

FÜR EINEN BESSEREN DATENZUGANG

STELLUNGNAHME DER ARBEITSGRUPPE
UNTERNEHMENS DATEN

Verein für Socialpolitik

FÜR EINEN BESSEREN DATENZUGANG

STELLUNGNAHME DER ARBEITSGRUPPE

UNTERNEHMENS DATEN

Mitglieder der UAG Unternehmensdaten sind (in alphabetischer Reihenfolge):

Sandra Gottschalk, Carla Krolage, Georg Licht, Andreas Peichl, Sandra Schaffner, Sebastian Wichert*

Impressum

Verein für Socialpolitik

Mohrenstraße 59, 10117 Berlin

Layout: Caroline Tomanik | immo-grafik.at

September 2023

* Sebastian Wichert, Head of LMU-ifo Economics & Business Data Center (EBDC), wichert@ifo.de

INHALTSVERZEICHNIS

01	ZUSAMMENFASSUNG DER ZENTRALEN FORDERUNGEN ZUR VERBESSERUNG DES FORSCHUNGSZUGANGS ZU UNTERNEHMENS DATEN	4
02	EINLEITUNG	5
03	STATUS QUO - UNTERNEHMENS DATEN IN DEUTSCHLAND UND DEFIZITE BEIM FORSCHUNGSZUGANG	7
04	ZUGANGSBEDINGUNGEN IM INTERNATIONALEN VERGLEICH	10
05	EMPFEHLUNGEN	11

1. ZUSAMMENFASSUNG DER ZENTRALEN FORDERUNGEN ZUR VERBESSERUNG DES FORSCHUNGSZUGANGS ZU UNTERNEHMENSDATEN:

1. Bessere Nutzungsmöglichkeiten und vereinfachte Verknüpfung von amtlichen (Unternehmensdaten-) Daten untereinander sowie mit privaten Daten durch Ausweitung der gesetzlichen Regelungen im Rahmen eines Forschungsdatengesetzes

- a. Wissenschaftliche Forschung als Verknüpfungszweck
- b. Beweislastumkehr bei der Datenverknüpfung
- c. Schaffung einer Rechtsgrundlage für die Verknüpfung von Unternehmens- mit Personendaten
- d. Einführung einer einheitlichen Unternehmens-ID zur Verknüpfung von Datensätzen
- e. Schaffung einer Rechtsgrundlage für die Verknüpfung von Unternehmens- mit Befragungsdaten
- f. Langfristige sichere Speicherung von Identifikatoren und Schutz vor Löschung

2. Verbesserung der Nutzbarkeit von (bisher nur) gewerblichen Unternehmensdaten

- a. Forschungstauglicher Massendatenzugang zu amtlichen Registern
- b. Vereinfachte Nutzung bestehender Unternehmensdaten durch gemeinsame Verhandlungsführung
- c. Stärkere Berücksichtigung von Datenkosten in Förderprogrammen
- d. Aufbau einer institutsübergreifenden Unternehmensdaten-Infrastruktur

3. Kapazitätsausbau der Forschungsdatenzentren und zeitnahe Bereitstellung von Daten

- a. Ausbau der Personalkapazitäten
- b. Ausbau der IT-Infrastruktur
- c. Zeitnahe Bereitstellung von Daten
- d. Klare Zuständigkeiten, Harmonisierung und Straffung der Prozesse in den Statistischen Ämtern
- e. Eigener Forschungsauftrag für das Statistische Bundesamt und die Statistischen Landesämter

4. Ausweitung des Datenzugriffs per Remote Access

2. EINLEITUNG

Wie reagieren Unternehmen auf Steuererhöhungen oder Veränderungen beim Mindestlohn? Wo in Deutschland werden welche Unternehmen gegründet oder geschlossen? Welche Regional- und Wirtschaftsfördermaßnahmen sind effektiv und warum? Wie reagieren Firmen auf den zunehmenden Fachkräftemangel, Lieferkettenprobleme oder die steigende Inflation? Zuverlässige, zeitnahe Antworten auf diese (und viele weitere) wirtschaftspolitisch relevante Fragen zur wirtschaftlichen Lage von Unternehmen und damit der gesamten deutschen Volkswirtschaft sind fundamental wichtig und das nicht nur in Krisenzeiten. Nur so können private Haushalte und Unternehmen informierte ökonomische Entscheidungen treffen und die Politik geeignete Rahmenbedingungen setzen und Gesetze beschließen.

Wirtschaftsforscher*innen wiederum können diese und weitere Fragen nur evidenzbasiert analysieren und zukünftige Entwicklungen solide prognostizieren, wenn ihnen ausreichend qualitativ hochwertige Unternehmensdaten zeitnah zur Verfügung stehen. Gute Politikberatung und der Transfer in die Öffentlichkeit bedingt aber auch guten Datenzugang für die wirtschaftswissenschaftliche Spitzenforschung² und ist damit für den Wissenschaftsstandort Deutschland fundamental: Die besten Köpfe suchen sich ein geeignetes wissenschaftliches Umfeld, in dem sie ihre innovativen Forschungsideen am besten (mit geeigneten Daten) umsetzen können. Das betrifft die Wahl ihres eigenen permanenten Forschungsumfeldes und die Auswahl ihrer (inter-)nationalen Kooperationspartner*innen. Zudem untersuchen (internationale) Spitzenforscher*innen natürlich eher Unternehmen und Maßnahmen in den Ländern, in denen gute Daten zur Verfügung stehen. D.h. nur, wenn diese Voraussetzungen erfüllt sind, ist auch die bestmögliche wissenschaftliche Politikberatung für Deutschland möglich.³

