



Dr. Hanna Brenzel

hat internationale Volkswirtschaft studiert und leitet das Referat „Forschungsdatenzentrum, Methoden der Datenanalyse“ im Statistischen Bundesamt.



Prof. Dr. Markus Zwick

ist Volkswirt und leitet die Gruppe „Institut für Forschung und Entwicklung in der Bundesstatistik, Forschungsdatenzentrum“ des Statistischen Bundesamtes.

EINE INFORMATIONELLE INFRASTRUKTUR IN DEUTSCHLAND IST ERWACHSEN – DAS FORSCHUNGS-DATENZENTRUM DES STATISTISCHEN BUNDESAMTES

Hanna Brenzel, Markus Zwick

📌 **Schlüsselwörter:** Wissenschaftskooperation – Einzeldaten – Mikrodaten – Remote Access – Anonymisierung – empirische Wirtschaftsforschung

ZUSAMMENFASSUNG

Die informationelle Infrastruktur in Deutschland hat sich seit Ende der 1990er-Jahre beträchtlich fortentwickelt, wenngleich zuletzt die Coronakrise gezeigt hat, dass sie noch erhebliche Defizite aufweist. Die Erkenntnis, dass Einzeldaten einen Wert haben, und die Nachfrage der empirischen Forschung nach amtlichen Mikrodaten führten im Jahr 2001 zur Einrichtung des Forschungsdatenzentrums des Statistischen Bundesamtes. Der Beitrag zeichnet dessen Entwicklung in den letzten gut 20 Jahren nach und erläutert das künftig vorgesehene Remote-Access-System, das Datennutzenden einen verbesserten Datenzugang bieten wird.

📌 **Keywords:** Cooperation with the scientific community – individual data – microdata – remote access – anonymisation – empirical economic research

ABSTRACT

The information infrastructure in Germany has progressed significantly since the end of the 1990s although the Covid-19 crisis has shown recently that it still has considerable deficits. The insight that individual data have a value and the demand for official microdata on the part of empirical economic researchers led to the establishment of the Research Data Centre of the Federal Statistical Office in 2001. The article describes its development over the past 20 years and explains the envisaged remote access system, which will provide improved data access for data users.

1

Einleitung

Im Sommer 2001 hat das Statistische Bundesamt als erster amtlicher Datenproduzent ein Forschungsdatenzentrum (FDZ) eingerichtet. Seitdem bietet das FDZ des Statistischen Bundesamtes – ab 2002 gemeinsam mit dem FDZ der Statistischen Ämter der Länder¹ – eine Datendienstleistung zwischen zwei Verfassungszielen an, die nicht immer einfach simultan zu bedienen sind. Dies ist auf der einen Seite die Wissenschaftsfreiheit, also der Zugang zu öffentlich finanzierten Daten, und auf der anderen Seite die informationelle Selbstbestimmung.

Die Einrichtung des FDZ war eine unmittelbare Folge der Empfehlungen der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI), die diese im Frühjahr 2001 in einem Gutachten vorgelegt hat (KVI-Gutachten; KVI, 2001). Die Etablierung von Forschungsdatenzentren bei den öffentlichen Datenproduzenten war als 29. Empfehlung aus dem KVI-Gutachten die wohl erfolgreichste und nachhaltigste, neben der zweiten Empfehlung, einen Rat für Sozial- und Wirtschaftsdaten (RatSWD) einzurichten.

Im Jahr 2002 folgten dann die Statistischen Ämter der Länder, die Bundesagentur für Arbeit und die Deutsche Rentenversicherung diesem Schritt des Statistischen Bundesamtes und richteten ebenfalls Forschungsdatenzentren ein. Mittlerweile, rund 20 Jahre später, existieren [41 vom RatSWD akkreditierte Forschungsdatenzentren](#) und erlauben einen sehr breiten Zugang zu mit öffentlichen Mitteln produzierten Einzeldaten.

Die Coronakrise hat unter anderem gezeigt, dass die informationelle Infrastruktur in Deutschland trotz aller Fortschritte noch erhebliche Defizite aufweist. Einerseits sind 41 akkreditierte Forschungsdatenzentren eine beeindruckende Zahl, andererseits ist sie aber auch ein Ausdruck der Datensilos, von denen es in Deutschland weiterhin sehr viele gibt. Mehr denn je existieren Daten an verschiedensten Stellen, privat wie öffentlich, Teile davon durchaus verfügbar, aber häufig in mehr oder weniger unbekannter Qualität. Immer noch fehlen sowohl ein Überblick über diese Daten als auch ein

gesetzlicher Rahmen, um diese Daten im Bedarfsfall mit anderen, beispielsweise amtlichen Daten für die empirische Politikberatung zusammenzuführen.

Der Fokus des folgenden Artikels liegt auf den erheblichen Fortschritten, die bezüglich der informationellen Infrastruktur in Deutschland in den letzten 25 Jahren gemacht wurden. Keiner der Initiatoren der Entwicklung – weder aus dem Bereich der Wissenschaft noch vonseiten der Datenproduzenten – hätte erwartet, dass sich der Zugang zu öffentlich finanzierten Einzeldaten so entscheidend weiterentwickeln würde. Diese Entwicklung hat im Grunde mit der ersten Regelung zur Weitergabe von Einzeldaten im Bundesstatistikgesetz von 1981 begonnen und wird mit dem nun vorgesehenen automatisierten Fernzugriff auf diese Daten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (Remote Access) nicht enden.

