

## Positionspapier des RatSWD zur Datenstrategie der Bundesregierung

Der Rat für Sozial und Wirtschaftsdaten (RatSWD) begrüßt die Initiative einer Datenstrategie der Bundesregierung nachdrücklich. Als institutionalisiertes Forum des Austausches zwischen Wissenschaft und Datenproduktion bringt sich der RatSWD gerne in die anstehenden Ausarbeitungen ein.

Die **Digitalisierung birgt große Chancen und Innovationspotenziale**, da neue Datenquellen entstehen und von Akteuren der Zivilgesellschaft, Verbänden, Wirtschaft, Wissenschaft und Verwaltung prinzipiell genutzt werden können: Neben wirtschaftlichen Chancen bieten die neuen Möglichkeiten der Datengewinnung und -nutzung auch der wissenschaftlichen Forschung immense Potenziale, um aktuelle gesellschaftliche Herausforderungen anzugehen. Durch wissenschaftliche Datennutzung wird die Informationsbasis für Gesellschaft und Politik verbessert. Außerdem können Politikmaßnahmen effizienter evaluiert und effektiver gestaltet werden. Wirtschaftliche und soziale Herausforderungen können verlässlicher identifiziert und öffentliche Mittel gezielter eingesetzt werden. Die aktive Rolle der Wissenschaft sollte daher in der Datenstrategie der Bundesregierung von Beginn an mitgedacht werden.

Bereits seit vielen Jahren engagiert sich der RatSWD für eine Infrastruktur, die der Wissenschaft einen flexiblen und datenschutzkonformen Datenzugang ermöglicht – ein wichtiger Schritt, um die Potenziale der Digitalisierung zu nutzen. In den letzten Jahren konnten wichtige Fortschritte erzielt werden. Dennoch lassen sich vier Problemfelder definieren, für die konkrete Handlungsbedarfe bestehen:

### 1) Schaffung von Datenzugängen

Der RatSWD setzt sich seit vielen Jahren für den verbesserten **Zugang zu amtlichen Mikrodaten für die unabhängige Wissenschaft** ein. Zuletzt hat sich der RatSWD intensiv für das kürzlich beschlossene **Digitale-Versorgung-Gesetz** und somit für die Einrichtung eines Forschungsdatenzentrums (FDZ) für Versorgungsdaten der gesetzlichen Krankenversicherungen (GKV) engagiert. Auch die Bereitstellung amtlicher Daten im Bereich Bildung und Migration forciert der RatSWD.<sup>1</sup> Ein Augenmerk liegt dabei stets auf den Besonderheiten der Anonymisierung und Pseudonymisierung von sensiblen, personenbeziehbaren Mikrodaten wie ihrer Langzeitarchivierung und Replizierbarkeit.

---

<sup>1</sup> Eine Zusammenfassung der Aktivitäten finden Sie unter: <https://www.ratswd.de/themen/datenzugang>

Noch immer sind entscheidende, bereits vorhandene Informationsquellen nur eingeschränkt oder überhaupt nicht für die wissenschaftliche Forschung zugänglich. Im Bereich der Verwaltungsdaten fehlt eine **integrierte Dateninfrastruktur**, sodass viele Daten ungenutzt in Daten-Silos liegen. Sie sind somit weder für die Verwaltung selbst, noch für die wissenschaftliche Forschung nutzbar.

Einige Bereiche staatlichen Handelns sind von den einschlägigen Registern und Statistiken ungenügend abgedeckt oder es fehlen klar geregelte Zugänge für die Forschung. Dies gilt insbesondere für Melderegister, Ausländerzentralregister, Bildungsregister, Mortalitätsregister, Steuerdaten, Transaktionsdaten im Immobilienbereich aber auch für die Kriminal- und Strafrechtspflegestatistik. Durch fehlende bzw. verbesserungsbedürftige Datenbereitstellung wird nicht nur datengestützte Forschung in Deutschland erschwert, auch für internationale Vergleiche sind die Daten aus Deutschland vielfach nur unzureichend nutzbar. Dies gilt beispielsweise in der 'Europäischen Gemeinschaftsstatistik über Einkommen und Lebensbedingungen' (EU-SILC – Paneldaten) und in den Kriminalitätserhebungen der Vereinten Nationen.