Eine Umfrage mit fast 500 Teilnehmer*innen unter den Mitgliedern des Vereins für Socialpolitik, der größte Zusammenschluss von akademisch arbeitenden Wirtschaftswissenschaftler*innen im deutschsprachigen Raum, im Frühjahr 2023 ergab allerdings Defizite beim (Forschungs-)Zugang zu Unternehmensdaten in Deutschland, denn:

- 66% der Befragten sind eher unzufrieden bis sehr unzufrieden mit den Datenzugangsmöglichkeiten in Deutschland. Unter denen, die Unternehmensdaten nutzen oder bereits genutzt haben, ist der Anteil mit 68% sogar noch etwas höher.
- Mehr als 70% derer, die mit Unternehmensdaten arbeiten, geben an, dass der Forschungserfolg unter den schlechten Datenzugangsmöglichkeiten leide und Forschungsprojekte nicht durchgeführt werden konnten.
- Die Mehrheit der Forschenden hat auch schon mit ausländischen Daten geforscht, wobei mehr als die Hälfte der Forschenden, die Unternehmensdaten nutzen, angibt, dass einer der Gründe auch war, dass geeignete deutsche Daten nicht verfügbar waren.

Unbestritten ist, dass sich auch in Deutschland in den letzten Jahrzehnten der Zugang zu Forschungsdaten, auch zu Unternehmensdaten, verbessert hat. Insbesondere sind hier die immer weiter steigende Anzahl von akkreditierten und qualitätsgesicherten Forschungsdatenzentren (FDZ) an staatlichen Behörden und Forschungseinrichtungen, die Gründung der Nationalen Forschungsdateninfrastruktur (NFDI), die Aktivitäten des Rats für Sozial- und Wirtschaftsdaten (RatSWD) und die strenger werdenden Verpflichtungen von Forschungsförderorganisationen zur Datenarchivierung und Datenveröffentlichung zu nennen.⁴ Zudem treten vermehrt kommerzielle (Unternehmens-)Datenhändler auf den Plan.

Auch die Bundesregierung hat die dringende Notwendigkeit eines besseren Zugangs zu Forschungsdaten erkannt und in ihren im Januar 2021 und im August 2023 veröffentlichten Datenstrategien⁵ zahlreiche relevante Gesetzesvorhaben und Maßnahmen angekündigt.⁶

Die sehr zu begrüßenden Absichtserklärungen bedürfen jedoch noch der konkreten Umsetzung im politischen Gesetzgebungsprozess, damit sich die Lage für die Forschung verbessern kann. Erschwerend kommt hinzu, dass es in einigen Interessengruppen außerhalb der akademischen Forschung, wie z.B. bei kommerziellen Datenhändlern, Datenschutzinitiativen, und z.T. Behörden oder

² Mehrere Untersuchungen zeigen, dass besserer Datenzugang zu qualitativ besseren Forschungspublikationen führt (siehe u.a. Nagaraj, Abhishek, und Matteo Trancheo. "How Does Data Access Shape Science? Evidence from the Impact of U.S. Census's Research Data Centers on Economics Research." NBER Working Paper Series, Nr. 31372, Juni 2023. doi: 10.3386/w31372.)

³ Der Fakt, dass die skandinavischen Länder seit längerer Zeit sehr forschungsfreundliche Datenzugangsregeln haben, führt eben dazu, dass internationale Forscher*innen ihre Forschungsvorhaben dort umsetzen. Forschungspapiere über diese Länder werden zudem häufig sehr gut publiziert und das, obwohl die dort erzielten Ergebnisse aufgrund der relativ geringen Größe und wirtschaftlichen Bedeutung dieser Länder oftmals nur bedingt auf andere Länder übertragbar sind.

⁴ siehe <https://www.konsortswd.de/datenzentren/>; <https://www.nfdi.de/>; <https://www.konsortswd.de/ratswd/> und https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf [zuletzt aufgerufen am 08.09.2023].

⁵ <https://www.bundesregierung.de/breg-de/suche/datenstrategie-der-bundesregierung-1845632> und https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2023/datenstrategie.html;jsessionid=BB0F62822D24F3AA03F6CF135A8418901_cid287

⁶ Hinzu kommen zahlreiche laufende europäische Gesetzgebungsinitiativen bzw. deren Umsetzung in nationales Recht, wie z.B. der Data Act, Data Governance Act, Data Service Act, AI Act,..!

den Unternehmen selbst, aus verschiedensten Gründen Vorbehalte gegen eine Erweiterung des Datenzugangs für die akademische Forschung gibt.

Gerade in dieser Phase der Veränderung, bevor eine Vielzahl an Gesetzen (endlich) geändert und eingeführt werden, müssen Forscher*innen hier und jetzt ihre spezifischen Bedarfe klar artikulieren und aktiv Verbesserungsvorschläge einbringen. Dieses Positionspapier soll einen Beitrag im Bereich Unternehmensdaten dazu leisten.⁷ In Kapitel 2 wird zunächst beispielhaft erklärt, was Unternehmensdaten sind, insbesondere welche Granularität sie haben (müssen), wer sie wie bereitstellt und warum sie zwar zugänglich, aber dennoch oftmals geschützt sein müssen. Das Kapitel schließt mit einer Zusammenfassung der Defizite beim Zugang zu Unternehmensdaten in Deutschland. Kapitel 3 gibt dann eine Übersicht, wie (besserer) Forschungszugang zu (Unternehmens-)Daten in anderen Ländern organisiert ist, bevor in Kapitel 4 konkrete organisatorische, technische und rechtliche Verbesserungsvorschläge zum Datenzugang in Deutschland gegeben werden.

⁷ Das vorliegende Positionspapier ist Teil einer Serie von Positionspapieren von UAGs des VFS, die sich unterschiedlichen Datentypen widmen (u.a. Gesundheitsdaten und Arbeitsmarkt- und Sozialdaten) und die sich notwendigerweise inhaltlich an manchen Stellen überschneiden werden.