2

Einzeldaten haben einen Wert

Die Analyse von Einzeldaten und die damit verbundene Nachfrage seitens der Wissenschaft hat noch keine lange Tradition und geht einher mit der bemerkenswerten Entwicklung der Informationstechnologie.² Insbesondere mit dem seitens der Deutschen Forschungsgemeinschaft geförderten Sonderforschungsbereich 3 „Mikroanalytische Grundlagen der Gesellschaftspolitik“ (Hauser und andere, 1994) nahm der Bedarf der Wissenschaft an Einzeldaten immer weiter zu. Parallel dazu erfolgte die breitere Ausstattung der Universitäten mit Großrechneranlagen in den siebziger und achtziger Jahren des letzten Jahrhunderts mit der Möglichkeit, immer größere Datenmengen zu verarbeiten. Schon seit den späten 1960er-Jahren gab es verschiedentlich Kooperationen zwischen der Wissenschaft und der amtlichen Statistik (Krupp, 2004). Diese erfolgten jedoch auf der Rechtsgrundlage des Bundesstatistikgesetzes aus dem Jahr 1951, das einen Einzeldatenzugang noch nicht kannte. Zu dieser Zeit wurde Einzeldaten kein eigener Wert beigemessen – sie wurden zur Aggregation benötigt und danach schnell gelöscht.

1 Siehe hierzu auch Hlawatsch und andere (2022).

2 Als Startpunkt der mikroanalytischen Betrachtungen, also der Analyse einzelner Merkmalsträger, gilt Orcutt (1957).

Erst die Novellierung des Bundesstatistikgesetzes im Jahr 1981 hat den Zugang zu amtlichen Einzeldaten rechtlich geregelt. Dies geschah allerdings in der Form der sogenannten absoluten Anonymität, bei der einzelne Daten den Merkmalsträgern nach menschlichem Ermessen nicht mehr zuzuordnen sind. Sehr schnell zeigte sich, dass diese Form der Anonymisierung insbesondere für die wissenschaftliche Forschung viel zu rigide war. Die wenigen anonymisierten Daten, die das Statistische Bundesamt im Zusammenhang mit Sonderarbeiten erstellt hat, waren kostenintensiv und wiesen aufgrund der informationsreduzierenden Anonymisierungsverfahren nur noch geringes Analysepotenzial auf. Hier reagierte der Gesetzgeber ausgesprochen schnell mit der Änderung des Bundesstatistikgesetzes im Jahr 1987 und der Etablierung der sogenannten faktischen Anonymität³ im §16 Absatz 6 Bundesstatistikgesetz und dem damit verbundenen Wissenschaftsprivileg.⁴

Mit der Regelung zur faktischen Anonymität und den ersten Scientific-Use-Files (SUF) des Mikrozensus und der Einkommens- und Verbrauchsstichprobe glaubte man seitens der statistischen Ämter Mitte der 1990er-Jahre, einen sachgerechten Zugang zu amtlichen Einzeldaten gefunden zu haben. Dies war bei weitem nicht der Fall und zeigt auch, warum die Wissenschaft zu den sehr fordernden Nutzern amtlicher Daten gehört: Lösungen bedienen immer nur den aktuellen Zeitpunkt, während die Wissenschaft immer auf dem Weg und bei einer neuen Lösung häufig schon mit der nächsten Wissenschaftsgeneration einen Schritt weiter ist.

Den nächsten Meilenstein der Entwicklung leiteten die Professoren Hauser, Wagner und Zimmermann 1998 mit ihrem Memorandum zu den Erfolgsbedingungen empirischer Wirtschaftsforschung ein (Hauser und andere, 1998). Sie sprachen wesentliche offene Fragen wie den Zugang zu wirtschaftsstatistischen Daten, die Verknüpfung von Daten im Längs- und Querschnitt sowie weitere Zugangsmöglichkeiten wie das kontrollierte Fernrechnen an. Erstmals griff auch der damalige Präsident des Statistischen Bundesamtes, Johann Hahlen, in der Folge mit einem Kommentar in diese Diskussion mit ein (Hahlen, 1998).

3 Zu den verschiedenen Formen der Anonymität siehe auch den Exkurs in Kapitel 4.

4 Hierzu und zu den ersten Anwendungen zum Mikrozensus und zur Einkommens- und Verbrauchsstichprobe (EVS) siehe Müller und andere (1991).