Die durch den RatSWD akkreditierten FDZ der amtlichen Statistik, z. B. die FDZ der Statistischen Ämter des Bundes und der Länder, zeigen bereits heute, dass staatliche Akteure die Bereitstellung teilweise sensibler Mikrodaten routiniert umsetzen und somit **den Staat zum Vorreiter in der Bereitstellung von Daten machen können**. Das **FDZ-Modell** kann für viele weitere Behörden ein Weg sein, amtliche Mikrodaten für die wissenschaftliche Forschung unter strenger Einhaltung des Datenschutzes zur Verfügung zu stellen. Die Erfolgsgeschichte der Vernetzung von inzwischen 34 vom RatSWD akkreditierten FDZ, die aktuell 3.940 Datensätze anbieten,<sup>2</sup> kann hier als Vorbild dienen. Dies gilt in gleicher Weise für Modelle projektbezogener Anschubfinanzierungen durch das BMBF für einen raschen Aufbau eines FDZ.

Wichtig wären die Gründung (a) eines FDZ im Bundeskriminalamt, welches die polizeiliche Kriminalstatistik für die Forschung zur Verfügung stellt und (b) eines FDZ beim Bundesfinanzministerium für eine zentrale Aufbereitung der Steuerdaten der Landesfinanzbehörden.<sup>3</sup> Dabei benötigen FDZ ausreichend personelle Ressourcen und technische Infrastruktur, um ihre Arbeit effektiv auszuführen. Hier müsste gerade (c) das Statistische Bundesamt, das hervorragende Arbeit leistet, durch entsprechende Stellen und Mittel und einen expliziten gesetzlichen Forschungs- und Serviceauftrag gestärkt werden und

---

<sup>2</sup> Ein Überblick über die Fortschritte dieser Infrastruktur gibt der aktuelle Tätigkeitsbericht: RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019): Tätigkeitsbericht 2018 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.40>. Einen Gesamtüberblick bieten: Bug, Mathias; Liebig, Stefan; Oellers, Claudia; Riphahn, Regina T. (2018): Operative und strategische Elemente einer leistungsfähigen Forschungsdateninfrastruktur in den Sozial- und Wirtschaftswissenschaften. Under Debate; In: Journal of Economics and Statistics 2018; 238(6): 571–590; De Gruyter; Oldenburg; <https://doi.org/10.1515/jbnst-2018-0029>.

<sup>3</sup> Hierzu bedürfte es einer über § 21 Abs. 6 FVG hinausgehende Pflicht zur Weitergabe von Informationen seitens der Landesfinanzbehörden an den Bund. Eine Rechtsgrundlage für die Verknüpfung mit anderen Forschungsdatenzentren könnte im Gesetz über Steuerstatistiken aufgenommen werden.

auch in die Lage versetzt werden, selbst methodische wie technologische Innovationen im Bereich Datenbereitstellung vorzunehmen.

Eine gewichtige Limitation der amtlichen Statistik besteht darin, dass Deutschland im internationalen Vergleich in der Entwicklung eines **registerbasierten Zensus** hinterherhinkt. Speziell die Verbesserung des Identitätsmanagements im Zuge der Registermodernisierung ist notwendig, um die Daten besser nachnutzbar und verknüpfbar zu machen. Bei der Planung und Umsetzung eines registerbasierten Zensus kann die Wissenschaft wichtige Impulse geben und sollte eingebunden werden. Auch die Bereitstellung der so generierten Zensusdaten für die Wissenschaft ist frühzeitig einzuplanen.

In der Breite noch ungelöst ist der Zugang für die Wissenschaft zu „**kommerziellen**“ **Daten** – z.B. von Unternehmen der New Economy. Ein solcher Datenzugang ist aktuell meist von persönlichen Kontakten abhängig. Hier sollten Wege geschaffen werden, um geregelte Zugänge zu ermöglichen – ohne die kommerziellen Interessen der Dateneigner zu beschneiden. Dabei könnten die Einrichtung eines **Datenzentrums für Digitale Unternehmensdaten** und/oder die **Schaffung einer Treuhandstelle** Lösungen sein. Der RatSWD hat bereits einen Rahmen für eine Big Data Treuhandstelle ausgearbeitet.<sup>4</sup> Dieser Rahmen bietet Ideen sowie mögliche Referenzlösungen für die weitere Ausgestaltung des von der Bundesregierung angedachten Datentreuhänders.