3. STATUS QUO - UNTERNEHMENS DATEN IN DEUTSCHLAND UND DEFIZITE BEIM FORSCHUNGSZUGANG

Unternehmensdaten sind für viele Subdisziplinen der Wirtschaftswissenschaften fundamental und es existieren verschiedene Typen von Unternehmensdaten. Diese können - wie alle anderen Datenarten - auf verschiedene Weise nach umfangreichen Metadaten(-schemata) klassifiziert werden: nach Quelle, Erhebungsweg, Inhalt, Periodizität, Aktualität, Qualität, Sensitivität, Nutzungsbedingungen etc. Alle Punkte einzeln abzuhandeln, sprengt den Rahmen und Sinn dieses Papiers und daher wird hier nur auf einige wenige Aspekte eingegangen.

Für viele wirtschaftspolitisch relevante Forschungsfragen sind möglichst viele, detaillierte Informationen über Unternehmen, deren Geschäftstätigkeit und Kund*innen und Mitarbeiter*innen notwendig. Mit Hilfe mikroökonomischer Analysemethoden ermöglichen es diese sogenannten Mikrodaten, also umfangreiche Einzelangaben zu Unternehmen, u.a. Treiber von makroökonomischen Größen, wie Inflation oder Arbeitslosigkeit, genauer zu untersuchen. Darüber hinaus werden Unternehmensdaten, die auf Branchenebene oder nach Regionen aggregiert sind, oftmals durch die amtliche Statistik regelmäßig in Fachserien veröffentlicht. Allerdings ist der Nutzen der aggregierten Statistik für viele Fragestellungen beschränkt, da Muster und Zusammenhänge im Aggregat nicht erkennbar sind und Verknüpfungen von verschiedenen Datensätzen mit unterschiedlichen Informationen auf Unternehmensebene zur weiteren Steigerung des Analysepotentials unmöglich werden. Wichtig dabei ist jedoch zu verstehen, dass in der volkswirtschaftlichen Forschung, obwohl Einzeldaten genutzt werden müssen, das individuelle Unternehmen nicht im Zentrum des Analyseinteresses steht. Mit Hilfe dieser sehr granularen Daten werden generelle Zusammenhänge untersucht und keine sensiblen Informationen über einzelne Unternehmen veröffentlicht, sondern immer nur aggregierte Statistiken wie Mittelwerte, Perzentile einer Verteilung oder Regressionskoeffizienten. Natürlich ist sich die Forschung der damit einhergehenden Verantwortung bewusst und erkennt auch das Bedürfnis und Recht von Unternehmen an, dass deren Daten geschützt werden müssen, um Geschäftsgeheimnisse zu wahren und die Wettbewerbsfähigkeit nicht zu gefährden als auch Persönlichkeitsrechte der damit verbundenen Personen sicherzustellen. Zudem sollen Unternehmen möglichst nicht mit zusätzlichen zeitaufwendigen verpflichtenden Befragungen und statisti-

schen Datenerhebungen bürokratisch belastet werden. Aus diesem Grund strebt die Wissenschaft möglichst guten Datenzugang nach den sogenannten FAIR-Prinzipien⁸ an. Grob gesagt besagen diese, dass Forschungsdaten auffindbar, zugänglich, interoperable und nachnutzbar sein sollten. Neben guter Dokumentation und dauerhafter Verfügbarkeit der Daten bedeutet dies u.a., dass die Zugangswege und Nutzungsbedingungen transparent und diskriminierungsfrei kommuniziert werden. Im Falle von Unternehmensdaten kann das eben bedeuten, dass detaillierteste Daten nur für rein-wissenschaftliche, nicht-kommerzielle Forschung unter strengen vertraglichen und technischen Sicherheitsmaßnahmen verfügbar sind. Dies steht im Gegensatz zum "Open Data"-Konzept⁹, nach dem Daten für jedermann frei zugänglich (im Internet) verfügbar sein müssen. Oftmals ist der einzige rechtlich zulässige Weg, diesen absoluten offenen Zugang für die Allgemeinheit herzustellen, die Unternehmensdaten vollständig zu anonymisieren und stark zu aggregieren. Damit ist aber wieder das Analysepotential massiv eingeschränkt und eine Verknüpfung mit anderen Datensätzen ist nicht möglich.

Grob lassen sich zwei Arten von Unternehmensdaten unterscheiden¹⁰: "Datensätze von Unternehmen", die Firmen bzw. deren Kund*innen, Nutzer*innen und Mitarbeiter*innen in ihren Arbeits- und Geschäftsprozessen (nebenbei) erzeugen und "Daten über Unternehmen".

Beispiele für erstere sind Nutzungsdaten von sozialen Medien (z.B. X (vormals Twitter), Facebook,...) oder Transaktionsdaten von Finanzdienstleistern oder Geschäften sowie Mobilitätsdaten, Telekommunikationsdaten oder Personaldaten aus HR-Prozessen. Der Forschungszugang zu diesen aus Forschungssicht sehr vielversprechenden Daten beruht in der Regel meist auf projektspezifischen Unternehmenskooperationen, die in Deutschland aufgrund des hohen administrativen Aufwands und der rechtlichen Unsicherheit noch relativ selten sind.¹¹ Zudem werden diese Daten (in anonymisierter oder aufbereiteter Form) mehr und mehr von den Unternehmen selbst oder kommerziellen Datenhändlern zum Kauf angeboten - oftmals für in der Forschung prohibitiv hohe Summen und mit restriktiven Lizenzbedingungen. Positive Ausnahme in Deutschland sind die RWI-GEO-RED-Daten¹², die von der Immobilienplattform ImmobilienScout24 kom-

⁸ siehe <https://www.go-fair.org/fair-principles/> [zuletzt aufgerufen am 08.09.2023].