3

Der Weg über den Gründungsausschuss bis zum KonsortSWD

Wohl keiner der Autoren des oben genannten Memorandums hätte auch nur ansatzweise den Erfolg ihrer Anmerkungen erwartet. Das Statistische Bundesamt griff das Thema gemeinsam mit führenden empirischen Wissenschaftlerinnen und Wissenschaftlern auf und lud 1999 zum Symposium „Kooperation zwischen Wissenschaft und amtlicher Statistik“ nach Wiesbaden (Statistisches Bundesamt, 1999). Das Symposium war geprägt von einem in hohem Maße offenen und konstruktiven Austausch zwischen den Vertreterinnen und Vertretern der Wissenschaft und der amtlichen Datenproduktion.⁵ Die wesentlichen Inhalte und Schlussfolgerungen der Veranstaltung wurden in einem kurzen Arbeitspapier mit acht Thesen zusammengefasst und der damaligen Bundesministerin für Bildung und Forschung, Edelgard Bulmahn, überreicht. Noch im Jahr 1999 berief daraufhin die Ministerin die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik – damals häufig auch Bulmahn-Kommission genannt. Den gemeinsamen Vorsitz der Kommission übernahmen Johann Hahlen, der als damaliger Präsident des Statistischen Bundesamtes die Seite der Datenproduktion vertrat, und Hans-Jürgen Krupp als weltweit anerkannter empirischer Wissenschaftler.

Die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik legte im Jahr 2001 ihr vielbeachtetes KVI-Gutachten vor (KVI, 2001). Insgesamt umfasst das Gutachten 36 Empfehlungen, viele davon sind mittlerweile umgesetzt, wie die Einrichtung von Forschungsdatenzentren, die Zugang zu mit öffentlichen Mitteln entstandenen Einzeldaten anbieten, oder dem Rat für Sozial- und Wirtschaftsdaten. Eine erste größere Würdigung der erzielten Resultate erfolgte mit der Veröffentlichung „Fortschritte der informationellen Infrastruktur in Deutschland“ im Jahr 2008 (Rolf und andere, 2008). Die gemeinsame Festschrift zum 65. Geburtstag von Johann Hahlen und zum 75. Geburtstag von Hans-Jürgen Krupp zeichnet den hier skizzierten Weg in zahlreichen Beiträgen nach – über

5 Hierzu siehe insbesondere Statistisches Bundesamt (1999, hier: Seite 216 ff.) mit den Beiträgen der Paneldiskussion.

den Gründungsausschuss des Rats für Sozial- und Wirtschaftsdaten zum Rat für Sozial- und Wirtschaftsdaten (RatSWD) sowie zu den Arbeiten der Forschungsdaten- und Servicezentren.

Der 2004 eingerichtete [RatSWD](#) ist weiterhin eines der wichtigsten Beratungsgremien der Bundes- und Landesregierungen im Kontext der Forschungsdateninfrastruktur. Dass Daten einer der wichtigsten Rohstoffe der Gegenwart sind, ließ sich zu Beginn, trotz aller Bedeutung, kaum absehen. Dies hat sich über die Zeit gewandelt. Aufbauend auf den Erfahrungen des Rats für die Sozial- und Wirtschaftswissenschaften hat die Gemeinsame Wissenschaftskonferenz des Bundes und der Länder (GWK) im Jahr 2014 den [Rat für Informationsinfrastrukturen \(RfII\)](#) zur koordinierten Weiterentwicklung des Informationsinfrastrukturaufbaus eingesetzt. Eine zentrale Empfehlung des RfII war die Einrichtung einer [Nationalen Forschungsdateninfrastruktur \(NFDI\)](#); sie koordiniert seit 2020 Aktivitäten verschiedener Konsortien aus zahlreichen Wissenschaftsbereichen. Für die Sozialwissenschaften ist dies das [Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften \(KonsortSWD\)](#), ein Teil davon ist weiterhin der RatSWD mit seiner Forschungsdateninfrastruktur von 41 akkreditierten Forschungsdatenzentren.

4

Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

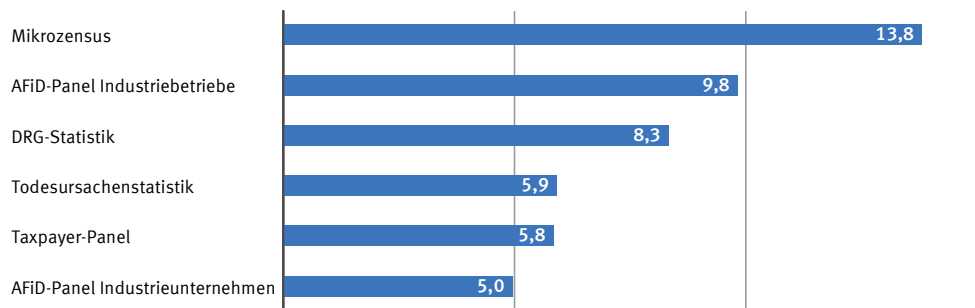
Das FDZ des Statistischen Bundesamtes bietet gemeinsam mit dem FDZ der Statistischen Ämter der Länder der empirisch arbeitenden Wissenschaft seit nunmehr über 20 Jahren ein abgestimmtes Daten- und Dienstleistungsangebot für die wissenschaftliche Nutzung von qualitativ hochwertigen Mikrodaten der amtlichen Statistiken an. Dieses Angebot ist eine Grundvoraussetzung für den Erfolg der empirisch arbeitenden Forschung, was sich nicht zuletzt in einer steigenden Nutzerzahl über die Jahre hinweg zeigt. Dabei arbeiten die FDZ fortlaufend daran, den Zugang zu den Mikrodaten zu verbessern und an die sich ändernden Bedarfe der Wissenschaft anzupassen.