## 2) Technische Hürden: Cloud-Lösungen und Remote Access

Durch die Digitalisierung entstehen neue und große Datenmengen. Damit diese auswertbar werden, sind europaweite **Cloud-Lösungen** nötig. Mit GAIA-X und der European Open Science Cloud (EOSC) werden hierzu momentan Schritte in die richtige Richtung getätigt, an welche die Datenstrategie der Bundesregierung anknüpfen sollte. Datenanbieter wie die Statistischen Ämter, die Deutsche Rentenversicherung, die Bundesagentur für Arbeit, die Bundesbank und viele mehr, könnten durch solche Initiativen ihre Daten sicher speichern und die Nutzbarkeit sicherstellen.

Für amtliche Daten, die grundsätzlich bereits zugänglich sind, bleiben in Deutschland die technischen Möglichkeiten des Zugangs hinter den Standards in vielen anderen EU-Mitgliedstaaten zurück und limitieren so ihre Verwendung und Innovationskraft: Zurzeit ist ein Datenzugang mehrheitlich vor Ort oder durch kontrollierte Datenfernverarbeitung möglich. Diese Optionen erfordern hohen zeitlichen und organisatorischen Aufwand auf Seiten der Datennutzenden wie auch auf Seiten der Datenanbietenden. Der RatSWD setzt sich daher für die Schaffung flexibler Datenzugänge ein. Große Potenziale bieten dabei sogenannte **Remote Access-Lösungen**.<sup>5</sup> Über Remote Access kann ein sicherer Fernzugriff

---

<sup>4</sup> RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019): Big Data in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Datenzugang und Forschungsdatenmanagement. RatSWD Output 4 (6). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.39>.

<sup>5</sup> RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019): Remote Access zu Daten der amtlichen Statistik und der Sozialversicherungsträger. RatSWD Output 5 (6). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.42>.

selbst auf sensible Einzeldaten ermöglicht werden. Vorbilder sind die statistischen Ämter der nordeuropäischen Staaten und anderer europäischer Staaten, die schon seit längerer Zeit die rechtlichen und technischen Voraussetzungen hierfür geschaffen haben. Die Datenstrategie der Bundesregierung sollte in diesem Zusammenhang ebenfalls Weichen stellen. Für Daten der amtlichen Statistik sind hierfür Gesetzesänderungen notwendig. Speziell sollte eine Anpassung des § 16 Abs. 6 Bundesstatistikgesetzes forciert werden, um Zugang über Remote Desktop zu formal anonymisierten Einzeldaten der amtlichen Statistik zu ermöglichen. Die Wissenschaft sollte in die Planung von zeitnahen Pilotprojekten eingebunden werden.

### 3) Datenqualität sichern

Wissenschaft und Verwaltung benötigen qualitativ hochwertige und detaillierte Daten. Nur so kann ein verlässlicher Erkenntnisgewinn realisiert werden. Der Charakter der Datengewinnung hat sich jedoch in den letzten Jahren gewandelt. Vermehrt werden Forschungsdaten **mit neuen Informationstechnologien,<sup>6</sup> aus Big Data<sup>7</sup> oder mittels Online-Studien erhoben**. Diese Daten haben große Potenziale für die Beantwortung bestehender und neuer Forschungsfragen. Ihre Qualität genügt jedoch nicht immer wissenschaftlichen Standards. Die unregelmäßige Nutzung und Weiterverbreitung solcher Daten birgt dabei nicht nur Gefahren für das Vertrauen in die Wissenschaft, sondern kann auch für Gesellschaft und Politik schädlich sein.

Die methodische und technische Qualität und damit der wissenschaftliche Wert neuer Daten sollten in der Datenstrategie der Bundesregierung Beachtung finden.<sup>8</sup> Zu überlegen ist z.B. die Einführung eines **Datenqualitätssiegels**. Die Bedeutung von Datenqualität kann auch durch die gezielte Förderung von Forschung zur Datenqualität aufgewertet werden.

Neben der Qualitätssicherung von automatisierten Datenerhebungen sollte die Strategie der Bundesregierung sicherstellen, dass bereitgestellte Verwaltungsdaten etablierten Qualitätsstandards entsprechen. So sollte sich die Bundesregierung in ihren Bemühungen um die Verbesserung von **Open Data Initiativen** nicht zuletzt auf die pragmatischen Grundsätze einer nachhaltig nachnutzbaren Forschungsdateninfrastruktur gemäß den international anerkannten **FAIR-Kriterien<sup>9</sup>** berufen. Die Daten der öffentlichen Verwaltung sollten folglich nicht nur maschinenlesbar bereitgestellt werden, sondern auch durch **entsprechende**

---

<sup>6</sup> Eine spezielle Handreichung zu diesem Feld veröffentlicht der RatSWD in Kürze.