⁹ siehe <https://opendefinition.org/od/2.1/en> [zuletzt aufgerufen am 08.09.2023].

¹⁰ Überlappungen sind natürlich dennoch möglich.

¹¹ Leider bietet der gerade beschlossene EU Data Act auch keine Rechtsgrundlage für einen verpflichtenden Forschungszugang zu diesen Unternehmensdaten außer in sehr eng definierten Notlagen. Internationale, groß angelegte Paradebeispiele für solche Kooperationen sind der Opportunity Insights Economic Tracker, in dem makroökonomische Entwicklungen in den USA mit privatwirtschaftlichen Daten fast in Realtime dargestellt werden können (<https://www.tracktherecovery.org/>) oder der Social Connectedness Index, der mit Hilfe von Facebook-Daten die weltweite gesellschaftliche und wirtschaftliche Vernetzung misst (<https://dataforgood.facebook.com/dfg/tools/social-connectedness-index>). Oftmals arbeiten zudem Forscher*innen im Rahmen von Informations-/Feldexperimenten mit den Unternehmen zusammen und untersuchen die Wirkung von geänderten Maßnahmen oder Produkten auf das Mitarbeiter-/Kundenverhalten.

¹² siehe <https://www.rwi-essen.de/forschung-beratung/weitere/forschungsdatenzentrum-ruhr/datenangebot/rwi-geo-red-real-estate-data> [zuletzt aufgerufen am 08.09.2023].

men und der wissenschaftlichen Forschung kostenfrei zur Verfügung gestellt werden dürfen. Um bessere Zugangswege für die Wissenschaft für solche Daten "von Unternehmen" auszuloten, wurden mit unterschiedlichen Schwerpunktsetzungen die NFDI Sektion "Industry Engagement", die Stifterverbandsinitiative "Datagroup Business 2 Science" und das Projekt "Zugang zu Firmendaten für die Wirtschaftswissenschaften" des NFDI-Konsortiums "KonsortSWD" ins Leben gerufen.¹³ Zwar bedarf es auch hier mehr Möglichkeiten für Kooperationen, damit die deutsche Forschung mit der internationalen Forschung (insbesondere für die USA) mithalten kann. Da diese Kooperationen aber immer auf Freiwilligkeit beruhen und ein Vertragsverhältnis zwischen Unternehmen und Forschenden abgeschlossen wird, sind die Möglichkeiten für eine gesetzliche Regulierung begrenzt. Einen gesetzlichen Zwang, dass alle Unternehmen gleichermaßen alle ihre Daten der Wissenschaft kostenfrei zur Verfügung stellen müssen, wird und sollte es nicht geben.¹⁴ Daher, auch wenn Daten aus Unternehmensprozessen eine immer größere Wichtigkeit erlangen, werden sie in diesem Dokument mit Verweis auf die laufenden Initiativen nicht weiter behandelt.

"Daten über Unternehmen" stellen die bisher in der Forschung weitaus häufiger genutzten Daten dar. Im Folgenden werden einige häufig genutzte Datenquellen aufgeführt:

- Viele Unternehmensdaten fallen im gesetzlich festgelegten Rahmen der amtlichen Statistik an. Bei den Forschungsdatenzentren der statistischen Ämter des Bundes und der Länder gibt es beispielsweise aktuell das AFiD Panel Unternehmensregister (aktuell nur verfügbar bis zum Jahr 2019), das vor allem allgemeine Informationen zu Sitz, Wirtschaftszweig, Beschäftigten und Umsatz beinhaltet. Darauf aufbauend gibt es noch tiefergehende Teilerhebungen wie das AFiD-Panel Industrieunternehmen (bis 2018) Industriebetriebe (bis 2021), das AFiD-Panel Energieunternehmen (bis 2017) oder das AFiD-Panel Dienstleistungen (bis 2020), die zusätzlich Informationen zu Investitionen, geleisteten Arbeitsstunden, gezahlten Löhnen und dem Absatzproduktionswert haben. Ab diesem Herbst wird das Business Tax Panel zudem verknüpfte Daten aus dem Unternehmensregister und verschiedenen Steuerstatistiken, wie der Gewerbe- und der Umsatzsteuer, in einer Paneldimension bereitstellen. Dies ist ein großer Schritt in der Bereitstellung administrativer Unternehmensdaten, da erstmals die Betrachtung einer Vollerhebung von Unternehmen im Längsschnitt und eine Verknüpfung

zwischen verschiedenen Unternehmensstatistiken ermöglicht wird. Auf der Seite der sozialversicherungspflichtigen Beschäftigung bietet das Institut für Arbeitsmarkt- und Berufsforschung (IAB) Daten über die sozialversicherungspflichtig Beschäftigten und deren Betriebe an. Diese Daten sind zwar untereinander (teilweise) verknüpfbar, aber nicht mit denen des AFiD Panels.