Von der Gründung im Jahr 2001 bis heute (August 2022) haben die FDZ der Statistischen Ämter des Bundes und der Länder 3 900 wissenschaftliche Projekte von mehr als 400 nutzungsberechtigten Einrichtungen betreut. Insgesamt stehen der Wissenschaft 3 438 verschiedene Produkte⁶ aus insgesamt 90 verschiedenen Statistiken zur Verfügung. Dabei werden nicht alle Statistiken mit

⁶ Ein Produkt ergibt sich aus einem einzelnen Berichtsjahr einer Statistik und Zugangsweg, beispielsweise Mikrozensus 2019 als Scientific-Use-File.

Grafik 1

Die am stärksten in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder nachgefragten Statistiken auf Produktebene von 2013 bis August 2022
Anteile an der Gesamtnachfrage in %



N = 17 569 Produkte. Referenzjahr ist 2013 aufgrund der Vergleichbarkeit, die mit der Neukonzeption der Nutzendendatenbank im Jahr 2013 einhergeht.

2022 - 0316

der gleichen Intensität nachgefragt: Mit knapp 14 % der Gesamtnachfrage nach allen Produkten hat der Mikrozensus den größten Anteil. Auf den Plätzen 2 und 3 liegen das AFiD-Panel Industriebetriebe und die DRG-Statistik. Jeweils knapp 6 % aller Nachfragen beziehen sich auf die Todesursachenstatistik und das Taxpayer-Panel. Mit 5,0 % aller Nachfragen ist ein weiteres Produkt aus der Produktreihe Amtliche Firmendaten für Deutschland vertreten, das AFiD-Panel Industrieunternehmen. Diese sechs Statistiken decken rund die Hälfte der Gesamtnachfrage im Zeitraum von 2013 bis August 2022 ab.

↳ Grafik 1

Zählt man alle Datennutzende über die Jahre zusammen, so wurden 6 365 Personen (Studierende, Professorinnen und Professoren sowie weitere Wissenschaftlerinnen und Wissenschaftler) durch die Mitarbeiterinnen und Mitarbeiter der FDZ über die Jahre hinweg beraten und betreut. Die große Vielfalt der Datenprodukte ergibt sich unter anderem auch durch die breite Abdeckung an unterschiedlichsten Datenzugangswegen, die der Wissenschaft durch die FDZ zur Verfügung stehen. Bereits seit ihrer Gründung existierten vier unterschiedliche, an den Bedarf der Nutzenden angepasste Zugangswege, die sich im Wesentlichen in ihrem Analysepotenzial und Anonymisierungsgrad unterscheiden (siehe hierzu auch Zühlke und andere, 2003):

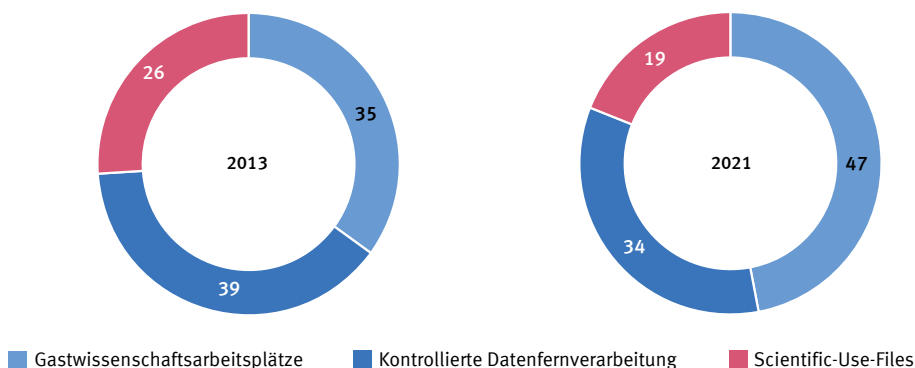
- › Public-Use-Files/Campus-Files
- › Scientific-Use-Files (SUF)
- › Gastwissenschaftsarbeitsplätze (GWAP)
- › Kontrollierte Datenfernverarbeitung (KDFV)

Insbesondere die letzten drei Zugangswege sind zentral für die empirische Wirtschafts- und Sozialforschung, da das Analysepotenzial aufgrund der geringeren Anonymisierungsmaßnahmen die Beantwortung einer Vielzahl an Forschungsfragen erlaubt. Beim Vergleich, wie sich die Anteile der je Zugangsweg genutzten Produkte von 2013 zu 2021 verändert haben, ist eine Verschiebung hin zur Nutzung von Gastwissenschaftsarbeitsplätzen zu erkennen. Sie geht einher mit einer Reduktion der Datennutzung von Scientific-Use-Files und der kontrollierten Datenfernverarbeitung. Zwei Gründe lassen sich hierfür vermuten: Zum einen erfordert die Mikrodatennutzung aufgrund der immer komplexer werdenden Analysemethoden und Verschneidungen von Datenprodukten eine direkte Einsichtnahme der Mikrodaten. Zum anderen hat sich der Anonymisierungsgrad des Datenmaterials am Gastwissenschaftsarbeitsplatz durch die Novellierung des Bundesstatistikgesetzes im Jahr 2016 von faktischer zu formaler Anonymität geändert. Dies hat zur Folge, dass nunmehr auch formal anonymisiertes Datenmaterial am Gastwissenschaftsarbeitsplatz der Wissenschaft zur Verfügung gestellt werden kann und damit das Analysepotenzial identisch zu dem bei der kontrollierten Datenfernverarbeitung ist. Im Unterschied zur kontrollierten Datenfernverarbeitung, die fortgeschrittene Programmierkenntnisse erfordert, ist dies beim Gastwissenschaftsarbeitsplatz nicht zwingend notwendig – hier ist es möglich, direkt mit den Einzeldaten zu arbeiten. ↳ Grafik 2