<sup>7</sup> RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2019): Big Data in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Datenzugang und Forschungsdatenmanagement. RatSWD Output 4 (6). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.39>.

<sup>8</sup> S. dazu auch: RfII [Rat für Informationsinfrastrukturen] (2019): Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel. Göttingen, Rat für Informationsinfrastrukturen. <http://www.rfii.de/?p=4043>.

<sup>9</sup> Die "FAIR Data Principles" formulieren Grundsätze, die nachhaltig nachnutzbare Forschungsdaten erfüllen müssen und die Forschungsdateninfrastrukturen dementsprechend im Rahmen der von ihnen angebotenen Services implementieren sollten. Gemäß der FAIR-Prinzipien sollen Daten "Findable, Accessible, Interoperable, and Reusable" sein. In diesem Kontext sind auch Fragen der Langzeitarchivierung sowie Replizierbarkeit von Forschungsdaten eine notwendige Voraussetzung.

**einheitliche Metadaten** einfacher recherchierbar werden.<sup>10</sup> Die Verbesserung des Datenangebots auf GovData, die Betriebsaufnahme der Kompetenzzentrum Open Data, aber auch die Unterstützung der kommunalen Verwaltung etwa durch datenbezogene Schulungen (sowie der Entwicklung von Curricula für Digitale Bildung) und Ressourcen wären Maßnahmen, die umgesetzt werden sollten.

#### 4) **Der Kultur des Misstrauens entgegenzutreten und Daten-Verknüpfungen ermöglichen**

Grundsätzlich ist Wissenschaft auf der Suche nach Erkenntnis, die dazu dient, einen Mehrwert für die Gesellschaft zu erschließen. Dies geschieht etwa durch die Analyse von gruppenbezogenen Korrelationen, Mustern und sozialen Regelmäßigkeiten. Ziel wissenschaftlicher Arbeit ist nie die Re-identifizierung von Individuen. Der wissenschaftliche Erkenntnisgewinn bereichert Politik und Verwaltung auch bzw. gerade, wenn hierdurch politische Handlungsbedarfe aufgedeckt und Verbesserungsmöglichkeiten identifiziert werden.

Die empirische wissenschaftliche Arbeit wird durch Beschränkungen bei Datenzugang und -nutzung erheblich erschwert. Nicht selten spielt hier ein in Deutschland besonders intensives Datenschutz-Misstrauen eine Rolle. Die Furcht vor Datenmissbrauch begrenzt zurzeit nicht nur den Datenzugang,<sup>11</sup> auch der Rahmen für **Datenverknüpfungen ist oft so restriktiv**, dass innovative Forschung nicht oder nur durch teure und aufwändige Doppelerhebungen durchgeführt werden kann. Das Bundesstatistikgesetz erlaubt beispielsweise nicht das projektunabhängige Zusammenspielen von Unternehmensdaten des Statistischen Bundesamtes mit Mikrodaten des Instituts für Arbeitsmarkt und Berufsforschung der Bundesagentur für Arbeit. Dieser Umstand be- und verhindert teilweise die Beantwortung gesamtgesellschaftlich wichtiger Forschungsfragen.

Der RatSWD adressiert dieses Datenschutz-Misstrauen seit vielen Jahren, z.B. indem er sich bei der Etablierung und Konsolidierung von **forschungsethischen und datenschutzrechtlichen Standards** als Ansprechpartner und Impulsgeber etabliert hat und **Datenkompetenzen** auf Seite der Datennutzenden und der Datenproduzenten forciert.<sup>12</sup> Er fordert dabei nicht nur Behörden und kommerzielle Akteure zum Teilen ihrer Daten auf, sondern appelliert auch an die Forschung selbst, ihre Daten in Repositorien zur Nachnutzung zur Verfügung zu stellen.

---

<sup>10</sup> Dies betrifft beispielsweise Daten wie die amtlichen Lärmkartierung, die nach europäischem Recht zwar publiziert werden müssen, in Deutschland allerdings in unterschiedlichsten Formaten, mit unterschiedlichen Lizenzen etc. bereitgestellt werden, was die Nachnutzung erschwert.

