedenen stromproduzierenden Unternehmen von der Krise profitiert haben.

- 532. Zahlreiche Gesetzesvorhaben auf nationaler und europäischer Ebene** zeigen, dass diese Probleme prinzipiell auf politischer Ebene erkannt wurden. Beispiele dafür sind das Registermodernisierungsgesetz, die aktuell diskutierte Weiterentwicklung der nationalen Datenstrategie, das im Konsultationsprozess stehende Forschungsdatengesetz und der europäische Data Governance Act. Diese und andere Initiativen haben das Potenzial, die Datenlage für die Forschung und damit die Entscheidungsgrundlage für die politischen Entscheidungsträgerinnen und -träger zu verbessern. Die **Fortschritte dabei sind jedoch langsam**. Ob bzw. wie sehr diese Gesetzesvorhaben die Datengrundlage verbessern, wird entscheidend davon abhängen, wie die bislang eher groben strategischen Zielsetzungen, beispielsweise der Datenstrategie, konkret gesetzlich umgesetzt werden. Damit die Stärkung der gesetzlichen Rahmenbedingungen tatsächlich zur erwarteten Verbesserung der Forschungsdateninfrastruktur führt, kommt es maßgeblich auf die praktische Umsetzung in der öffentlichen Verwaltung und die Unterstützung der Leitungsebene an.
- 533.** Die Dateninfrastruktur in Deutschland wird maßgeblich durch die **allgemeinen Rahmenbedingungen des Statistischen Verbunds** [↘ PLUSTEXT 22](#) bestimmt. Diese können insbesondere dadurch verbessert werden, dass sich die Statistikgesetzgebung weniger auf die Definition der Inputs der Statistik, also die konkreten Erhebungen, und stärker auf die Outputs, also die angeforderten Informationen fokussiert. Dadurch wäre die amtliche Statistik deutlich flexibler und reaktions-schneller. [↘ ZIFFER 553](#)

Darüber hinaus gibt es **fünf zentrale Handlungsfelder zur Verbesserung der Dateninfrastruktur**. Erstens sollten Datenlücken in Deutschland durch **neue Datenerhebungen** geschlossen werden. [↪ ZIFFER 557](#) Zweitens sollte der Zeitverzug bei der Bereitstellung von Daten verkürzt und die Reaktionsschnelligkeit der Statistik erhöht werden. [↪ ZIFFER 562](#) Drittens sollte die **Verfügbarkeit** bereits erhobener Daten verbessert werden. [↪ ZIFFER 564](#) Viertens sollten die **Möglichkeiten der Verknüpfung**, insbesondere zwischen Daten-sätzen verschiedener Datenproduzenten, erweitert werden. [↪ ZIFFER 568](#) Fünftens sollte der **Datenzugang deutlich einfacher, moderner und benutzerfreundlicher** ausgestaltet werden. [↪ ZIFFER 572](#)



[↪ PLUSTEXT 22](#)

Hintergrund: Der Statistische Verbund – gesetzliche Rahmenbedingungen

Das nationale System der öffentlichen Statistik besteht aus dem Verbund des Statistischen Bundesamts und der Statistischen Ämter der Länder (Statistischer Verbund), der anderen nationalen Datenproduzenten (Other National Authorities, ONAs) wie der Deutschen Bundesbank, die an der Erstellung von europäischen Statistiken mitwirken (Decker und Vorgrimler, 2021), sowie weiterer öffentlicher Institutionen, die Statistiken außerhalb des europäischen Statistiksystems erstellen. Der **Statistische Verbund**, seine Aufgaben sowie die rechtlichen Rahmenbedingungen seiner Arbeit sind **im Bundesstatistikgesetz (BStatG) verankert**. Er stellt den **Kern des nationalen Statistiksystems** dar und übernimmt sowohl national als auch international eine koordinierende Rolle bei der Erhebung, Aufbereitung und Übermittlung von Daten. Der Statistische Verbund **erstellt Bundesstatistiken**, die per Gesetz beauftragt werden (§ 5 BStatG). Darüber hinaus kann dem Statistischen Bundesamt die Aufbereitung von Daten aus dem Verwaltungsvollzug (Prozessdaten) und deren Verarbeitung zu **Geschäftsstatistiken** übertragen werden (§ 8 BStatG).

II. AUSGANGSLAGE: VERBESSERUNGSBEDARF BEI DATENERHEBUNG, -VERKNÜPFUNG UND -ZUGANG

534. Das Angebot und der Zugang zu Forschungsdaten hat sich in Deutschland **seit der Jahrtausendwende deutlich verbessert**. Davor standen den Forschenden hauptsächlich aggregierte Kennzahlen zur Verfügung. Zum Ausbau hat insbesondere der Rat für Sozial- und Wirtschaftsdaten (RatSWD) beigetragen, der seit seiner Gründung bei der Entwicklung der Dateninfrastruktur in diesen Bereichen eine Schlüsselrolle einnimmt. [↪ PLUSTEXT 23](#) Die Anzahl der vom RatSWD akkreditierten Einrichtungen, durch die die Datenproduzenten ihre Daten der Forschung zur Verfügung stellen (Forschungsdatenzentren, FDZ), ist auf nunmehr 42 FDZ im Jahr 2023 gestiegen. Diese **FDZ** decken insbesondere Daten in den Bereichen Sozial- und Wirtschaftswissenschaften, Bildung, Gesundheit und Psychologie ab, aber auch weitere Daten wie die des Kraftfahrtbundesamtes. Sie **sollen Daten aufbereiten**, dokumentieren **und unter Sicherstellung des Datenschutzes** Forschenden zweckgebunden, vertraulich, zeitnah und qualitätssicher **zur Verfügung stellen**. Zur Nutzung der Daten sollen sie verschiedene Zugangswege bereitstellen, im besten Fall über Datenfernzugriff. Die Arbeit des RatSWD ist ein **Teil der Nationalen Forschungsdateninfrastruktur**, die die Daten für das gesamte Wissenschaftssystem systematisch erschließt, vernetzt und nutzbar macht.
535. **Im internationalen Vergleich** ist die deutsche Forschungsdateninfrastruktur jedoch nach wie vor **rückständig**. In den skandinavischen Ländern sind beispielsweise umfassende Registerdaten, sowohl zu Personen als auch zu Unternehmen, verfügbar und auf Anfrage verknüpfbar. Diese Daten decken den Arbeitsmarkt ebenso ab wie das Bildungs- und Gesundheitswesen oder die Einkommens- und Vermögenssituation (Ludvigsson et al., 2019). Zudem können die Daten über Fernzugang genutzt werden. In Österreich werden ähnliche Möglichkeiten durch das Austrian Micro Data Center (AMDC) und in Frankreich durch das Centre d'accès sécurisé aux données (CASD) angeboten (Gottschalk et al., 2023). [↪ PLUSTEXT 24](#)
536. Trotz der Fortschritte, die in den vergangenen 20 Jahren gemacht wurden, **bestehen deshalb weiterhin erhebliche Verbesserungsbedarfe**. Vor dem Hintergrund des Leitbilds des RatSWD, [↪ PLUSTEXT 23](#) das der Sachverständigenrat teilt, gibt es fünf zentrale Probleme und damit **fünf Handlungsfelder**. Erstens werden einige wichtige Daten nicht erhoben (**Datenerhebung**). Zweitens stehen erhobene Daten am aktuellen Rand erst mit deutlicher Verzögerung zur Verfügung und Erhebungen können, insbesondere in Krisen, oftmals nicht schnell an neue Gegebenheiten angepasst werden (**Zeitverzug**). Drittens werden bereits erhobene Daten für die Forschung nur eingeschränkt zur Verfügung gestellt (**Verfügbarkeit**). Viertens sind die Möglichkeiten der Verknüpfung zwischen



➤ PLUSTEXT 23

Hintergrund: Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) und sein Leitbild

Die für Wirtschafts- und Sozialpolitik relevante Forschungsdateninfrastruktur in Deutschland wird maßgeblich vom RatSWD koordiniert. Dieser wurde im Jahr 2004 auf Empfehlung der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik als unabhängiger Beirat der Bundesregierung eingerichtet. Er ist als **gemeinsame Interessenvertretung von Datenproduzierenden**, wie beispielsweise der FDZ der Statistischen Ämter des Bundes und der Länder, der Deutschen Bundesbank und des Sozio-oekonomischen Panels (SOEP) **sowie der Datennutzenden konzipiert**. Damit nimmt er eine zentrale Rolle bei der Weiterentwicklung der Forschungsdateninfrastruktur in den Sozialwissenschaften ein. [➤ ABBILDUNG 143](#) Der Rat übernimmt insbesondere die Aufgabe der Akkreditierung und Evaluation der FDZ in Deutschland nach einheitlichen Kriterien und Mindeststandards. **Leitbild** des RatSWD für moderne und zukunftsfähige Forschungsdatenzentren ist dabei das Konzept von **Open Data für die Forschung**, nämlich der uneingeschränkte Zugang zu administrativen, wirtschaftlichen und wissenschaftlichen Daten für die Forschung, ohne Bezahlung, für Mensch und Maschinen lesbar, mit Metadaten und mit transparenter Lizenz versehen. Erhobene **Daten sollten dabei auffindbar, leicht zugänglich, automatisiert verknüpfbar und nachnutzbar sein**.

1. Lückenhafte Erhebung von Daten

- 537.** In vielen Bereichen ist die Forschung eingeschränkt, weil **Daten gar nicht** oder nicht in ausreichender Qualität **erhoben** werden. So fehlt **beispielsweise** ein länderübergreifendes **Register für Schülerinnen und Schüler oder ein Bildungsverlaufsregister**, das die Daten zu Noten von Zwischen- und Abschlussprüfungen, Schuleingangsuntersuchungen und Ähnlichem für wissenschaftliche Zwecke zugänglich und im Längsschnitt auswertbar machen würde (RatSWD, 2022a; JG 2021 Ziffern 376 ff.). Dies liegt einerseits daran, dass viele Länder diese Daten überhaupt nicht erheben, und andererseits daran, dass die Daten, die in den Ländern verfügbar sind, nicht zusammengebracht werden (Hertweck et al., 2023). Dies macht jedoch eine aussagekräftige Evaluation der Bildungspolitik unmöglich.
- 538.** Für die Verteilungspolitik sind Informationen über die Vermögensverteilung essenziell. Seit der Aussetzung der Vermögensteuer im Jahr 1997 werden jedoch keine verlässlichen **Daten zu Haushaltsvermögen** mehr erhoben. Zwar stellen das Statistische Bundesamt (Destatis) über die Einkommens- und Verbrauchsstichprobe, das Deutsche Institut für Wirtschaftsforschung (DIW) über das SOEP und die Deutsche Bundesbank über die Befragung Private Haushalte und ihre Finanzen (PHF) umfragebasierte Daten zu Vermögen in Deutschland zur Verfügung. Diese Erhebungen **weisen** jedoch einige **seit Längerem bekannte Defizite auf** (Löffler et al., 2015). Die Freiwilligkeit der Teilnahme an den Befragungen dürfte zu Selektionseffekten führen, die zu einer Verzerrung zwischen der gemessenen und der tatsächlichen Vermögensverteilung führen – insbesondere aufgrund einer mangelhaften Abdeckung von Hochvermögenshaushalten. Selbsteinschätzungen können insbesondere bei wenig liquidem Vermögen deutlich vom

tatsächlichen Vermögenswert abweichen. Zudem erschwert die geringe Stichprobengröße eine differenzierte Betrachtung der Vermögenssituation nach sozioökonomischen Charakteristika.