Die Zahlen verdeutlichen den hohen Stellenwert, den sich die FDZ über die Jahre in der wissenschaftlichen

Grafik 2

Je Zugangsweg genutzte Produkte in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder in %



2022 - 0317

Eine informationelle Infrastruktur in Deutschland ist erwachsen – das Forschungsdatenzentrum des Statistischen Bundesamtes

Landschaft Deutschlands aufgebaut haben und untermauern deren Erfolgsgeschichte. Mit Fortbestand der FDZ haben sich jedoch die Erwartungen an sie grundlegend weiterentwickelt und Akteure aus dem politischen und wissenschaftlichen Raum drängen seit Längerem auf eine substantielle Verbesserung der Datenzugangs- und Datennutzungsmöglichkeiten. Die technische Infrastruktur des Datenzugangs hat sich seit der Entstehung der FDZ nicht wesentlich weiterentwickelt, die Nutzerbedarfe sowie die Rahmenbedingungen haben sich hingegen grundlegend gewandelt. Datennutzende erwarten heute flexiblere Zugangsmöglichkeiten, um Mikrodaten unabhängig von den Servicezeiten der FDZ nutzen zu können, wie dies bei einigen anderen FDZ national sowie in verschiedenen europäischen Ländern bereits Standard ist. Die Coronakrise hat im Besonderen die Grenzen der bestehenden Infrastrukturen aufgezeigt. Neben den sich verändernden Erwartungen führen steigende Nutzendenzahlen und Erweiterungen des Datenangebots zunehmend zu Engpässen im Bereich der personellen Kapazitäten, da diese nicht im gleichen Umfang zugenommen haben. Zudem sind mit den aktuellen Datenzugangswegen hohe organisatorische Aufwände für die FDZ verbunden. Das aktuell in den statistischen Ämtern laufende gemeinsame Großprojekt zur Verbesserung der IT-Infrastruktur und des Datenzugangs soll diesen

Herausforderungen begegnen. Ziel ist, die FDZ der Statistischen Ämter des Bundes und der Länder wieder in eine Vorreiterrolle in der deutschen Forschungsdateninfrastrukturlandschaft zu versetzen. [↘ Grafik 3](#)

Kernelement einer Verbesserung des Datenzugangs ist die Umsetzung eines Remote-Access-Systems, das den rechtlichen Rahmenbedingungen und den geltenden IT-Sicherheitsstandards entspricht. Aktuell stehen gesetzliche Herausforderungen der Umsetzung eines Remote Access für formal anonymisierte Daten entgegen. Ein Remote Access für formal anonymisierte Daten ist unter den derzeitigen rechtlichen Bedingungen daher derzeit nicht realisierbar. Eine Umsetzungsmöglichkeit, für die keine Rechtsänderung notwendig ist, wäre, einen Remote Access für faktisch anonymisierte Daten mit geringer Datenveränderung anzubieten. In diesem Fall ist der Grad der Datenveränderung von höchster Relevanz: Bei einer zu hohen Datenveränderung werden die angebotenen Daten dem Bedarf der Wissenschaft nicht gerecht, bei einer zu niedrigen Datenveränderung kann die Geheimhaltung nicht mehr gewahrt werden. Der Grad der faktischen Anonymisierung bestimmt daher maßgeblich den Nutzen und die Abdeckung des Bedarfs der Wissenschaft.

Grafik 3

Datenzugangswege in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder

Zugangsweg	On-Site-Nutzung		Im Aufbau	Off-Site-Nutzung		Anonymisierung
	Kontrollierte Datenfernverarbeitung	Gastwissenschafts-arbeitsplätze	Remote Access	Scientific-Use-Files	Public-Use-Files/ Campus-Files	
Anonymisierungsgrad der Daten	formal anonym	formal anonym	faktisch anonym	faktisch anonym	absolut anonym	
Nutzungsberechtigt	Wissenschaftliche Einrichtung	Wissenschaftliche Einrichtung	Wissenschaftliche Einrichtung	Wissenschaftliche Einrichtung	alle	
Datenhaltung während der Nutzung	Statistische Ämter	Statistische Ämter	Statistische Ämter	Wissenschaftliche Einrichtung	beliebig	
Aufenthaltort der Nutzenden während der Nutzung	beliebig	Standorte der Forschungsdatenzentren	Wissenschaftliche Einrichtung	Wissenschaftliche Einrichtung	beliebig	
Analysepotenzial						