<sup>11</sup> Beispielsweise sind die Paneldaten der Europäischen Gemeinschaftsstatistik über Einkommen und Lebensbedingungen aus Datenschutzgründen in Deutschland für die Forschung nicht zugänglich und die Bereitstellung der Daten des Mikrozensus als Scientific Use File ist erheblich verzögert.

<sup>12</sup> Zu nennen sind hier die Handreichung Datenschutz (RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2017): Handreichung Datenschutz. RatSWD Output 5 (5). Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.6>) und Materialien zur Forschungs(daten)ethik <https://www.ratswd.de/themen/forschungsethik>).

Die Datenstrategie der Bundesregierung sollte Rahmenbedingungen setzen, um den Datenkulturwandel zum Teilen von Daten weiter zu befördern. Konkrete Maßnahmen wären die **Förderung von Repositorien**, aber auch eine **Erweiterung der Verwaltungs- und Rechtsgrundlagen für Datenverknüpfungen für wissenschaftliche Zwecke**. Auch die **Einführung eines Forschungsgeheimnisses** für unabhängige Forschung könnte einige Beschränkungen beheben. Dem Prinzip der Datensparsamkeit im Sinne des Datenschutzes würde durch solche Privilegien Rechnung getragen werden, da Doppelerhebungen vermieden werden. Nicht zuletzt würden durch Verknüpfungen sowohl bei Datennutzenden als auch bei Datenproduzenten Zeit und Ressourcen gespart und die Integrität der Datenbestände erhalten werden.

Bei dem angedachten Datenkulturwandel sollte Deutschland auch auf Entwicklungen in Europa aufbauen. Auf der Basis der Europäischen Datenschutz-Grundverordnung (DS-GVO) schlug Finnland beispielsweise vor, den Datenzugang der Forschung länderübergreifend zu harmonisieren. Ziel ist ein transnationaler Zugang innerhalb eines Länderverbundes.<sup>13</sup> Ein **Europäischer Datenraum** könnte langfristig den Zugang der Forschung zu staatlichen und möglichst auch kommerziellen Daten regeln und so zur Stärkung europaweiter Forschungsverbände und zu mehr international vergleichender Forschung beitragen. Hierdurch würden neue Potenziale für Wertschöpfung, Innovation und evidenzbasierte politische Entscheidungen erschlossen.

Zusammenfassend begrüßt der RatSWD die Initiative einer Datenstrategie der Bunderegierung. Er wird gerne **zu ihrem Gelingen aktiv beitragen** und umfassend mitarbeiten, wenn es etwa um die Schaffung der rechtlichen Rahmenbedingungen, die Identifikation von Datenquellen, die Sicherung der Datenqualität, die kreative Ausgestaltung von Datenzugängen und die notwendige Anonymisierung bzw. Pseudonymisierung der Daten bis hin zur technischen Ausgestaltung und Entwicklung der notwendigen Infrastruktur geht. Auch die im Entstehen begriffene Nationale Forschungsdateninfrastruktur (NFDI) kann hierfür ein Dach bieten.

Wir freuen uns auf die Zusammenarbeit.

---

<sup>13</sup> Dadurch können Forschende z.B. an einer dänischen Universität an ihrem Arbeitsplatz nicht nur mit dänischen, sondern auch mit norwegischen, schwedischen und finnischen amtlichen Daten arbeiten.

### Der Rat für Sozial- und Wirtschaftsdaten (RatSWD)

berät seit 2004 die Bundesregierung und die Regierungen der Länder in Fragen der Forschungsdateninfrastruktur für die empirischen Sozial-, Verhaltens- und Wirtschaftswissenschaften. Im RatSWD arbeiten acht durch Wahl legitimierte Vertreterinnen und Vertreter der sozial-, verhaltens- und wirtschaftswissenschaftlichen Fachdisziplinen mit acht Vertreterinnen und Vertretern der wichtigsten Datenproduzenten zusammen. Er versteht sich als institutionalisiertes Forum des Dialoges zwischen Wissenschaft und Datenproduzenten und erarbeitet Empfehlungen und Stellungnahmen. Der RatSWD engagiert sich für eine Infrastruktur, die der Wissenschaft einen breiten, flexiblen und sicheren Datenzugang ermöglicht. Solche Daten werden von staatlichen, wissenschaftsgetragenen und privatwirtschaftlichen Akteuren bereitgestellt. Der RatSWD hat 34 Forschungsdatenzentren akkreditiert, deren Kooperationen er fördert.