539. Für eine gezielte Verteilungspolitik, aber auch für die Klima- und Energiepolitik, sind derzeit **Daten zum Gebäudebestand** sehr relevant. In der öffentlichen Verwaltung werden zwar Daten zu Immobilien gepflegt, z. B. im Liegenschaftskataster oder in den Grundbüchern. Die darin enthaltenen Gebäude- und Wohnungsdaten werden jedoch bisher nicht vollständig und bundeseinheitlich erfasst. Zudem fehlt ein einheitlicher Identifikator in Form einer Gebäude- und Wohnungsnummer, wie z. B. in Norwegen, um die verschiedenen Daten zu verknüpfen (Krause et al., 2022). Die schlechte Datenlage in diesem Bereich wurde besonders in der Energiekrise auffällig, als kaum Daten zur energetischen Effizienz von Gebäuden vorhanden waren, auf deren Basis man gezielte Unterstützungsmaßnahmen hätte entwickeln und umsetzen können.
540. Neben Daten, die kontinuierlich erhoben werden sollten, gibt es **anlassbezogene Daten**, die nicht systematisch erhoben werden. Sie können **zur Evaluation diskretionärer Politikmaßnahmen** aber erforderlich werden. Beispielfähig dafür steht die Evaluation der Corona-Hilfspakete. In Frankreich wurde unmittelbar mit Umsetzung der Unterstützungsmaßnahmen im Jahr 2020 deren Evaluation, inklusive der dafür erforderlichen Datenerhebung, beschlossen. Die erste Evaluation erfolgte bereits mit dem Abschlussbericht im Sommer 2021 (JG 2021 Kasten 25). In Deutschland werden erst jetzt, mehr als drei Jahre nach dem Beginn der Corona-Pandemie, Daten zur Evaluation der Unterstützungsmaßnahmen für Unternehmen miteinander verknüpft und eine entsprechende Dateninfrastruktur aufgebaut.

2. Zeitverzug bei der Datenbereitstellung

541. Wenn keine anlassbezogenen Daten erhoben werden, könnten Politikmaßnahmen zeitnah mit Daten untersucht werden, die ohnehin regelmäßig erhoben werden. **Amtlich erhobene Mikrodaten für Forschungszwecke** stehen allerdings, wenn überhaupt, erst **mit erheblicher Verzögerung** zur Verfügung. Das liegt sowohl an langen gesetzlichen Meldefristen als auch an mangelnden Ressourcen zur Datenaufbereitung. Die Verzögerung macht **zeitnahe Auswertungen am aktuellen Rand nahezu unmöglich**. Die Auswirkungen der Corona-Krise oder der Energiekrise auf Unternehmen sowie die Wirkung der Unterstützungsmaßnahmen hätten beispielsweise mithilfe der Amtlichen Firmendaten für Deutschland (AFiD) untersucht werden können. Diese werden jedoch erst mit einem Verzug von zwei bis drei Jahren bereitgestellt – je nach Datenmodul ist das aktuellste verfügbare Jahr derzeit 2021, 2020 oder 2019. Auch beim Mikrozensus, der umfangreichsten Haushaltsbefragung in Deutschland, und der Lohn- und Einkommensteuerstatistik stammen die aktuellsten verfügbaren Daten aus den Berichtsjahren 2020 bzw. 2017.
542. **Makroökonomische Kennzahlen**, wie etwa die Volkswirtschaftlichen Gesamtrechnungen oder die Indikatoren der Konjunkturstatistik, liegen **ebenfalls**

erst mit deutlicher Verzögerung vor. Vor diesem Hintergrund sind die Bestrebungen von Destatis zu begrüßen, die Schnellschätzung des Bruttoinlandsprodukts (BIP) durch Scanner-, Satelliten-, Mobilitäts-, Finanztransaktions- und Transportdaten zu beschleunigen und die Qualität zu verbessern (Dickopf et al., 2019). Gleichwohl besteht Bedarf an einem erweiterten Angebot an hochfrequenten und zeitnah verfügbaren Konjunkturindikatoren. Diese Indikatoren hätten zwar zwangsläufig eine leicht geringere Qualität als die fertigen Datenprodukte der amtlichen Statistik, würden aber deutlich zeitnäher ein ausreichend genaues Bild von der aktuellen Lage zeichnen und somit vor allem in volatilen Krisenzeiten eine sehr hilfreiche Informationsquelle darstellen.

3. Verfügbarkeit existierender Daten eingeschränkt

543. Ein weiteres Problem ist, dass **bereits erhobene Daten teils gar nicht oder nur beschränkt für die Forschung zur Verfügung gestellt werden**. Aus diesem Grund sind beispielsweise Ländervergleiche von Bildungsergebnissen nahezu unmöglich. Die zwischen den Jahren 2000 und 2006 im Zuge der PISA-Erhebung durchgeführte Erweiterungsstichprobe, die Ländervergleiche ermöglichte, wurde zum Erhebungsjahr 2009 eingestellt. Seitdem enthält der PISA-Datensatz keine Ländervariablen, da Ländervergleiche eigentlich mit den neu eingeführten Daten des Instituts für Qualitätsentwicklung im Bildungswesen (IQB) durchgeführt werden sollten. Die qualitativ hochwertigen Daten des IQB enthalten zwar prinzipiell Identifikatoren für die einzelnen Länder. Allerdings ist deren Nutzung gesetzlich eingeschränkt. Die Verwendung der Identifikatoren unterliegt einem Gutachterprozess und Vergleiche einzelner Länder sind explizit verboten.
544. Noch problematischer ist der **erschwerter Zugang zu amtlichen Prozess- oder Registerdaten**. Zum einen fehlen in vielen Bereichen noch Register. Beispielsweise ist das Unternehmensbasisregister erst im Aufbau und ein Verwaltungsregister mit Gebäude- und Wohnungsdaten ist schlichtweg nicht existent. Ein registerbasierter Zensus ist erstmals für das Jahr 2031 geplant. Zum anderen sind **viele** der vorhandenen **Registerdaten bislang nicht oder nur eingeschränkt für die Nutzung durch die Forschung vorgesehen** (RatSWD, 2022a, 2023a). Darunter fallen etwa das Gewerbeverzeichnis, das Melderegister oder das Ausländerzentralregister.
545. Oftmals stehen bei Registerdaten nicht die vollständigen Datensätze, sondern **nur Stichproben zur Verfügung**. Dies ist beispielsweise beim Ausländerzentralregister des FDZ Bundesamt für Migration und Flüchtlinge (BAMF) oder bei der Stichprobe der integrierten Arbeitsmarktbiografien (SIAB) beim FDZ des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB) der Fall. Grund für diese Einschränkung sind oftmals Datenschutzbedenken, teils aber auch Beschränkungen der IT-Infrastruktur. Die Einschränkung auf Stichproben kann bestimmte Analysen verhindern, beispielsweise Untersuchungen für kleine Subgruppen oder Schätzungen auf Basis von Netzwerken zwischen den individuellen Beobachtungen, für die ein größtmögliches verbundenes Netzwerk notwendig ist (Bonhomme et al., 2023). Teils werden die vollständigen

Datensätze zudem nur wenigen Forscherinnen und Forschern zur Verfügung gestellt, was das Level Playing Field im Wissenschaftsprozess verzerrt.

546. Ein zentrales Problem stellen zudem Löschvorschriften, wie die **Löschvorschrift nach § 13a BStatG** dar, die abgeschafft werden sollten. [↪ ZIFFER 567](#) § 13a BStatG schreibt derzeit vor, dass Identifikatoren von Unternehmen nach 30 Jahren gelöscht werden müssen. Dies kann je nach Auslegung des Gesetzes als Pflicht zur Löschung aller auf Basis der Verknüpfungen erstellten Datensätze nach 30 Jahren ausgelegt werden. Untersuchungen auf Basis längerer Zeitreihen oder eine Anwendung neuer Methoden auf ältere Daten, die für die Gegenwart neue Erkenntnisse verspricht, werden damit unmöglich gemacht. Solche Untersuchungen werden ebenfalls erschwert, wenn analog vorhandene Daten der Vergangenheit nicht digitalisiert und der Forschung zugänglich gemacht werden. Dies ist beispielsweise bei den Daten des Ausländerzentralregisters der Fall (Altmann et al., 2023).

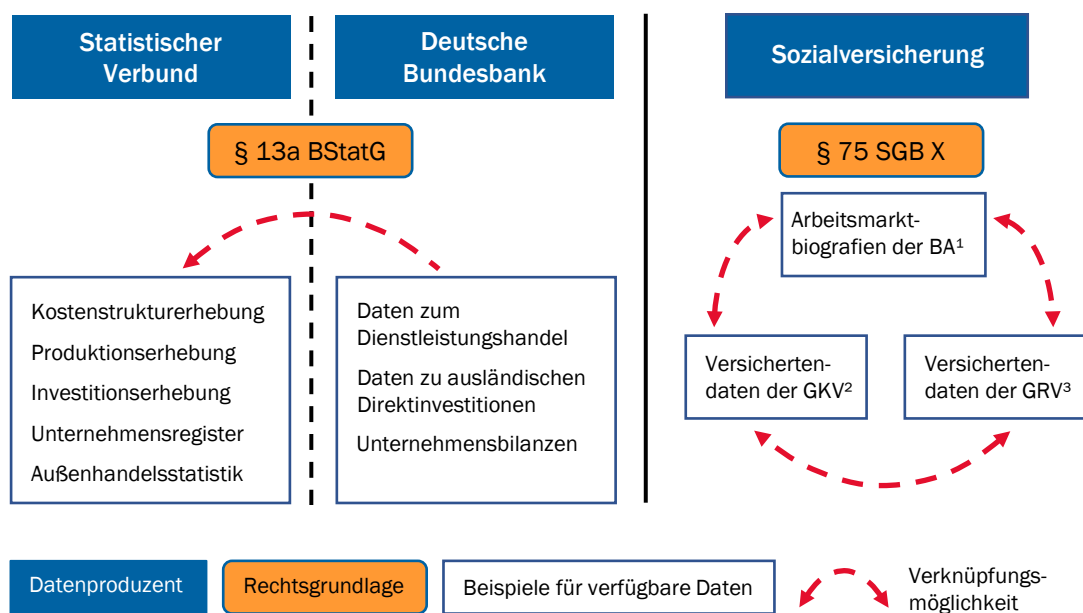
4. Datenverknüpfung bestenfalls auf Einzelfallbasis