- Zahlreiche Verwaltungsdaten, die eben auch für Forschung genutzt werden könnten, fallen - laufend oder einmalig - bei Behörden und Aufsichtsgremien im Rahmen von Unternehmensförderprogrammen, der Umweltberichterstattung oder Kontrollmaßnahmen bei Firmen an.
- Daneben gibt es zahlreiche private Datenanbieter und -aggregatoren, die Unternehmensdaten aus verschiedensten Quellen (u.a. auch amtlichen Registern) sammeln, verknüpfen und aufbereiten, und dann auch der Wissenschaft - allerdings zu teils enormen Kosten - zur Verfügung stellen. So können beispielsweise aufbereitete Varianten des deutschen Handelsregisters käuflich erworben werden, da das offizielle Handelsregister keinen praktikablen Zugang für die Forschung bietet. Desweiteren werden in den Wirtschaftswissenschaften häufig internationale Firmendatenbanken genutzt (z.B. die Orbis-Datenbank des Anbieters Bureau van Dijk und die Finanzmarktdatenbanken von Macrobond, Refinitiv oder Bloomberg). Zudem existieren zahlreiche Firmenadressdatenanbieter, deren Daten häufig zum Versand von Umfragen oder zur genauen Lokalisierung von Unternehmen(-sfilialen) genutzt werden können. Zur Zeit drängen zudem zahlreiche Anbieter von Umwelt- und Nachhaltigkeitsdaten auf den Markt, die ESG- und Emissionsdaten bereitstellen.
- Zahlreiche Unternehmensforschungsdaten stammen zudem aus Befragungen, die oftmals durch öffentliche Forschungseinrichtungen durchgeführt werden. Z.B. werden die Konjunkturumfrage des ifo Instituts, das Mannheimer Innovationspanel des ZEW oder das IAB/ZEW-Gründungspanel der Wissenschaft (in anonymisierter Form und unter zahlreichen Auflagen) über die vom RatSWD akkreditierten und qualitätsgesicherten FDZ kostenfrei zur Verfügung gestellt.¹⁵

Die obige beispielhafte Aufstellung zeigt die große Vielfalt und die unbestrittenen Verbesserungen der letzten Jahre beim Forschungszugang zu Unternehmensdaten in Deutschland. Allerdings ließe sich das Auswertungs-

¹³ siehe <https://www.nfdi.de/sektion-industry-engagement/> und <https://www.stifterverband.org/datagroup> [zuletzt aufgerufen am 08.09.2023].

¹⁴ Nichtsdestotrotz sind einige (Digital-)Unternehmen, die eine marktbeherrschende Stellung einnehmen, schon jetzt gesetzlich zur Datenbereitstellung verpflichtet und weitere Datenteilungsverpflichtungen könnten sinnvoll sein, um negative Auswirkungen auf Wettbewerb und die Gesellschaft unabhängig zu untersuchen.¹⁵ Das Mannheimer Unternehmenspanel (MUP) des ZEW, das auf den Unternehmensdaten der Kreditauskunftei Creditreform basiert und die Grundgesamtheit der deutschen Unternehmen umfasst, wurde mit den Daten des Betriebshistorikpanels (BHP) des IAB verknüpft (Die Verknüpfung wurde durch die amtliche Statistik erlaubt, da die Daten des MUP grundsätzlich öffentlich verfügbar sind und in ähnlicher Form auch durch den Datenanbieter Moody's vertrieben werden.) Durch diese Verknüpfung können Unternehmensdaten mit Betriebsdaten kombiniert werden. Dazu gehören auch Informationen über die sozialversicherungspflichtig und geringfügig Beschäftigten. Diese Verknüpfung beinhaltet eine Unternehmens-ID, die auch in kommerziell vertriebenen Datenbeständen (z.B. Dafne-/ Orbis-Datenbank des Herstellers Bureau van Dijk) vorhanden ist. Daher können Forschende von diesen kommerziellen Anbietern erworbene Unternehmensdaten im FDZ des IAB mit dem BHP verknüpfen und Analysen auf Unternehmensebene mit den kombinierten Variablenbeständen durchführen. Darüber hinaus wird zurzeit ein kombinierter Datensatz generiert, der die Datenbestände des MUP, des BHP und die zugehörigen Beschäftigtendaten zu einem Gesamtdatensatz kombiniert. Dieser wird im FDZ des IAB für Forschende bereitgestellt.

potenzial noch deutlich steigern, insbesondere durch Verknüpfungen der Datensätze untereinander, wie es in vielen anderen europäischen Ländern bereits der Fall ist. Vor dem Hintergrund der anstehenden Transformation der Privatwirtschaft infolge oder durch Digitalisierung (insbesondere Anwendungen der künstlichen Intelligenz), Klimawandel, Umweltschutzmaßnahmen, steigenden Energiekosten, Fachkräftemangel und die Lieferkettenproblematik ist die Zusammenführung von Informationen aus verschiedenen Quellen wichtig. Nur so kann qualitativ hochwertige, international konkurrenzfähige wissenschaftliche Forschung betrieben werden, aus der dann auch evidenzbasierte Politikberatung entsteht. Eine Verknüpfung von Daten der amtlichen Statistik (z.B. AfID, BHP) mit Befragungsdaten zu diesen unterschiedlichen Themen würde einen großen Mehrwert bieten, da dann verschiedene Informationen über Unternehmen miteinander zu einem neuen reichhaltigen Forschungsdatensatz kombiniert werden könnten.¹⁶

Aktuell gibt es verschiedene Probleme, die mit der Nutzung und/oder Verknüpfung von Unternehmensdaten für Forschungsvorhaben in Deutschland einhergehen und die in anderen europäischen Ländern entweder gelöst sind oder gar nicht erst auftreten:

- Ein zentrales Problem beim Forschungszugang zu amtlichen Daten (von verschiedenen Behörden) stellen datenschutzrechtliche und sonstige gesetzliche Nutzungsvorschriften dar. Dies betrifft sowohl die Vielzahl an nichtharmonisierten Regelungen, aber auch die vermutlich historisch bedingte strenge deutsche Interpretation dieser Gesetze. Hinzu kommen oftmals divergierende Auslegungen durch die 18! staatlichen Bundes- und Landesdatenschutzbehörden (Bund, und 17 der Länder - Bayern hat 2) sowie zahlreiche Datenschutzabteilungen in sonstigen Behörden und Einrichtungen. Klar ist: Je mehr Informationen über ein Unternehmen bereitgestellt werden, desto leichter wird es, dieses zu identifizieren. Dies trifft in noch stärkerem Maße zu, wenn Daten verknüpft werden. Deswegen sind geeignete technische und organisatorische Maßnahmen zu treffen, um eine Identifikation zu verhindern (sofern dies geboten ist).¹⁷ Aufwendige Zugangswege zu Daten - wie Gastaufenthalte in FDZs - können andererseits aber auch dazu führen, dass Forschung nicht oder weniger tiefgehend betrieben wird, als dies mit weniger restriktiven Zugangsbeschränkungen wäre. Daher gilt es, die entsprechenden Regelungen mit Augenmaß zu gestalten.
- Verknüpfungen von Daten der amtlichen Statistik mit Daten aus wissenschaftlichen Unternehmensbefragungen dürfen in Deutschland - im Gegensatz zur Verknüpfung mit öffentlich verfügbaren Unter-