2022 - 0318

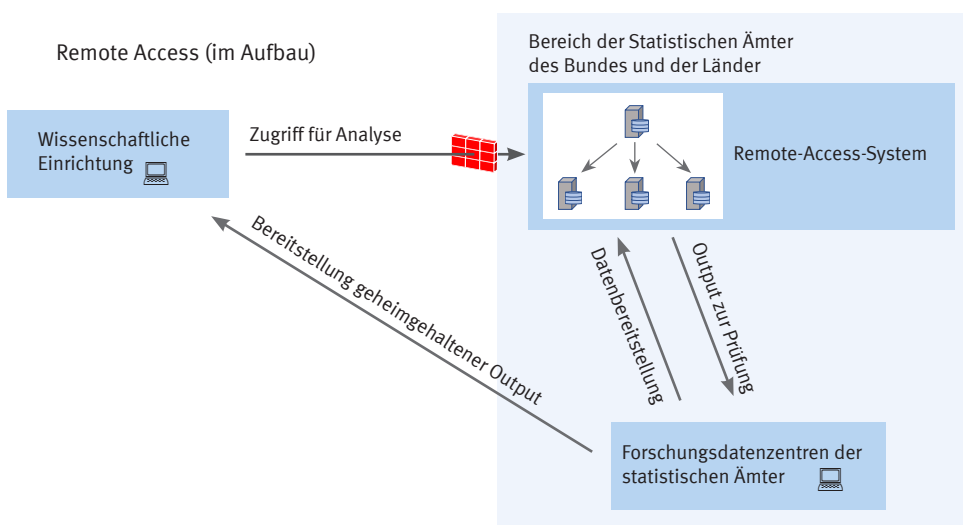
Das geplante Remote-Access-System besitzt Innovationspotenzial und hat das Potenzial, die aktuelle Arbeitsbelastung in den FDZ-Standorten zu verringern, indem bestehende Datenzugangswege (zum Beispiel die arbeitsaufwendige kontrollierte Datenfernverarbeitung) substituiert beziehungsweise reduziert werden können. Mit dieser Maßnahme greifen die FDZ die [Empfehlungen des Statistischen Beirats](#) (Statistisches Bundesamt, 2022) nach einem verbesserten Datenzugang und Remotezugriff auf. Gleichzeitig ist das Remote-Access-System aus Sicht der Datennutzenden ein innovativer, moderner und an ihren Bedarfen ausgerichteter Datenzugang. Durch das Remote-Access-System wird – mit der Investition in eine anschlussfähige Infrastruktur – die kontinuierliche Weiterentwicklung der FDZ vorangetrieben. Die Orientierung an den Bedarfen der Wissenschaft festigt dabei den Status der FDZ als moderner Datenanbieter.

Im November 2021 haben die Aufbauarbeiten für ein solches System begonnen, welches mit Blick auf Ausbau- und Weiterentwicklungsmöglichkeiten zukunftsfähig konzipiert wird. So ermöglicht das System prinzipiell die Bereitstellung von allen Daten mit einem hohen Schutzbedarf – bei einer künftigen Gesetzesänderung somit auch für formal anonymisiertes Datenmaterial. Die Forschenden können dabei direkt aus ihrer wissenschaftlichen Institution heraus auf einen geschützten Datenbereich zugreifen und Analysen mit ihrer prä-

ferierten Statistiksoftware RStudio, Stata oder SPSS durchführen. Der große Vorteil gegenüber der kontrollierten Datenfernverarbeitung ist dabei, dass die Forschenden die Mikrodaten sehen können und nicht wie bisher „blind“ ihre Syntaxen programmieren müssen. Zudem fördert das System die Datensparsamkeit: Durch die Dateneinsicht muss nicht jedes Zwischenergebnis freigegeben werden, sondern lediglich finale Ergebnisoutputs nach entsprechender Geheimhaltungskontrolle nach § 16 Bundesstatistikgesetz als absolut anonymisiertes Ergebnis. Der große Vorteil gegenüber den bisherigen Scientific-Use-Files ist, dass die Remote-Scientific-Use-Files dabei in den gesicherten Bereichen der Statistischen Ämter des Bundes und der Länder verbleiben. Durch die vom Remote-Access-System geschaffenen technisch-organisatorischen Sicherheitsmaßnahmen ist es möglich, im Vergleich zu den bisher per DVD versendeten Scientific-Use-Files den Anonymisierungsgrad der Remote-Scientific-Use-Files deutlich herabzuschrauben. Dadurch lässt sich das Analysepotenzial der Daten erhöhen und gleichzeitig im Ergebnis faktisch anonymisiertes Datenmaterial erhalten. Es entsteht somit ein neues faktisch anonymes Datenmaterial, das hinsichtlich Analysepotenzial und Anonymisierungsgrad zwischen dem bisher per DVD verschickten Scientific-Use-File und den an den Gastwissenschaftsarbetsplätzen bereitgestellten Mikrodaten einzuordnen ist.