547. Für viele wirtschaftspolitische Fragestellungen ist eine Verknüpfung von Daten aus mehreren Quellen notwendig, weil die Datenerhebung und -haltung in Deutschland sehr dezentral stattfindet. [↪ PLUSTEXT 22](#) Will man beispielsweise untersuchen, wie sich Finanzierungsbeschränkungen auf das Investitionsverhalten deutscher Unternehmen auswirken und damit Löhne und Produktivität beeinflussen, müsste man drei Datenquellen verknüpfen. So werden die meisten Arbeitsmarktdaten von der Bundesagentur für Arbeit und dem IAB registergestützt erhoben. Destatis erhebt Produktions- und Investitionsdaten und die Deutsche Bundesbank erhebt Bilanzdaten größerer Unternehmen. Die **Verknüpfungsmöglichkeiten sind jedoch sehr stark eingeschränkt, da Verknüpfungen zwischen verschiedenen Datenproduzenten nur möglich sind, wenn diese explizit gesetzlich erlaubt werden.** [↪ ABBILDUNG 144](#) Bislang fehlt eine entsprechende Erlaubnis jedoch in den meisten Fällen. So dürfen beispielsweise keine Verknüpfungen von Daten der amtlichen Statistik mit Daten aus wissenschaftlichen Unternehmensbefragungen in Deutschland durchgeführt werden (Gottschalk et al., 2023). Zusätzlich gilt bei personenbezogenen Daten in der Regel das Erfordernis der zuvor eingeholten informierten Einwilligung der Betroffenen. Verknüpfungen zwischen Daten der amtlichen Statistik und den Sozialversicherungen sind daher ebenfalls praktisch nicht möglich (Altmann et al., 2023). In der Bildungs- und Gesundheitsforschung stellt die fehlende Erlaubnis zur Datenverknüpfung ebenfalls ein erhebliches Forschungshindernis dar (Altmann et al., 2023; Hertweck et al., 2023).

Neben gesetzlichen Hürden, die durch die Reform des BStatG im Jahr 2005 nur teilweise ausgeräumt wurden (Brandt et al., 2007; RatSWD, 2023a), bestehen **institutionelle und organisatorische Herausforderungen** bei der Datenverknüpfung. Beispielsweise **fehlen einheitliche Identifikatoren zur Verknüpfung** der Datensätze zwischen den verschiedenen Datenproduzenten. Darüber hinaus sind die Regeln zur Datennutzung nicht einheitlich. Daher werden

▸ ABBILDUNG 144

Derzeitige Verknüpfungsmöglichkeiten wichtiger Datenquellen im öffentlichen Statistiksistem



1 – Bundesagentur für Arbeit. 2 – Gesetzliche Krankenversicherung. 3 – Gesetzliche Rentenversicherung.

Quelle: eigene Darstellung

© Sachverständigenrat | 23-468-01

Datensätze aktuell vorwiegend als voneinander getrennte Silos aufgebaut und genutzt.

548. **In der Vergangenheit gab es Anstrengungen**, eine standardisierte **Verknüpfung** verschiedener Datensätze **zu erproben**. Dazu gehörte unter anderem das Projekt „Kombinierte Firmendaten für Deutschland“, das von den Statistischen Ämtern des Bundes und der Länder, dem IAB, der Bundesagentur für Arbeit (BA) sowie von der Deutschen Bundesbank durchgeführt und im Jahr 2012 abgeschlossen wurde (Bender et al., 2007; Gürke et al., 2011). **Es fehlen jedoch neben den erforderlichen gesetzlichen Grundlagen breit anwendbare, standardisierte Prozesse**, sodass eine Datenverknüpfung weiterhin nur in einem sehr begrenzten Maß möglich und mit einem hohen Aufwand verbunden ist. So werden trotz der bisherigen Anstrengungen Statistiken verschiedener Institutionen **bislang ausschließlich auf Einzelfallbasis** für konkrete Forschungsvorhaben miteinander verknüpft. Das liegt an den rechtlichen Rahmenbedingungen.

Beispielsweise erlaubt **§ 75 Sozialgesetzbuch (SGB) X** zwar implizit die **Verknüpfung verschiedener Daten der Sozialversicherungen** für konkrete Forschungsvorhaben. ▸ ABBILDUNG 144 Diese müssen jedoch auf Einzelfallbasis durch die zuständige Bundesbehörde genehmigt werden. Auf Basis von **§ 13a BStatG** können zudem **Daten der Deutschen Bundesbank** an den Statistischen Verbund übermittelt werden und von diesem **mit Daten der Bundesstatistik verknüpft** werden. ▸ ABBILDUNG 144 Von dieser Möglichkeit wird allerdings nur im Rahmen spezifischer Forschungsprojekte Gebrauch gemacht (Kruse et al., 2023). Zumindest werden die über diesen Weg verknüpften Daten aber über das FDZ des Statistischen Bundesamts der Wissenschaft zur Verfügung gestellt.

5. Datenzugang nicht benutzerfreundlich

549. Nicht zuletzt ist der **Zugang zu den prinzipiell verfügbaren Daten aufwendig und komplex**. So ist die Beantragung des Datenzugangs zeitaufwendig, während der Datenzugriff kostspielig und wenig benutzerfreundlich ausgestaltet ist. Insbesondere die Einschränkung des Zugangswegs auf Gastwissenschaftlerarbeitsplätze und die kontrollierte Datenfernverarbeitung, die für viele Datensätze der amtlichen Statistik, aber auch anderer FDZ gilt, ist problematisch. Eine kontrollierte Datenfernverarbeitung sieht z. B. eine Vorbereitung der Programmroutinen durch Forschende vor. Diese Routinen werden an die Forschungsdatenzentren versandt, dort geprüft und ausgeführt, wonach die Überprüfung der Ergebnisse und schließlich die Freigabe der Ergebnisse an die Forschenden erfolgt. Dies ist ein langwieriger Prozess, bei dem aufgrund technischer Restriktionen zusätzlich Beschränkungen für die Anzahl der Variablen vorliegen. So können etwa beim Taxpayer-Panel maximal 25 Variablen von über 1 000 verfügbaren Variablen auf einmal verwendet werden (Wissenschaftlicher Beirat beim BMF, 2021). Diese **umständlichen Zugangswege führen zu höheren Kosten, Engpässen bei der physischen Infrastruktur und zu einer geringeren Nutzerfreundlichkeit** als ein vollständiger Datenfernzugriff durch einen gesicherten Onlinezugang (RatSWD, 2019, 2022b).
550. Zu der **geringen Nutzerfreundlichkeit des Datenzugangs** tragen wesentlich **datenschutzrechtliche Einschränkungen** bei. Die datenschutzrechtlichen Grundlagen werden in den einzelnen Ländern unterschiedlich ausgelegt, was den Zugang zu Registerdaten für Forschungszwecke erschwert (RatSWD, 2023a). Dabei wird das Forschungsprivileg nach Art. 89 DSGVO (Datenschutz-Grundverordnung) bislang sehr zurückhaltend interpretiert und die wissenschaftliche Forschung in der Güterabwägung gegenüber Datenschutzinteressen als nachrangig behandelt. So führt beispielsweise eine unterschiedliche Interpretation der formalen Anonymität der Daten dazu, dass die bayerischen Unternehmensdaten nicht einmal bei einem physischen Aufenthalt im Forschungsdatenzentrum (On-Site-Access) direkt durch die Forscherinnen und Forscher einsehbar sind. Stattdessen wird hierfür nur die kontrollierte Datenfernverarbeitung mit all ihren Einschränkungen angeboten. [↪ ZIFFER 549](#)
551. Darüber hinaus liegen häufig nur **begrenzt Informationen über die Struktur der einzelnen Datensätze oder hinreichend guter Metadaten und Datensatzdokumentationen** vor, sodass die Nutzerinnen und Nutzer viel Zeit mit der Datenerkundung und -bearbeitung verbringen müssen. Dies ist insbesondere bei einem fehlenden Fernzugang kostspielig. Beispielsweise werden Datenstrukturfiles nicht für alle Datensätze verfügbar gemacht. Solche Datenstrukturfiles würden den Forschenden eine bessere Vorbereitung auf die Datenarbeit ermöglichen, weil sie zufällig generierte Daten enthalten, die die Struktur des echten Datensatzes widerspiegeln. Dadurch können schon im Vorfeld Analysedateien vorbereitet werden und die Zeit vor Ort im Forschungsdatenzentrum deutlich effizienter genutzt werden. Dies entlastet die physische Infrastruktur vor Ort.
552. **Auch für weniger sensible Daten, die als aggregierte oder mikroaggregierte Datensätze zur Verfügung stehen, ist der Zugang in Deutschland**

wenig benutzerfreundlich. So stellt der von Destatis angebotene Standardzugang GENESIS-ONLINE insbesondere für Zeitreihendaten eine deutlich weniger benutzerfreundliche Oberfläche bereit als andere Datenanbieter (Bachmann et al., 2023a). Zeitreihendaten, selbst wenn sie originär von Destatis stammen, lassen sich deutlich einfacher über FRED, das Datenportal der St. Louis Federal Reserve, oder die Zeitreihendatenbank der Deutschen Bundesbank abrufen und visualisieren.