nehmensdaten kommerzieller Datenanbieter - nicht durchgeführt werden. Die amtliche Statistik beruft sich bei ihrer ablehnenden Haltung u.a. auf das Bundesstatistikgesetz bzw. die Datenschutzgrundverordnung (DSGVO).

- Die verschiedenen Datensätze verfügen nicht über einen einheitlichen Identifikator (Unternehmensnummer) für Unternehmen. D.h. auch wenn die Daten verknüpft werden können, ist dies sehr aufwändig.
- Daten, die generell erhoben werden, stehen nicht für die Forschung zur Verfügung. Im Bereich der Unternehmensdaten sind dies beispielsweise Daten zum genauen Standort und Umzug der Unternehmen, Daten zu staatlichen Subventionen (Kurzarbeitergeld, Corona-Hilfen), Arbeitnehmerüberlassung, Daten aus KfW-Förderprogrammen. Manchmal stehen diese für eine gesetzlich vorgeschriebene Evaluation einmalig zur Verfügung, aber müssen dann gelöscht oder vollständig anonymisiert werden, obwohl fruchtbare Sekundärforschung möglich wäre.
- Die Nutzung von Daten privater kommerzieller Anbieter birgt gleich mehrere Probleme. Die Nachvollziehbarkeit/Replizierbarkeit kann nicht sichergestellt werden, da eine Versionierung, Vergabe von DOI etc. nach den FAIR-Prinzipien nicht wie an einem FDZ sichergestellt wird. Zudem kann meistens keine kostenfreie Nutzung zu Replikationszwecken eingeräumt werden und die Forschung macht sich von kommerziellen Anbietern abhängig, die ihre Monopolstellung ausnutzen können. Die Preisgestaltung kann zudem dazu führen, dass insbesondere Nachwuchswissenschaftler*innen keinen Zugang zu den Daten bekommen und dadurch weniger Möglichkeiten als Forschende haben, die über größere finanzielle Möglichkeiten verfügen. Kooperationen mit Unternehmen auf freiwilliger Basis helfen dabei nur bedingt, da die Unabhängigkeit der Forschung eingeschränkt sein wird. Anders formuliert: Kein Unternehmen wird seine Daten für eine Forschungsfrage zur Verfügung stellen, wenn dadurch sein Geschäftsmodell in Gefahr gerät.

¹⁶ Ein Nebeneffekt dieser Verknüpfungen wäre die Reduktion des Bürokratieaufwands für Unternehmen, wenn sie nicht mehrmals im Rahmen unterschiedlicher Befragungen und statistischer Erhebungen dieselben Auskünfte geben müssten.

¹⁷ Dabei stellt sich die Frage, ob Daten, die nur Unternehmen und deren Geschäftstätigkeit an sich beschreiben, wirklich den selben strikten datenschutzrechtlichen Maßstäben für Individuen und personenbezogene Daten unterliegen sollten.

4. ZUGANGSBEDINGUNGEN IM INTERNATIONALEN VERGLEICH

In anderen (demokratischen) Ländern, die ähnlich hohen Datenschutzstandards wie Deutschland unterliegen, ergeben sich bereits bessere Zugangsbedingungen für Forschende zu Unternehmensdaten. Beispielhaft soll hier kurz auf einige EU-Länder eingegangen werden, in denen sich in letzten Jahren der Zugang zu Forschungsdaten massiv verbessert hat.¹⁸

Nordische Länder

In Dänemark, Finnland¹⁹, Norwegen und Schweden sind umfassende Registerdaten verfügbar. Zudem besteht die Möglichkeit verschiedene Datenquellen miteinander zu verbinden, beispielsweise Daten aus Unternehmensbefragungen mit administrativen Informationen zu Firmencharakteristika. Die Verknüpfungen erlauben wichtige empirische Fragen überzeugend zu beantworten. Forschende können das jeweilige statistische Amt beauftragen, Datensätze aus verschiedenen Quellen zu verknüpfen und in geeigneter Weise anonymisieren. Eine weitere Besonderheit in diesen Ländern ist, dass der Datenzugang für Wissenschaftlerinnen und Wissenschaftler zu Daten der administrativen Statistik „remote“ – also über sichere Fernzugänge auf Server beim Datengeber – vom eigenen Arbeitsplatz erfolgt.

Österreich: Austrian Micro Data Center

Seit Juli 2022 ermöglicht Statistik Austria in Österreich über das „Austrian Micro Data Center“ einen umfangreichen Remote-Zugriff auf eine stetig wachsende Zahl an verknüpfbare Unternehmensdaten.²⁰

Dies ermöglicht die Verknüpfung von sämtlichen Mikrodaten von Statistik Austria einschließlich Register-, Steuer- und Befragungsdaten mit Daten weiterer öffentlicher Stellen sowie mit Mikrodaten der Forschungseinrichtungen selbst. Ggf. ist hierfür ein einmaliger Erlass des zuständigen Ministeriums erforderlich, das Daten aus seinem Geschäftsbereich für die Forschung freigibt. Die Pseudonymisierung und die Verknüpfung untersteht der unabhängigen Stammzahlenregisterbehörde, die verschiedene Identifikationsnummern vorhält und nur für das jeweilige Forschungsprojekt zusammenbringt, sodass Inhaltsdaten und Identifikationsnummern strikt getrennt verwahrt werden. Für die Datennutzung ist eine Akkreditierung der jeweiligen Forschungseinrichtung, auf die das Amtsgeheimnis übergeht, sowie die Bewilligung eines Projektantrags unter Beachtung des Datenminimierungsgrundsatzes erforderlich.