➤ Grafik 4

Grafik 4
Das Remote-Access-System



2022 - 0322

↳ Exkurs

Mikrodaten werden als faktisch anonym bezeichnet, wenn eine Deanonymisierung zwar nicht gänzlich ausgeschlossen werden kann, die Angaben jedoch „nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft“ der jeweiligen Merkmalsträgerin beziehungsweise dem jeweiligen Merkmalsträger zugeordnet werden können (§ 16 Absatz 6 Bundesstatistikgesetz). Nach Maßgabe des Bundesstatistikgesetzes dürfen faktisch anonymisierte Daten allerdings nur von wissenschaftlichen Einrichtungen und nur zur Durchführung wissenschaftlicher Vorhaben genutzt werden. Bei der Herstellung der faktischen Anonymität besteht das Ziel darin, die Wahrscheinlichkeit einer korrekten Zuordnung von Angaben zu Auskunftgebenden nahezu auszuschließen, dabei jedoch den statistischen Informationsgehalt möglichst weitgehend zu erhalten. Dabei können unterschiedliche Anonymisierungsverfahren zur Anwendung kommen. Gängig sind Verfahren zur Informationsreduktion (zum Beispiel Aggregation, Klassenbildung, Zensorisierung) beziehungsweise zur Informationsveränderung (zum Beispiel Swapping-Verfahren). Für die Bestimmung der faktischen Anonymität müssen der Aufwand und der Nutzen einer Deanonymisierung bewertet werden. In den Forschungsdatenzentren resultiert die faktische Anonymität allerdings nicht allein aus dem verbleibenden Informationsgehalt der Daten, sondern auch aus den Rahmenbedingungen der Datennutzung und den damit einhergehenden Möglichkeiten zur Deanonymisierung. Wann ein Mikrodatsatz als faktisch anonym bezeichnet werden kann, hängt somit auch von den Zugangsbedingungen ab. Dabei ist von entscheidender Bedeutung, welches Zusatzwissen über die Merkmalsträgerinnen und -träger verfügbar ist und wo die Datennutzung stattfindet. Abhängig davon, ob die Mikrodaten außerhalb oder innerhalb der statistischen Ämter genutzt werden, kann die faktische Anonymität mit mehr oder weniger starken Informationseinbußen erreicht werden.

5

Ausblick


Die Forschungsdateninfrastruktur hat sich in den 25 Jahren seit dem Memorandum zu den Erfolgsbedingungen empirischer Wirtschaftsforschung rasant weiterentwickelt. Wohl kaum jemand, der die ersten Diskussionen getragen und verfolgt hat, wird diesen Fortschritt erwartet haben. Trotz allem ist die Frage des sachgerechten Zugangs zu den mit öffentlichen Mitteln produzierten Einzeldaten, ob für die Wissenschaft oder interessierte Dritte, noch nicht zufriedenstellend gelöst.

Dies liegt an der Natur der Sache, nämlich dem Wissensdurst der Wissenschaft. Eine zeitreisende Wissenschaftlerin aus dem Jahr 1995 wäre heute überwältigt vom Datenangebot der akkreditierten Forschungsdatenzentren. Die heutige Generation der Wissenschaft ist es nicht annähernd. Jede Wissenschaftsgeneration – und diese tauschen sich schnell aus – kommt mit neuen Fragestellungen. Niemand möchte die Arbeiten der Doktormütter und -väter in gleicher Art, nur mit aktualisierten Daten, fortführen. Dies führt zur Weiterentwicklung: von standardisierten statischen Scientific-Use-Files für Erhebungsjahre zum Paneldatensatz, zum integrierten Unternehmensdatensatz über die Zeit, zu einem kombinierten Paneldatensatz von Unternehmens- und Haushaltsdaten, angereichert mit Informationen von Sensoren, einem Datensatz, der bisher noch fehlt.

Die empirische Wissenschaft war, ist und wird mit ihrer Nachfrage nach Daten immer weiter sein als ein Datenproduzent. Dies führt zu Innovationen in den FDZ und verbessert die amtliche Statistik insgesamt. Aber es führt auch zu Frustrationen: auf der Seite der Datennutzer, die gerne mehr Möglichkeiten hätten, und auf der Seite der Produzenten in ihrem Bestreben, die Forschungsdateninfrastruktur immer weiter auszubauen. Daher ist eine permanente Kommunikation von Wissenschaft und Statistik notwendig, wie sie der RatSWD seit fast 20 Jahren erfolgreich betreibt.

Die gegenwärtigen Herausforderungen sind technisch wie rechtlich begründet. An dieser Stelle hat sich nicht viel verändert. Technisch gilt es für die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, den Zugangsweg „Remote Access“ zu etablieren. Andere akkreditierte FDZ haben gezeigt,

dass dies auch unter datenschutzrechtlichen Bedingungen möglich ist. Remote Access wäre eine deutliche Entlastung der immer knappen Ressourcen, auf Seiten der Datennutzenden wie auf der Seite der Datenproduzenten. Die statistischen Ämter erproben diese Zugangsform derzeit in einem Pilotprojekt und mit Lösung der drängendsten Frage, nämlich der Finanzierung, ist dieses Verfahren absehbar auch einsatzbereit.