III. HANDLUNGSOPTIONEN: EINE DATENINFRASTRUKTUR FÜR DAS 21. JAHRHUNDERT

1. Rahmenbedingungen des Statistischen Verbunds verbessern

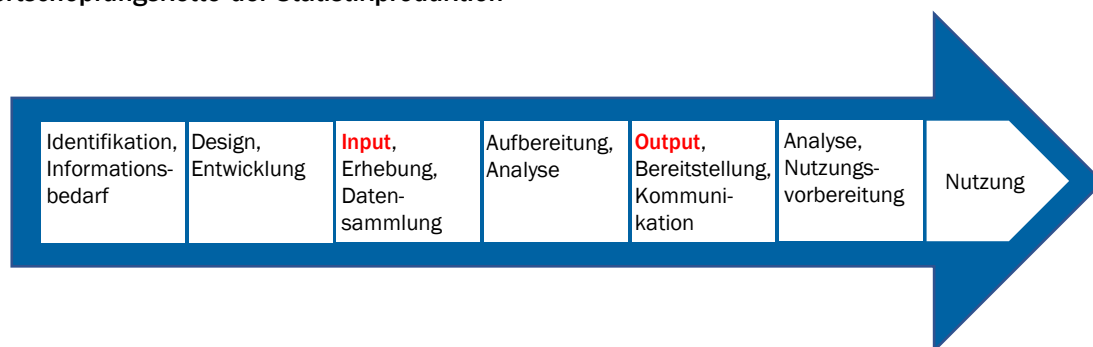
553. Der seit der Jahrtausendwende angestrebte Paradigmenwechsel der amtlichen Statistik von der Inputorientierung zur Outputorientierung (Hahlen, 2009) sollte endlich konkretisiert und umgesetzt werden. Dies bedeutet eine Fokussierung auf Informationsbedarfe und Dienstleistungen, die bereitgestellt werden sollen, statt auf Datenquellen, die genutzt werden dürfen. Der **gesetzliche Rahmen der amtlichen Statistik** ist allerdings **noch immer inputorientiert**, fokussiert also auf die genaue Definition der Daten, die die amtliche Statistik erheben darf. [↘ ABBILDUNG 145](#) Bislang müssen gemäß § 9 BStatG für amtliche Statistiken Rechtsvorschriften – oft fachstatistische Einzelgesetze – beschlossen werden, die detailliert definieren, welche Merkmale von welchen Subjekten erhoben werden dürfen. Das bedeutet allerdings, dass zur Anpassung von Statistiken und Erhebungen sowie für neue Nutzungswege oftmals gesetzliche Anpassungen notwendig sind. Dies reduziert die Anpassungsfähigkeit der amtlichen Statistik erheblich. [↘ ZIF-FER 562](#)

Aus diesem Grund sollte die **Statistikgesetzgebung**, insbesondere § 9 BStatG, **hin zu einer Outputorientierung angepasst werden**. Eine solche Anpassung würde den Statistischen Verbund in die Lage versetzen, die im Gesetz allgemein definierten Aufgaben durch passende Erhebungen zu erfüllen, ohne diese über fachstatistische Einzelgesetze detailliert legitimieren zu müssen. So sollte der Statistische Verbund befugt sein, die notwendigen Erhebungen, inklusive Merkmalen und Stichproben, sowie den notwendigen Zugriff auf Register- und Verwaltungsdaten selbst zu definieren, die zur Erfüllung eines per Gesetz definierten Informationsbedarfs notwendig sind. Dies würde die Flexibilität und Reaktionsgeschwindigkeit des Statistischen Verbunds deutlich erhöhen.

554. Die **Verbesserung der Dateninfrastruktur erfordert zusätzliche finanzielle und personelle Ressourcen**. Die benötigten Ressourcen für zusätzliche Dienstleistungen und insbesondere für zusätzliche Datenprodukte können

↳ ABBILDUNG 145

Wertschöpfungskette der Statistikproduktion



Quelle: KomZS (2023)
© Sachverständigenrat | 23-470-01

reduziert werden, wenn bereits bestehende Produkte effizienter ausgestaltet werden. Viele der folgenden Vorschläge setzen daher an einer Anpassung und Verbesserung existierender Produkte oder der Nutzung bereits bestehender, aber noch nicht genutzter Datenquellen sowie der Umstellung von Erhebungen auf registergestützte Methoden an. Einige Produkte und Dienstleistungen werden auf diese Weise jedoch nicht bereitzustellen und mit deutlichem Mehraufwand verbunden sein. Insbesondere die Verbesserung der digitalen Dateninfrastruktur sowie die Erschließung und Nutzung administrativer Daten und Register werden Mehraufwand generieren.

Daher sollte das **Budget** der entsprechenden Institutionen, insbesondere der amtlichen Statistik, ausgeweitet werden. Dies kann dennoch kostensparend für den Staat sein, da die zusätzlichen Mittel als finanzielle **Investition in eine bessere Entscheidungsgrundlage für Politikmaßnahmen** zu sehen sind, die durch eine zielgenauere Ausgestaltung und fortlaufende Evaluation die **Kosten von Politikmaßnahmen reduzieren kann**. Setzt man beispielsweise die zusätzlichen Kosten großzügig mit 50 % des durchschnittlichen Budgets von Destatis in den Jahren 2021 bis 2023 von 260 Mio Euro jährlich an, würde sich ein jährlicher Mehraufwand von 130 Mio Euro ergeben. Würde man beispielsweise annehmen, dass bei Unternehmenshilfen, wie im Fall der Coronahilfen, eine Reduktion der Mitnahmeeffekte im Umfang von 5 % der Gesamtkosten möglich wäre, wären alleine die Kosten für die Corona-Hilfsmaßnahmen für Unternehmen, deren Auszahlungsvolumen bislang 46 Mrd Euro betrug, um 2,3 Mrd Euro geringer ausgefallen. Damit hätten die oben angesetzten Zusatzkosten für knapp 18 Jahre getragen werden können.

- 555. Um das Dienstleistungsangebot für die Wissenschaft zu verbessern, ist eine **gesetzliche Verankerung eines Forschungsauftrags für Destatis** im BStatG sinnvoll. Die Übertragung eines Forschungsauftrags an Destatis könnte sich dabei am Beispiel von Forschungsaufträgen an das US Census Bureau (gemäß Department of Commerce DOO 35-2B), der Bundesagentur für Arbeit (gemäß § 280 SGB III) oder der Deutschen Bundesbank (implizit abgeleitet aus dem geldpolitischen Auftrag und dem Finanzstabilitätsauftrag nach § 3 BBankG (Gesetz über die Deutsche Bundesbank) und dem Erfordernis einer soliden Forschungsgrundlage zur Erfüllung des Auftrags) orientieren. Diese betreiben mit dem

Center for Economic Studies, dem IAB und dem Forschungszentrum der Deutschen Bundesbank eigene Forschungszentren.

Ein **in § 3 BStatG verankerter gesetzlicher Forschungsauftrag für Destatis** und die **Einrichtung eines Forschungszentrums bei Destatis** hätten mindestens drei Vorteile. Erstens würden Forscherinnen und Forscher direkt bei Destatis mit den Forschungsdaten an konkreten Forschungsfragen arbeiten und könnten so unmittelbar Lösungen für die unweigerlich in der Praxis auftretenden Datenprobleme entwickeln. Dies würde zu einer schnellen und kontinuierlichen Verbesserung der verfügbaren Forschungsdaten beitragen. Zweitens würde das Eigeninteresse von Destatis an der Aufbereitung, Bereitstellung und Nutzung von Forschungsdaten steigen, wenn es auch hinsichtlich des Forschungsauftrags evaluiert werden würde. Dies hätte positive Effekte für die verfügbaren Forschungsdaten. Drittens ist anzunehmen, dass die Forscherinnen und Forscher des Forschungszentrums zur qualitativ hochwertigen wissenschaftlichen Forschung mit deutschen Daten und somit zum besseren Verständnis der deutschen Volkswirtschaft beitragen würden. [↘ KASTEN 30](#)

556. Darüber hinaus sollten **die FDZ des Bundes und der Länder als Kernaufgabe** des Statistischen Verbunds **im BStatG gesetzlich verankert** werden. Bisher werden die FDZ im BStatG nicht explizit erwähnt. Die Bereitstellung der Daten an externe Forschende wird lediglich indirekt über das BStatG geregelt. Mit einer gesetzlichen Verankerung würde die Bedeutung der FDZ bei Destatis und den Statistischen Landesämtern hervorgehoben und explizit darauf hingewiesen, dass für die Datenbereitstellung eine entsprechende Einrichtung (FDZ) und entsprechende personelle und finanzielle Ressourcen notwendig sind. Damit würde erstens der langfristige Betrieb der FDZ gesichert. Zweitens würde das aktuell bestehende Problem der Selbstfinanzierung der FDZ durch die Datennutzerinnen und -nutzer behoben, das zu vergleichsweise hohen Kosten des Datenzugangs führt. Eine öffentliche Finanzierung wäre insofern gerechtfertigt, als dass die Forschungsergebnisse nicht nur den Datennutzerinnen und -nutzern selbst, sondern der gesamten Gesellschaft zugutekommen. Es entstehen also positive externe Effekte. Eine direkte Finanzierung würde dazu beitragen, diese externen Effekte zu internalisieren.

2. Datenlücken schließen

557. Der Mangel an vergleichenden Analysen der Bildungssysteme verschiedener Länder könnte über eine **bundesweite Schülerindividualstatistik** adressiert werden (Hertweck et al., 2023). Sie sollte die Individualdaten **zur gesamten Schullaufbahn von Schülerinnen und Schülern als Panel** bereitstellen. Verwaltungsdaten sollten um Kompetenzmessungen ergänzt und über Schulübergänge hinweg verknüpft werden. Solch eine Statistik könnte den **Kern eines Bildungsverlaufsregisters** vom Kindergarten über die Schule bis zur Hochschule bilden. Um die Effektivität der verschiedenen Bildungssysteme und die Qualität verschiedener Schulen vergleichen zu können, ist es essenziell, regelmäßige und flächendeckende Kompetenzmessungen zu etablieren. Deren Ergebnisse könnten im Rahmen der Schülerindividualstatistik für die Forschung zur Verfügung gestellt werden.