Daneben streben privatwirtschaftliche Akteure beispielsweise über eine Kooperationsplattform die Errichtung eines sog. „Data Spaces“ an, in dem ein sicherer abteilungs-, organisations- und/oder branchenübergreifender Datenaustausch stattfinden kann.²¹ Allerdings müssen Forschende die laufenden Nutzungskosten des cloudbasierten Remote Access-Systems sowie von Verwaltungsdienstleistungen des AMDC selbst tragen. Da diese Kosten sehr hoch ausfallen können, insbesondere wenn man mit riesigen Unternehmensregisterdaten arbeiten will, wurden Stipendienprogramme (u.a. für Nachwuchswissenschaftler*innen) eingeführt.

Frankreich: Centre d'accès sécurisé aux données

Administrative (und ähnliche vertrauliche) Daten werden in Frankreich über das Centre d'accès sécurisé aux données (CASD) angeboten, das ein Konsortium bestehend aus unterschiedlichen nationalen Instituten repräsentiert (INSEE, GENES, CNRS Polytechnique und HEC Paris).²² Die Hauptaufgabe des CASD besteht darin, einen sicheren Zugang zu vertraulichen Daten für Forschungs- und andere Non-profit-Zwecke zu gewährleisten. Das CASD beherbergt aktuell ca. 500 unterschiedliche Datensätze. Diese stammen aus dem öffentlichen Bereich, wie dem französischen nationalen Amt für Statistik (INSEE) (wie etwa Unternehmens- und Beschäftigungsdaten), dem Gesundheitssektor sowie aus dem privaten Bereich, in dem beispielsweise private Unternehmen mit dem CASD kollaborieren und sensible Daten für die genannten Zwecke zu Verfügung stellen. Sollte ein Projekt genehmigt werden und Zugang zu den Daten (gegen eine entsprechende monetäre Kompensation) gewährt werden, so können innerhalb des Projekts alle beantragten Datenquellen exploriert und verbunden werden. Beispielsweise ist es so möglich, die Sozialversicherungsdaten (DADS, Beschäftigung) mit den jeweiligen Unternehmensdaten (FICUS/FARE) zu verbinden, da jeweils die Unternehmensregisternummer (Code Siren) identifiziert, bei welchem Unternehmen eine bestimmte sozialversicherte Person erwerbstätig ist. Grundsätzlich ist es auch möglich, eigene Daten, wie etwa selbst erhobene Umfragedaten, mit jeglichen CASD-Datenprodukten zu verbinden.

Der Datenschutz beim Arbeiten mit Datenprodukten vom CASD wird durch die sogenannte SD-Box gewährleistet. Diese Hardware erlaubt es den Nutzer*innen via einer projektspezifischen Zugangskarte in Verbindung mit dem Fingerabdruck über eine gesicherte Leitung auf den Server des CASD zuzugreifen (ohne direkte Exportmöglichkeit). Auch hier werden die Nutzungskosten auf die Forschenden umgelegt.²³

¹⁸ Den Autor*innen ist auch nach intensiver Recherche kein einziger (veröffentlichter) Fall bekannt, in dem der einfachere Zugang zu besseren Forschungsdaten in diesen Ländern, zu einem signifikanten Datenleck und/oder Schaden für die in den Daten enthaltenen Personen oder Unternehmen geführt hat.

¹⁹ Finnland hat im Zuge der Umsetzung der EU-Verordnung Nr. 223/2009 über europäische Statistiken große Schritte hin zu weitgehenden Datenzugangsmöglichkeiten unternommen. In der Folge wurde eine Kommission gegründet, die insbesondere zum Ziel hatte, der Wissenschaft besseren Datenzugang zu bieten, um auf dieser Grundlage die wirtschaftspolitische Beratung zu verbessern. Auf dieser Grundlage wurde eine Reform des Statistikgesetzes beschlossen, die der amtlichen Statistik explizit die Aufgabe auferlegt, der Wissenschaft Daten zur Verfügung zu stellen. Zudem wurde eine Verknüpfung von Datensätzen für wissenschaftliche Zwecke unter Berücksichtigung des Datenschutzes ermöglicht.

5. EMPFEHLUNGEN

Aufbauend auf der Darstellung des Status Quo und den daraus resultierenden Unzulänglichkeiten und Problemen ergeben sich für uns vier Handlungsfelder zur Verbesserung des Zugangs zu Unternehmensforschungsdaten in Deutschland: Vereinfachte Verknüpfbarkeit, Verbesserung der Nutzbarkeit gewerblicher Unternehmensdaten, Kapazitätsausbau der Forschungsdatenzentren und Ausweitung des Remote Access.