Eine weitere große Herausforderung für die informationelle Infrastruktur in Deutschland liegt darin, die verschiedensten Datenbestände für einzelne Analysefragestellungen zu integrieren – seien es amtliche oder private Daten, traditionelle oder nicht traditionelle Daten (beispielsweise Mobilfunksignale). Die Coronakrise hat vor allem gezeigt, dass häufig nicht die Daten fehlen, sondern sie nur an vielen verschiedenen Stellen und in oft unbekannter Qualität vorliegen. Was an dieser Stelle fehlt ist eine Institution, die als Datenkompetenzzentrum das Datenökosystem überschaubar macht und die Fähigkeit besitzt, als Datentreuhänder die verschiedenen Daten datenschutzkonform für einzelne Fragestellungen, vielleicht nur temporär, zusammenzuführen und gegebenenfalls zu analysieren (Hölscher und andere, 2022). Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder bringen für diese Herausforderungen eine mittlerweile langjährige Expertise mit; was ihnen hierzu fehlt, ist ein gesetzlicher Auftrag und die dazugehörige Finanzierung. 

LITERATURVERZEICHNIS

- Hahlen, Johann. *Amtliche Statistik als zentraler Teil der statistischen Infrastruktur. Anmerkungen zum Memorandum*. In: Allgemeines Statistisches Archiv. Band 82, 1998, Seite 387 ff.
- Hauser, Richard/Hochmuth, Uwe/Schwarze, Johannes. *Mikroanalytische Grundlagen der Gesellschaftspolitik*. 1994. Band 1 und 2, Berlin.
- Hauser, Richard/Wagner, Gert G./Zimmermann, Klaus F. *Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung*. In: Allgemeines Statistisches Archiv. Band 82, 1998, Seite 369 ff.
- Hlawatsch, Anja/Meyer, Karen/Rothe, Patrick. *20 Jahre Forschungsdatenzentrum der Statistischen Ämter der Länder*. In: Bayern in Zahlen. Jahrgang 153. Ausgabe 11/2022, Seite 25 ff. [Zugriff am 14. November 2022]. Verfügbar unter: [statistik.bayern.de](https://www.statistik.bayern.de)
- Hölscher, Michael/Hofacker, Angelina/Münstermann, Dirk/Netz, Gabi/Rosengart, Nathalie/Schüller, Katharina/Seidel, Gerald/Zwick, Markus. [*Warum wir ein nationales Datenkompetenzzentrum / Kompetenzzentrum für Data Literacy brauchen*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2022, Seite 73 ff.
- Krupp, Hans-Jürgen. [*Mikroanalysen und amtliche Statistik – gestern, heute, morgen*](#). In: Statistisches Bundesamt (Herausgeber). Mikroanalysen und amtliche Statistik. Band 1 der Schriftenreihe Statistik und Wissenschaft, Seite 27 ff. Wiesbaden 2004.
- KVI (Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik). *Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*. Baden-Baden 2001.
- Müller, Walter/Blien, Uwe/Knoche, Peter. *Die faktische Anonymität von Mikrodaten*. Band 19 der Schriftenreihe Forum der Bundesstatistik. Wiesbaden 1991.
- Orcutt, Guy H. *A new typ of socio-economic systems*. In: Review of Economics and Statistics. Jahrgang 39. Ausgabe 2/1957, Seite 116 ff.
- Rendtel, Ulrich/Seidel, Willi/Müller, Christine/Meinfelder, Florian/Wagner, Joachim/Chlumsky, Jürgen/Zwick, Markus. *Statistik zwischen Data Science, Artificial Intelligence und Big Data: Beiträge aus dem Kolloquium „Make Statistics great again“*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 16. Ausgabe 2/2022, Seite 97 ff. Verfügbar unter: doi.org/10.1007/s11943-022-00305-7
- Rolf, Gabriele/Zwick, Markus/Wagner, Gert G. (Herausgeber). *Fortschritte der informationellen Infrastruktur in Deutschland*. Baden-Baden 2008.
- Statistisches Bundesamt. *Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven*. Band 34 der Schriftenreihe Forum der Bundesstatistik. Wiesbaden 1999.

LITERATURVERZEICHNIS

Statistisches Bundesamt. *Empfehlungen des Statistischen Beirats für die 20. Legislaturperiode*. Pressemitteilung Nr. 034 vom 26. Januar 2022. [Zugriff am 15. November 2022]. Verfügbar unter: www.destatis.de

Wagner, Joachim. *25 Jahre Nutzung vertraulicher Firmenpaneldaten der amtlichen Statistik für wirtschaftswissenschaftliche Forschung: Produkte, Projekte, Probleme, Perspektiven*. In: AStA Wirtschafts- und Sozialstatistisches Archiv. Jahrgang 9. Ausgabe 2/2015, Seite 83 ff.

Zühlke, Sylvia/Zwick, Markus/Scharnhorst, Sebastian/Wende, Thomas. [Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder](#). In: *Wirtschaft und Statistik*. Ausgabe 10/2003, Seite 907 ff.

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Dezember 2022
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-22006-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2022
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.