558. Das Problem fehlender **amtlicher Daten zur Vermögensverteilung** könnte vergleichsweise kostengünstig durch eine **Verknüpfung des Mikrozensus mit administrativen Daten** und die Entwicklung geeigneter Schätzverfahren zur Abschätzung des Vermögens anhand von administrativen Daten adressiert werden (Bachmann et al., 2023b). Die Verknüpfung mit Lohn- und Einkommenssteuerdaten könnte die Auskunftgebenden entlasten, die Qualität der Angaben zum Einkommen erhöhen und die Abschätzung von Finanz- und Immobilienvermögen ermöglichen. Dadurch würde neben den fehlenden Vermögensdaten auch das Problem der verzerrten Darstellung der Einkommenssituation adressiert, das bei Daten aus Eigenangaben entsteht (Emmenegger und Münnich, 2023). Betriebsvermögen können wiederum auf Basis der Daten zu Unternehmensabschlüssen der Deutschen Bundesbank ermittelt werden. Daten aus einem künftigen Gebäude- und Wohnungsregister könnten zur Ermittlung von Immobilienvermögen herangezogen werden. Um dies zu ermöglichen, sollte schnellstmöglich darauf hingewirkt werden, datensatzübergreifende Identifikatoren festzulegen. Dazu dürfte die Steuer-ID geeignet sein. Eine Auskunftspflicht über direkte Unternehmensbeteiligungen der Haushalte und eine Nennung der Handelsregisternummer dürften für eine Verknüpfung mit Unternehmensbilanzdaten geeignet sein. Zudem muss die Verknüpfung der Befragungsdaten des Mikrozensus mit anderen administrativen Daten gesetzlich erlaubt werden.
559. Die schlechte Informationslage zum Gebäudebestand in Deutschland könnte sich mit der geplanten Novelle der EU-Gebäuderichtlinie deutlich verbessern (Europäisches Parlament, 2023). Diese Novelle sieht die Einrichtung nationaler Datenbanken für Energieausweisdaten sowie die Einführung von Gebäudelogbüchern vor, die die Grundlage für ein Gebäuderegister bilden könnten. Ein **Gebäuderegister wäre sinnvoll und die darin enthaltenen Informationen sollten sorgfältig ausgewählt werden, da sie die Datengrundlage für Forschung im Bereich der Energie- und Sozialpolitik bilden** können. So sind etwa die Informationen zu Mietwohnungen, die mehr als die Hälfte des Wohnungsbestands in Deutschland ausmachen (Krause et al., 2022), bisher kaum erfasst, aber hochrelevant etwa für die Sozialpolitik. Zentral sind zudem Daten zu bestehenden Heizanlagen, in denen ersichtlich ist, welche Energieträger zur Wärmeerzeugung verwendet werden.
560. Aus makroökonomischer und wirtschaftspolitischer Sicht besteht darüber hinaus Bedarf an einer **hochfrequenten Haushaltsbefragung**, die nahezu in Echtzeit die ökonomische Lage (Arbeitsmarkt, Konsum, Erwartungen, Einkommen, Vermögen, Verschuldung etc.) der deutschen Haushalte abbildet (Bachmann et al., 2021). Gerade während der Corona-Pandemie war diese Lage der Politik nur in groben Zügen bekannt und wirtschaftspolitische Maßnahmen deshalb oft nicht genau dosierbar. Ein Schritt in Richtung einer solchen Befragung könnte etwa eine systematische Umstellung des SOEP auf eine rollierende (vierteljährliche) Erhebungsfrequenz sein. Ein ähnliches System besteht bei der rollierenden Erhebung des CPS in den USA. Der Mikrozensus, als größte Haushaltsbefragung in Deutschland, enthält gleichmäßig auf das Jahr verteilte Berichtswochen. Bislang werden die Informationen zu den Berichtswochen im Mikrozensus kaum genutzt. Der Rechtsrahmen für den Mikrozensus sollte zudem so gestaltet werden, dass kurzfristig und anlassbezogen neue Umfragemodule einbezogen werden können.

Darüber hinaus wäre eine Haushaltsbefragung in noch höherer Frequenz (wöchentlich, zweiwöchentlich), die in Krisenzeiten aktiviert werden könnte, hilfreich. Analog könnte eine hochfrequente Unternehmensbefragung nach Vorbild des Business Pulse Survey des US Census Bureau oder des Bundesbank Online Panel - Firmen (BOP-F) der Deutschen Bundesbank konzipiert werden.

561. Bei neuen Gesetzgebungsvorhaben sollte künftig ab einem zu definierenden Volumen der dafür erforderlichen Mittel eine Pflicht zur begleitenden Datenerhebung, Datenbereitstellung und zeitnahen wissenschaftlichen Evaluation verankert werden. Beispielsweise könnte eine **wissenschaftliche Begleitung und Evaluierung inklusive Datenerhebung für Maßnahmen mit einem Volumen von mehr als 100 Mio Euro jährlich** vorgeschrieben werden. Bisher ist bei Regelungsvorhaben lediglich die Pflicht einer internen Evaluierung ab einem Erfüllungsaufwand von mehr als 5 Mio Euro verankert (Bundesregierung, 2013, 2019). Zudem sollten Forschungsklauseln, also forschungsfreundliche, barrierefreie Zugangsregeln für unabhängige wissenschaftliche Forschung geschaffen werden. Idealerweise sollte bei Maßnahmen, die Unternehmen oder Haushalte betreffen, der Identifikator der Betroffenen ermittelt werden. So ließen sich die Daten mit Informationen aus amtlichen Prozess- oder Registerdaten verknüpfen.

3. Geringerer Zeitverzug und höhere Anpassungsfähigkeit

562. Einige der zur Neuerhebung vorgeschlagenen Datenprodukte könnten im Zusammenspiel mit einer outputorientierten Statistikgesetzgebung [↘ ZIFFER 553](#) dazu beitragen, neben der Verbesserung der Menge an verfügbaren Daten auch die **Schnelligkeit und die Reaktionsfähigkeit der amtlichen Statistik zu erhöhen**. So könnte das Erhebungsprogramm einer hochfrequenten Haushaltsbefragung [↘ ZIFFER 560](#) kurzfristig angepasst werden, um nützliche Informationen in Krisensituationen bereitzustellen. Dazu wurden beispielsweise in den USA im Zuge der Corona-Pandemie die Fragen des monatlichen CPS kurzfristig angepasst, um die Auswirkungen der Corona-Pandemie zu messen. [↘ ZIFFER 531](#) Analog könnte eine Verpflichtung zur Datenerhebung und zu einer von Wissenschaftlerinnen und Wissenschaftlern konzipierten Evaluation bei Politikmaßnahmen ab einem bestimmten Maßnahmenvolumen dafür sorgen, dass Daten zur Evaluation dieser Maßnahmen zeitnah zur Verfügung stehen.
563. Um den **Zeitverzug** bei der Bereitstellung von **Forschungsdaten und** der Produktion **öffentlicher Datenprodukte der amtlichen Statistik zu verringern**, müsste bereits eine zeitnähere Datenerhebung erfolgen. Dazu kann in sehr vielen Fällen der stärkere Rückgriff auf administrative Prozessdaten dienen, die mit einheitlichen Identifikatoren miteinander verknüpfbar gemacht werden sollten. Insbesondere bestehen hier Potenziale, durch eine zunehmende Digitalisierung von Verwaltungsprozessen mehr administrative Prozessdaten verfügbar zu machen. Im Bereich der Gesundheitsdaten könnten beispielsweise Daten aus der elektronischen Patientenakte genutzt werden. Eine weitere Überlegung wäre es, für ausgewählte Statistiken Meldefristen zu verkürzen und die Datenübermittlung

zu beschleunigen. Dabei müssten allerdings zusätzliche Belastungen für die meldenden Stellen, beispielsweise Unternehmen, im Blick behalten werden. Um Erhebungsdaten schneller als amtliche Statistiken zu Verfügung zu stellen, ist eine zeitnähere Aufbereitung der Daten, insbesondere im Fall des Mikrozensus, notwendig. Gleiches gilt für die zeitnähere Verfügbarkeit von Forschungsdaten, für die die Aufbereitung der Daten für die Wissenschaft beschleunigt werden müsste. Für beides werden zusätzliche finanzielle und personelle Ressourcen benötigt.

4. Verfügbarkeit bereits erhobener Daten verbessern

564. Prinzipiell sollten alle Datenquellen, die öffentlich finanziert wurden, der unabhängigen wissenschaftlichen Forschung zur Verfügung stehen. Für die Wirtschaftsforschung sind personenbezogene und weniger aggregierte Daten besonders wichtig, weil mit ihnen Wirkungskanäle von wirtschaftspolitischen Maßnahmen und deren heterogene Auswirkungen analysiert werden können (RatSWD, 2023a). Deshalb sollte eine **Erweiterung der für die Forschung verfügbaren administrativen Mikrodaten priorisiert werden**. Die Bereitstellung solcher Daten dürfte kosteneffizienter sein als die Erhebung neuer Daten und ist insbesondere für Unternehmen und Individuen nicht mit zusätzlichem Erhebungsaufwand verbunden. Der Aufwand kann weiter gesenkt werden, wenn die **Datenerhebung der Datenproduzenten harmonisiert** wird, etwa durch die Verwendung von gemeinsamen Datenportalen. Die Harmonisierung könnte mit dem **Once-Only-Prinzip** verbunden werden, um das mehrmalige Eintragen der gleichen Daten zu vermeiden.
565. Aufgrund ihrer umfassenden Abdeckung und der hohen Qualität der Datenerhebung sind **administrative Daten für die empirische Forschung** besonders wertvoll (Currie et al., 2020), da sie besonders zuverlässig sind. Bereiche mit guter Verfügbarkeit solcher Daten bieten Wissenschaftlerinnen und Wissenschaftlern ideale Bedingungen, um diese Themen besonders tief und gründlich zu erforschen (Nagaraj und Tranchero, 2023). Dies nutzen internationale Forscherinnen und Forscher beispielsweise bei den häufig untersuchten Daten zum deutschen Arbeitsmarkt. ↘ **KASTEN 30** Diese qualitativ hochwertige empirische Forschung wiederum liefert **gut belegte Erkenntnisse, auf deren Grundlage die Politik schnelle und informierte Entscheidungen treffen kann**.

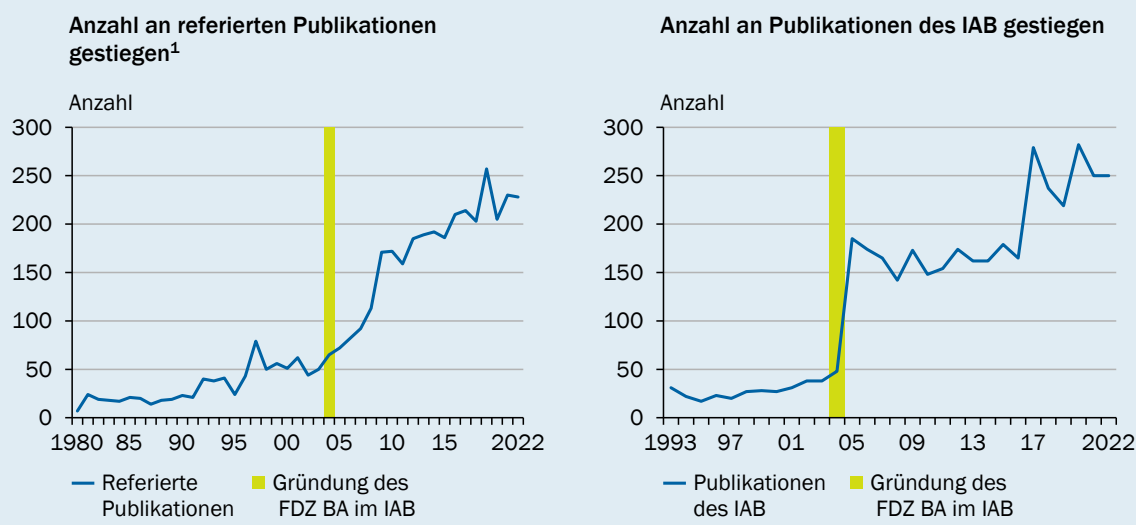
▸ KASTEN 30

Hintergrund: Forschungsdatenzugang und Forschung des IAB – eine Erfolgsgeschichte

Das **Institut für Arbeitsmarkt- und Berufsforschung (IAB)** ist eine seit dem Jahr 1967 tätige Dienststelle der Bundesagentur für Arbeit (BA). Das IAB **wurde mit der Arbeitsmarkt- und Berufsforschung beauftragt** und verfügt dafür über die aus dem Geschäftsbereich der Bundesagentur und aus anderen statistischen Quellen stammenden administrativen Daten. Das IAB hat umfangreiche Beiträge zur Arbeitsmarkt- und Berufsforschung geleistet, die sich in einer Vielzahl eigener Publikationen widerspiegeln. [▸ ABBILDUNG 146 RECHTS](#)

▸ ABBILDUNG 146

Publikationen zum deutschen Arbeitsmarkt



1 – Jährliche Anzahl der Publikationen in wissenschaftlichen Zeitschriften, die beim Portal Web of Science gezeigt werden, wenn folgende Kriterien eingegeben werden: „Germany“ und („employment“ oder „unemployment“ oder „wage“) im Feld „Thema“ sowie „Volkswirtschaftslehre und Business“ im Feld „Bereich“

Quellen: IAB, Web of Science

© Sachverständigenrat | 23-465-01

Daten zur Abbildung

Neben der Forschung mit Prozessdaten der BA erhebt das IAB auch eigene Daten. Die **Datenauswertungen wurden ursprünglich innerhalb des IAB durchgeführt** und arbeitsmarktrelevante Mikrodaten wurden nur selten mit externen Akteuren geteilt. Dies führte zu einer Unterlastung der Datenressourcen und zu sehr begrenzten Kooperationsprojekten mit nationalen und internationalen akademischen Wissenschaftlerinnen und Wissenschaftlern. Die daraus resultierenden Analysen waren meist deskriptiv und wurden lediglich in wenigen Fällen in referierten Zeitschriften veröffentlicht (Müller und Möller, 2019). Mit der Zeit jedoch stieg die Nachfrage nach empirischer Evidenz zu Arbeitsmärkten deutlich, insbesondere durch die Hartz-Reformen. Darüber hinaus hat die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001) Vorschläge zur effizienteren Verwendung statistischer Erhebungen gemacht, darunter einen zur Einrichtung von FDZ bei jedem staatlichen Datenproduzenten, z. B. auch beim IAB.

Das **FDZ BA im IAB** wurde **im Jahr 2004 gegründet** und **stellt seitdem externen Wissenschaftlerinnen und Wissenschaftlern die im IAB vorhandenen Daten in Form von standardisierten Datenprodukten zur Verfügung**. Seitdem ist die Anzahl der Forschenden, die mit den IAB-Mikrodaten den deutschen Arbeitsmarkt untersuchen, rasant gestiegen. Beispielsweise arbeiten derzeit rund 1 400 Forschende an mehr als 650 Forschungsprojekten. Dies resultierte in einer hohen Anzahl an referierten Publikationen zum deutschen Arbeitsmarkt. [▸ ABBILDUNG 146](#)

LINKS Durch die Schaffung von Zugangspunkten im europäischen Ausland und in den USA ab dem Jahr 2012 ist die Anzahl ausländischer Forscherinnen und Forscher, die die Daten des IAB nutzen, deutlich gestiegen (Müller und Möller, 2019).

566. Darüber hinaus wäre eine **Aufbereitung historischer makroökonomischer Daten wertvoll**. So ist der Zugang zu manchen Daten der Volkswirtschaftlichen Gesamtrechnungen erst ab dem Jahr 1991 benutzerfreundlich verfügbar, obwohl ihre Erhebung durch Destatis zum Teil viel weiter zurückliegt. Eine Bereitstellung der Daten ab den 1950er-Jahren, wie in den USA, wäre für ein verbessertes Verständnis von makroökonomischen Zusammenhängen wichtig. Zudem wären beispielsweise für die makroökonomische Investitions- und Kapitalstockforschung langlaufende Investitions- und Kapitalstockreihen hilfreich, die teils bis in das 19. oder sogar 18. Jahrhundert zurückgehen. Diese Daten könnten beispielsweise im Rahmen des Programms „EXSTAT - Experimentelle Statistiken“ von Destatis zur Verfügung gestellt werden.
567. Die Verfügbarkeit in der Vergangenheit erhobener Daten wird durch Löschvorschriften wie der **Löschvorschrift im Bundesstatistikgesetz (§ 13a BStatG)** erschwert. [↘ ZIFFER 546](#) Dies ist für die empirische Wirtschaftsforschung und die Politikberatung problematisch, da langlaufende Zeitreihen für die Identifikation von Wirkungsmechanismen erforderlich sind. Löschvorschriften für Daten sollten aufgehoben werden und insbesondere die in § 13a BStatG verankerte Löschvorschrift **abgeschafft werden**. Das Forschungsprivileg der DSGVO sollte, wie vom RatSWD (2023b) vorgeschlagen, genutzt werden, um zeitnah langfristige Speicherdauern für Forschungsdaten festzulegen.

5. Bessere Verknüpfbarkeit über verschiedene Datenanbieter hinweg

568. **Zur Verknüpfung von Mikrodaten über die bestehenden Datensilos der einzelnen Datenanbieter hinweg sollten Datentreuhänder eingerichtet werden**. Datentreuhänder sind Datenintermediäre, die Daten aus verschiedenen Quellen zusammenführen und sicher sowie gesetzeskonform den berechtigten Nutzerinnen und Nutzern zur Verfügung stellen. Dies betrifft insbesondere Daten der öffentlichen Verwaltung, der amtlichen Statistik und von öffentlich mitfinanzierten Forschungseinrichtungen. Daten aus der Wirtschaft und von Unternehmen sollten ebenfalls integriert werden. Die Verknüpfung von Daten aus verschiedenen Quellen kann durch Datentreuhänder ermöglicht werden. **Idealerweise erfolgt dies durch eine Erweiterung des gesetzlichen Auftrags von Destatis, beispielsweise durch die Gründung eines innerhalb Destatis angesiedelten Mikrodatenzentrums nach Vorbild des AMDC in Österreich.** [↘ PLUS-TEXT 24](#) Dabei sollte sichergestellt werden, dass ausreichend Mittel vorhanden sind, um Forscherinnen und Forschern einen kostengünstigen Datenzugang zu ermöglichen, und dass die Verwaltung und die registerführenden Stellen kooperieren. Das von der Bundesregierung gegründete Dateninstitut (BMWK und BMI, 2023)

ist anders als ursprünglich angedacht nicht darauf ausgerichtet, die Aufgabe eines Datentreuhänders für die Forschung zu übernehmen.



PLUSTEXT 24

Hintergrund: Austrian Micro Data Center als Vorbild für Deutschland

In Österreich können alle Forschungseinrichtungen, die die Zugangsvoraussetzungen gemäß Forschungsorganisationsgesetz erfüllen, über das **Austrian Micro Data Center (AMDC)** einen Antrag auf Nutzung und Verknüpfung administrativer Daten stellen (Fuchs et al., 2023). Das Datenangebot ist sehr umfangreich: Neben allen von Statistik Austria verwendeten Mikro- und Registerdaten sind zahlreiche weitere Verwaltungsdaten des Bundes verfügbar (Altmann et al., 2023). Darüber hinaus können die Forscherinnen und Forscher eine Verknüpfung administrativer Daten mit eigenen Daten beantragen. Für die genehmigten Projekte bestehen verschiedene Möglichkeiten des Datenzugriffs, darunter ein gesicherter Remote-Access. Das AMDC fungiert dabei als eine **zentrale Informationsstelle** nach Art. 8 des Data Governance Act sowie als **Datentreuhänder**, der insbesondere die Verknüpfung der verschiedenen Datensätze ermöglicht. Die Gründung des AMDC im Juli 2022 wurde durch zahlreiche Gesetzesänderungen vorbereitet, die die **Aufgaben und Rechte des AMDC gesetzlich verankert** haben (Altmann et al., 2023).

569. Bis die Datentreuhänderstrategie vollständig umgesetzt wird, bleibt der **Abbau der rechtlichen Hindernisse bei der Verknüpfung von Registern** mit anderen Daten aus der Verwaltung oder der Wissenschaft eine wichtige Aufgabe. Dabei ist die **Erweiterung des § 13a BStatG besonders relevant**, da hier die Grenzen für die Zusammenführung von amtlichen Daten definiert sind. So können nach der aktuellen Ausgestaltung nur die Wirtschaftsstatistiken der Deutschen Bundesbank an Destatis übermittelt werden, aber nicht z. B. die Daten der Deutschen Rentenversicherung. In das Forschungsdatengesetz sollte explizit die Möglichkeit der Datenverknüpfung für die Forschung aufgenommen werden und die Verknüpfbarkeit auf weitere Datenproduzenten neben der Deutschen Bundesbank und Destatis ausgeweitet werden. Dazu müsste wie vom RatSWD (2023b) angemerkt eine **Reihe von Einzelgesetzen geändert werden**. In **§ 7a** des Gesetzes über Steuerstatistiken (**StStatG**) müsste die Zusammenführung von Einzeldaten der Steuerstatistik mit Daten anderer öffentlicher Datenproduzenten erlaubt werden. Im Verwaltungsdatenverwendungsgesetz (**VwDVG**) und im Statistikregistergesetz (**StatRegG**) müsste ebenfalls der Grundsatz verankert werden, dass alle Datenproduzenten der öffentlichen Statistik ihre Daten für wissenschaftliche Forschungszwecke untereinander tauschen und miteinander verknüpfen dürfen, zumindest über einen Datentreuhänder.
570. Um den Aufwand bei der Datenverknüpfung, sowohl innerhalb der amtlichen Statistiken als auch mit externen Datensätzen, zu verringern, bedarf es einer möglichst schnellen Einführung **eindeutiger Identifikatoren über Datensätze hinweg**. Bei Unternehmen könnten beispielsweise die Handelsregisternummer oder die Umsatzsteuernummer verwendet werden (Gottschalk et al., 2023) und für die Betriebs-ID die vom Betriebsnummern-Service (BNS) erfasste Betriebsnummer dienen. Diese Vereinheitlichung wird besonders zukünftig relevant, wenn die Verknüpfungsmöglichkeiten erweitert und neue Register, etwa

Bildungsverlaufsregister, aufgebaut werden (RatSWD, 2022a). Aber auch für bereits zugelassene Datenverknüpfungen erhöht die mangelnde Konsistenz der Identifikatoren die Bearbeitungskosten (Gottschalk et al., 2023).

571. Da die verknüpften Forschungsdaten Wissenschaftlerinnen und Wissenschaftlern, aber nicht den politischen Entscheidungsträgern oder der breiten Öffentlichkeit, zur Verfügung stehen, dürfte die **soziale Akzeptanz für Datenverknüpfungen** höher sein als bei der Verknüpfung für Verwaltungszwecke. Wenn die Bevölkerung über den **gesellschaftlichen Nutzen der ermöglichten Forschung** informiert wird, dürfte die soziale Akzeptanz für Datenverknüpfungen steigen.

6. Einfacher und benutzerfreundlicher Datenzugang

572. Zur Vereinfachung des Datenzugangs für die Wissenschaft sollte ein **Wissenschafts- oder Forschungsprivileg** als zentraler Bestandteil des geplanten Forschungsdatengesetzes verankert werden. Dieses **sollte bestehende Einzelregelungen zum Datenzugang, -verwendung und -schutz ersetzen** und bundeseinheitlich regeln. Insbesondere sollte das Forschungsprivileg der wissenschaftlichen Forschung gegenüber dem Datenschutz ein hohes Gewicht bei der Datennutzung sicherstellen. Zudem sollten damit großzügige Verknüpfungsregeln für Daten erreicht werden. ↘ ZIFFER 569 Da die wissenschaftliche Forschung maßgeblich zum Gemeinwohl beiträgt, sollte im Bereich des Datenzugangs, der -verknüpfung und -verwendung für die Wissenschaft alles erlaubt sein, was nicht explizit und begründet verboten ist. Ein Forschungsdatengesetz würde maßgeblich zur Rechtssicherheit beitragen, da zurzeit oft unterschiedliche Rechtsauffassungen der beteiligten Stellen zur Auslegung des bestehenden Datenschutzrechts Datenverknüpfungen verhindern.
573. Die sehr dezentralisierte Forschungsdateninfrastruktur in Deutschland führt zu unnötigen Hürden beim Datenzugang. So muss für Forschungsvorhaben bei den FDZ des Bundes und der Länder bislang teilweise die Zustimmung aller Länder einzeln eingeholt werden. Stattdessen sollte **der Zugang zu Forschungsdaten im Statistischen Verbund stärker koordiniert und zentralisiert werden**. Das Standardverfahren sollte von der Notwendigkeit einer Zustimmung aller beteiligter Stellen zu einem Widerspruchsverfahren verschoben oder ein zentralisiertes Verfahren eingeführt werden. So könnte zur Genehmigung von Forschungsvorhaben die Zustimmung einer zentralen Stelle, beispielsweise bei Destatis, ausreichen. Zudem sollte es eine bundeseinheitliche Auslegung der Datenschutzregeln geben, deren bislang uneinheitliche Auslegung ein Zugangshindernis darstellt (RatSWD, 2023b). Um diese Ziele zu erreichen, sollte **§ 16 Abs. 6 BStatG so formuliert werden, dass alle Akteure des Statistischen Verbunds den Forschungsdatenzugang, die Geheimhaltung und den Datenschutz einheitlich regeln**. Dazu könnte beispielsweise Destatis in diesem Bereich die Kompetenz zur Erarbeitung einheitlicher, für den gesamten Statistischen Verbund geltender Zugangs- und Geheimhaltungsrichtlinien erhalten.

574. Die Nutzerfreundlichkeit des **Datenzugangs zu Mikrodaten** für die Wissenschaft könnte durch eine **zentrale Registrierung von Wissenschaftler:innen und Wissenschaftlern** sowie von Forschungseinrichtungen erhöht werden. Dabei könnte eine neue Einrichtung beim RatSWD als zentrale Registrierungsstelle für alle FDZ fungieren. Alternativ könnte die Akkreditierung bei einem FDZ automatisch den Zugang zu allen anderen gewähren. Voraussetzung für solche Lösungen ist jedoch eine **Vereinheitlichung der Rechtsgrundlagen für den Datenzugang**. Mit der Einführung einer neuen Indikatorik, etwa zur Bearbeitungsdauer der Anträge, könnten ausgesuchte Aspekte der Qualität des Datenzugangs nachverfolgt werden, um Problemstellen zu identifizieren. Damit die Sicherheit der Daten nicht von der Erhöhung der Nutzerfreundlichkeit beeinflusst wird, ist Sorgfalt bei Überprüfung der Nutzungsberechtigung der antragstellenden Einrichtungen entscheidend. Denkbar sind in diesem Kontext auch **Zugriffseinschränkungen für Institutionen, deren Forscher:innen und Forscher nachweislich die Datengeheimhaltungsregeln nicht einhalten**. Ein solches Vorgehen ist beispielsweise in Dänemark üblich.
575. Um eine übersichtliche Dateninfrastruktur zu schaffen, ist eine **öffentlich finanzierte, nicht zugangsbeschränkte Plattform zur Datenübersicht** notwendig. Eine solche Plattform wurde als zentrale Informationsstelle **im europäischen Data Governance Act vorgeschrieben**. Diese Verordnung hätte bis September 2023 in nationales Recht umgesetzt werden müssen. Deutschland hat den Data Governance Act bislang nicht umgesetzt und ist somit säumig. Dies könnte zu einem Vertragsverletzungsverfahren nach Art. 258 AEUV führen. Eine schnelle, forschungsorientierte Umsetzung durch die Bundesregierung ist also dringend geboten. Parallel dazu wäre eine Verbesserung der verfügbaren Metadaten hilfreich. Zudem sollten Datenstrukturfiles bereitgestellt werden, die Off-site zur Vorbereitung von Forschungsaufenthalten genutzt werden können. ↘ [ZIF-FER 551](#)
576. Um einen flächendeckenden **Zugriff auf formal anonymisierte Daten über ein Remote Access System** zu ermöglichen, sind **Anpassungen der Statistikgesetzgebung** notwendig. So müsste ein Fernzugriff beispielsweise in § 16 Abs. 6 BStatG für den Zugang zu amtlichen Daten oder in § 75 SGB X für den Zugang zu Daten der BA geregelt werden (RatSWD, 2023a). Die rechtlichen Voraussetzungen für die Verknüpfung von Registern untereinander müssen geschaffen werden und für die Umsetzung muss es einen eindeutigen Identifikator geben, wie dies in vielen europäischen Ländern der Fall ist.
577. Die **Nutzerfreundlichkeit des Zugangs zu höher aggregierten, öffentlich verfügbaren Daten sollte ebenfalls verbessert werden**. Bei makroökonomischen Zeitreihen können insbesondere die öffentlich zugänglichen Datenbanken der Federal Reserve Bank of St. Louis (FRED), die Zeitreihendatenbank der Deutschen Bundesbank, sowie das erst kürzlich überholte Datenangebot der OECD als Vorbilder für eine Verbesserung von GENESIS-ONLINE oder für die Schaffung eines zusätzlichen Zugangswegs für Zeitreihendaten dienen (Bachmann et al., 2023a). Zusätzlich sollen die Daten in einer Form bereitgestellt werden, die eine möglichst vielseitige Nutzung erlaubt. Hier sind insbesondere **Datenbanken, die keine fertigen Aggregate, sondern anfragespezifisch**

aggregierbare Mikrodaten beinhalten, besonders hilfreich. Dadurch wird es den Nutzerinnen und Nutzern ermöglicht, eigene Statistiken zu konstruieren. So wäre es z. B. möglich, die Momente der statistischen Verteilung – etwa den Median oder die Standardabweichung und nicht nur den Mittelwert – auszugeben sowie die Daten entsprechend der Forschungsfrage zu aggregieren. Eine entsprechende Infrastruktur wird derzeit gemeinsam von den Statistischen Landesämtern Berlin-Brandenburg, Nord (Hamburg und Schleswig-Holstein) sowie Rheinland-Pfalz erprobt und könnte im Jahr 2024 in den Regelbetrieb gehen. Um Skaleneffekte zu realisieren und den Datenzugang insgesamt benutzerfreundlicher zu gestalten, wäre es aus Sicht des Sachverständigenrates zielführend, eine entsprechende Infrastruktur auf den gesamten Statistischen Verbund auszuweiten.

ANHANG

▸ TABELLE 25

Akkreditierte Forschungsdatenzentren in Deutschland im Jahr 2023

Abkürzung	Bezeichnung
BAMF-FDZ	Forschungsdatenzentrum des Bundesamtes für Migration und Flüchtlinge
BIBB-FDZ	Forschungsdatenzentrum im Bundesinstitut für Berufsbildung
DeZIM.fdz	Forschungsdatenzentrum des Deutschen Zentrums für Integrations- und Migrationsforschung
EBDC	LMU-ifo Economics & Business Data Center
FDSZ Bundesbank	Forschungsdaten- und Servicezentrum der Deutschen Bundesbank
FDZ ALLBUS	Forschungsdatenzentrum ALLBUS bei GESIS
FDZ BA im IAB	Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung
FDZ Bildung	Forschungsdatenzentrum Bildung am DIPF (Leibniz-Institut für Bildungsforschung und Bildungsinformation)
FDZ BZgA	Forschungsdatenzentrum der Bundeszentrale für gesundheitliche Aufklärung
FDZ eLabour	Zentrum für qualitative arbeitssoziologische Forschungsdaten
FDZ GePaRD	Pharmakoepidemiologische Forschungsdatenbank
FDZ GML	Forschungsdatenzentrum German Microdata Lab bei GESIS
FDZ im KBA	Forschungsdatenzentrum im Kraftfahrt-Bundesamt
FDZ Internationale Umfrageprogramme	Forschungsdatenzentrum Internationale Umfrageprogramme bei GESIS
FDZ IZA, IDSC	Internationales Forschungsdatenzentrum des Forschungsinstituts zur Zukunft der Arbeit
FDZ pairfam	Forschungsdatenzentrum des Beziehungs- und Familienpanels
FDZ PIAAC	Forschungsdatenzentrum PIAAC bei GESIS
FDZ PsychData am ZPID	Forschungsdatenzentrum des Leibniz-Instituts für Psychologie
FDZ Qualiservice	Forschungsdatenzentrum Qualiservice
FDZ Ruhr am RWI	Forschungsdatenzentrum Ruhr am RWI – Leibniz-Institut für Wirtschaftsforschung
FDZ SHARE	Forschungsdatenzentrum des Survey of Health, Ageing and Retirement in Europe
FDZ WAHLEN	Forschungsdatenzentrum Wahlen bei GESIS
FDZ Wissenschaftsstatistik	Forschungsdatenzentrum Wissenschaftsstatistik
FDZ-AGD	Forschungsdatenzentrum Archiv für Gesprochenes Deutsch am Leibniz-Institut für Deutsche Sprache
FDZ-aviDA	Forschungsdatenzentrum für audio-visuelle Daten der qualitativen Sozialforschung
FDZ-BAuA	Forschungsdatenzentrum der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
FDZ-BO	Forschungsdatenzentrum Betriebs- und Organisationsdaten
FDZ-Bund	Forschungsdatenzentrum des Statistischen Bundesamtes
FDZ-DJJ	Forschungsdatenzentrum des Deutschen Jugendinstituts
FDZ-DZA	Forschungsdatenzentrum des Deutschen Zentrums für Altersfragen
FDZ-DZHW	Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung
FDZ-FGZ	Forschungsdatenzentrum des Forschungsinstituts Gesellschaftlicher Zusammenhalt
FDZ-IQB	Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen
FDZ-IWH	Forschungsdatenzentrum des Leibniz-Instituts für Wirtschaftsforschung Halle
FDZ-Land	Forschungsdatenzentrum der Statistischen Ämter der Länder
FDZ-LifBi	Forschungsdatenzentrum des Leibniz-Instituts für Bildungsverläufe
FDZ-RKI	Forschungsdatenzentrum des Robert Koch-Instituts
FDZ-RV	Forschungsdatenzentrum der Rentenversicherung
FDZ-SOEP	Forschungsdatenzentrum des Sozio-oekonomischen Panels am DIW Berlin
IÖR-Monitor	Monitor der Siedlungs- und Freiraumentwicklung
SAFE FDZ	Forschungsdatenzentrum des Leibniz-Instituts für Finanzmarktforschung SAFE
ZEW-FDZ	ZEW-Forschungsdatenzentrum

Quelle: RatSWD

© Sachverständigenrat | 23-319-01

Daten zur Tabelle

LITERATUR

- [Alstadsæter, A. et al. \(2020\)](#), The first weeks of the coronavirus crisis: Who got hit, when and why? Evidence from Norway, NBER Working Paper 27131, National Bureau of Economic Research, Cambridge, MA.
- [Altmann, S. et al. \(2023\)](#), Stellungnahme der UAG Arbeitsmarkt und Sozialversicherung, RatSWD Working Paper, Rat für Sozial- und Wirtschaftsdaten, Berlin, im Erscheinen.
- [Bachmann, R., A. Balleer, Z. Enders, N. Garnadt und H. Stüber \(2023a\)](#), Antworten zur öffentlichen Konsultation zum Forschungsdatengesetz, Fragebogen zum Forschungsdatengesetz, Verein für Socialpolitik, AG Daten, Unter-AG Makroökonomik, Berlin.
- [Bachmann, R., A. Balleer, Z. Enders, N. Garnadt und H. Stüber \(2023b\)](#), Für einen besseren Datenzugang: Stellungnahme der Arbeitsgruppe Makroökonomische Daten, RatSWD Working Paper, Rat für Sozial- und Wirtschaftsdaten, Berlin, im Erscheinen.
- [Bachmann, R., A. Peichl und R.T. Riphahn \(2021\)](#), Bessere Daten – bessere Politik, Frankfurter Allgemeine Zeitung, Frankfurt am Main, 6. August.
- [Bender, S., J. Wagner und M. Zwick \(2007\)](#), 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 5/2007, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- [BMWK und BMI \(2023\)](#), Konzept zum Aufbau des Dateninstituts, Bundesministerium für Wirtschaft und Klimaschutz sowie Bundesministerium des Innern und für Heimat, Berlin.
- [Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad und B. Setzler \(2023\)](#), How much should we trust estimates of firm effects and worker sorting?, *Journal of Labor Economics* 41 (2), 291–322.
- [Brandt, M., D. Oberschachtsiek und R. Pohl \(2007\)](#), Neue Datenangebote in den Forschungsdatenzentren – Betriebs- und Unternehmensdaten im Längsschnitt –, FDZ-Arbeitspapier 23, Statistische Ämter des Bundes und der Länder, Wiesbaden.
- [Bundesregierung \(2019\)](#), Fortentwicklung der Evaluierungskonzeption der Bundesregierung, Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 26. November 2019, Pressemitteilung, Berlin, 27. November.
- [Bundesregierung \(2013\)](#), Konzeption zur Evaluierung neuer Regelungsvorhaben, Vom St-Ausschuss Bürokratieabbau am 23.01.2013 beschlossene Fassung, Berlin.
- [Currie, J., H. Kleven und E. Zwiars \(2020\)](#), Technology and big data are changing economics: Mining text to track methods, *AEA Papers and Proceedings* 110, 42–48.
- [Decker, J. und D. Vorgrimler \(2021\)](#), Die Rolle der amtlichen Statistik innerhalb des Datenökosystems, *WISTA – Wirtschaft und Statistik* 3/2021, 17–26.
- [Dickopf, X., C. Janz und T. Mucha \(2019\)](#), Vom BIP-Flash zum BIP-Nowcast: Erste Ergebnisse einer Machbarkeitsstudie zur weiteren Beschleunigung der BIP-Schnellschätzung, *WISTA – Wirtschaft und Statistik* 6/2019, 47–58.
- [Emmenegger, J. und R. Münnich \(2023\)](#), Localising the upper tail: How top income corrections affect measures of regional inequality, *Jahrbücher für Nationalökonomie und Statistik* 243 (3–4), 285–317.
- [Europäisches Parlament \(2023\)](#), Energy performance of buildings (recast), Amendments adopted by the European Parliament on 14 March 2023 on the proposal for a directive of the European Parliament and of the Council on the energy performance of buildings (recast) (COM(2021)0802 – C9-0469/2021 – 2021/0426(COD)).
- [Fuchs, R., T. Göllner, S. Hartmann und T. Thomas \(2023\)](#), Fostering excellent research by the Austrian Micro Data center (AMDC), *Jahrbücher für Nationalökonomie und Statistik*, im Erscheinen, <https://doi.org/10.1515/jbnst-2023-0043>.
- [Gottschalk, S., C. Krolage, G. Licht, A. Peichl, S. Schaffner und S. Wichert \(2023\)](#), Besserer Zugang zu Unternehmensdaten für die empirische Wirtschaftsforschung in Deutschland“, RatSWD Working Paper, Rat für Sozial- und Wirtschaftsdaten, Berlin, im Erscheinen.

- Gürke, C., A. Gruhl und T. Hethey-Maier (2011), Verknüpfung von Unternehmensdaten verschiedener Datenproduzenten Das Projekt „Kombinierte Firmendaten für Deutschland“, WISTA – Wirtschaft und Statistik 2/2011, 91–97.
- Hahlen, J. (2009), Zur Rolle der amtlichen Statistik für eine evidenzbasierte Wirtschaftsforschung und -politik, WISTA – Wirtschaft und Statistik 10/2009, 1021–1030.
- Hertweck, F., I.E. Isphording, S.H. Matthewes, K. Schneider und C.K. Spieß (2023), Bildungspolitik braucht Forschung – doch Forschung braucht bessere Daten, RatSWD Working Paper, Rat für Sozial- und Wirtschaftsdaten, Berlin, im Erscheinen.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.) (2001), Wege zu einer besseren informationellen Infrastruktur: Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der Informationellen Infrastruktur zwischen Wissenschaft und Statistik, 1. Auflage, Nomos Verlagsgesellschaft, Baden-Baden.
- KomZS (2023), Abschlussbericht, Kommission Zukunft Statistik, Wiesbaden, im Erscheinen.
- Krause, A., M. Zimmermann und I. Herda (2022), Überlegungen zu einem Gebäude- und Wohnungsregister: Aufbau, Pflege und Nutzung, WISTA – Wirtschaft und Statistik 4/2022, 25–38.
- Kruse, H.W., F. Hieber, H. Limberg, B. Zapf und D. Boddin (2023), Außenhandelsaktive Unternehmen: neue Analysemöglichkeiten durch Mikrodatenverknüpfung, WISTA – Wirtschaft und Statistik 5/2023, 13–26.
- Löffler, M., A. Peichl, C. Wittneben und C. Woodage (2015), Möglichkeiten zur Verbesserung der statistischen Datengrundlage zur Beschreibung höchster Einkommen und Vermögen: Abschlussbericht, Research Report, ZEW Kurzexpertise 58, Zentrum für Europäische Wirtschaftsforschung, Mannheim.
- Ludvigsson, J.F., P. Svedberg, O. Olén, G. Bruze und M. Neovius (2019), The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research, European Journal of Epidemiology 34, 423–437.
- Müller, D. und J. Möller (2019), Giving the international scientific community access to German labor market data: A success story, in: Crato, N. und P. Paruolo (Hrsg.), Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design, Springer International Publishing, Cham, 101–117.
- Nagaraj, A. und M. Tranchero (2023), How does data access shape science? Evidence from the impact of U.S. census's research data centers on economics research, NBER Working Paper 31372, National Bureau of Economic Research, Cambridge, MA.
- Palmer, K., B. Prest, S. Iler und S. Villanueva (2022), Options for EIA to publish CO2 emissions rates for electricity, RFF Report 22–8, Resources for the Future, Washington, DC.
- Profit, S. (2023), Eine moderne Dateninfrastruktur für die evidenzbasierte Politikberatung in Deutschland, Rede, Wissenschaftliche Fachtagung „Daten.Forschung.Zukunft“, Session "Datenkompetenz und Nationale Dateninfrastruktur: Herausforderungen und Schlüssel für Demokratie und Entscheidungen", Wiesbaden, 6. Juli.
- RatSWD (2023a), Positionspapier des RatSWD: Nutzung von Registerdaten für Zwecke der Forschung sichern, Rat für Sozial- und Wirtschaftsdaten, Berlin.
- RatSWD (2023b), Forschungsdatengesetz: Was zentral ist, Rat für Sozial- und Wirtschaftsdaten, Berlin.
- RatSWD (2022a), Positionspapier des RatSWD: Aufbau eines Bildungsverlaufsregisters: Datenschutzkonform und forschungsfreundlich, Rat für Sozial- und Wirtschaftsdaten, Berlin.
- RatSWD (2022b), Positionspapier des RatSWD: Eckpunkte für ein Forschungsdatengesetz, Empfehlungen des Rates für Sozial- und Wirtschaftsdaten für die Erarbeitung eines Forschungsdatengesetzes, Rat für Sozial- und Wirtschaftsdaten, Berlin.
- RatSWD (2019), Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger, RatSWD Output 5(6), Rat für Sozial- und Wirtschaftsdaten, Berlin.
- U.S. Bureau of Labor Statistics (2020), Measuring the effects of the coronavirus (COVID-19) pandemic using the current population survey, <https://www.bls.gov/bls/measuring-the-effects-of-the-coronavirus-covid-19-pandemic-using-the-current-population-survey.htm>, abgerufen am 28.10.2023.
- Wissenschaftlicher Beirat beim BMF (2021), Gutachten des Wissenschaftlichen Beirats beim BMF zur Verbesserung der Dateninfrastruktur für die Steuerpolitik, Monatsbericht des BMF Januar 2021, 30–36.