5.1 Bessere Nutzungsmöglichkeiten und vereinfachte Verknüpfung von amtlichen untereinander sowie mit privaten Daten durch Ausweitung der gesetzlichen Regelungen im Rahmen eines Forschungsdatengesetzes

§13a BStatG erlaubt die Zusammenführung verschiedener Datensätze zu Unternehmen, weist jedoch Limitationen für die wissenschaftliche Arbeit auf. Durch ein Forschungsdatengesetz sollten die gesetzlichen Rahmenbedingungen wissenschaftsfreundlicher gestaltet werden.

a. Wissenschaftliche Forschung als Verknüpfungszweck: Bisher wird wissenschaftliche Forschung nicht explizit als Verknüpfungszweck in §13a BStatG genannt. Dadurch besteht das Risiko, dass Datenverknüpfungen mit Verweis auf den Gesetzeswortlaut abgelehnt werden. Es ist daher zu empfehlen, wissenschaftliche Forschung explizit als Zweck der Verknüpfung in den Gesetzestext aufzunehmen.

b. Beweislastumkehr bei der Datenverknüpfung: Eine restriktive Auslegung der aktuellen Gesetzeslage ermöglicht ausschließlich die Verknüpfung von Datensätzen untereinander, bei denen dies explizit erlaubt ist (z.B. durch das Verwaltungsdatenverwendungsgesetz). Eine Art "Beweislastumkehr" wäre hier wünschenswert, d.h. eine gesetzliche Regelung, die explizit festschreibt, welche Datenverknüpfungen für Forschungszwecke nicht erlaubt sind. Zusammenführungen von neuen Datensätzen, die nicht in expliziten Regelungen erfasst sind, wären hiermit grundsätzlich für die Forschung möglich.

c. Schaffung einer Rechtsgrundlage für die Verknüpfung von Unternehmens- mit Personendaten. Bisher besteht im Rahmen des BStatG (anders als z.B. im Sozialgesetzbuch für Daten der Sozialversicherung) keine Rechtsgrundlage, um Unternehmensdaten mit Personendaten zu verknüpfen. Hierdurch ist es beispielsweise nicht möglich, Auswirkungen von makroökonomischen Schocks auf Unternehmen mit den

Reaktionen der Angestellten z.B. in der Einkommensteuerstatistik zu verknüpfen. §13a BStatG sollte erweitert werden, um solche Analysen zu Forschungszwecken zu ermöglichen.

d. Einführung einer einheitlichen Unternehmens-ID zur Verknüpfung von Datensätzen. Eine einheitliche Unternehmens-ID, auf die von Forschenden auch außerhalb des Statistischen Systems zugegriffen werden kann, könnte die Verknüpfung von administrativen Daten untereinander und auch mit externen Daten (z.B. über die Handelsregisternummer oder die Umsatzsteuernummer) vereinfachen. Hierbei ist das ab Herbst zur Verfügung stehende Business Tax Panel als positives Beispiel zu nennen, bei dem administrative Steuerdatensätze zusammengeführt werden und perspektivisch mit weiteren Datenquellen verknüpft werden können.

e. Schaffung einer Rechtsgrundlage für die Verknüpfung von Unternehmens- mit Befragungsdaten. Bisher besteht im Rahmen des BStatG keine Rechtsgrundlage, um Unternehmensdaten mit (externen) Befragungsdaten wie z.B. der ifo Konjunkturumfrage zu verknüpfen. Hintergrund ist einerseits, dass diese Forschungsdaten nicht öffentlich sind (da nicht für jedermann zugänglich) und andererseits oftmals eine konkrete Einwilligung (narrow consent) zum spezifischen Verknüpfungsvorhaben vorliegen muss, was aber bei Erhebung der Befragungsdaten oftmals noch nicht klar ist. Bei (Befragungs-)Daten, die nur Unternehmen und deren Geschäftstätigkeit an sich beschreiben, ist zu überdenken, ob diese wirklich den selben strikten datenschutzrechtlichen Maßstäben für Individuen und personenbezogene Daten (inklusive informierter Einwilligung im Einzelfall) unterliegen sollten. Stattdessen könnte klargestellt werden, dass die DSGVO nicht für diese Art von Unternehmensdaten gilt. Falls dies nicht möglich ist, könnte mindestens klargestellt werden, dass eine sogenannte "breite Einwilligung" (broad consent) zur Forschung generell ausreichend für diese Verknüpfungsvorhaben ist.

f. Langfristige sichere Speicherung von Identifikatoren und Schutz vor Löschung: Um Daten auch perspektivisch (für Forschungsprojekte die jetzt noch gar nicht absehbar sind) verknüpfen zu können, ist es essentiell, dass Identifikatoren keiner Löschungspflicht unterliegen und keine Anonymisierungsver-

²⁰ siehe <https://www.statistik.at/services/tools/services/center-wissenschaft/austrian-micro-data-center-amdc> [zuletzt aufgerufen am 08.09.2023].

²¹ siehe www.dataintelligence.at [zuletzt aufgerufen am 08.09.2023].

²² siehe <https://www.casd.eu/en/> [zuletzt aufgerufen am 08.09.2023].

²³ Ganz ähnlich funktioniert der Prozess in den Niederlanden. Dort können über eine eindeutige Identifikationsnummer verschiedene amtliche Register untereinander verbunden und mit Befragungsdaten verknüpft werden. Die resultierenden Forschungsdaten werden anschließend wieder anonymisiert und per Remote Access zur Verfügung gestellt.

pflichtung besteht. Stattdessen ist eine sichere, ggf. separate Speicherung von Identifikatoren anzuraten, um eine spätere Verknüpfung von Datensätzen zu ermöglichen. gleichermaßen sollen auch die Daten an sich langfristig gespeichert und vorgehalten werden. Dies dient zudem dazu die Nachvollziehbarkeit und Replizierbarkeit von einmal publizierten Ergebnissen zu sichern und damit die für die Wissenschaft unverhandelbar Prinzipien der guten wissenschaftlichen Praxis sicherzustellen. Des Weiteren sind bestimmte Entwicklungen (z.B. Klima-/Strukturwandel nur über einen langen Zeitraum sichtbar, sodass eine Löschung der Identifikatoren nach 30 Jahren eine Untersuchung dieser langfristigen Phänomene unmöglich macht. Zudem ist die Schutzbedürftigkeit von 30 Jahre alten Unternehmensdaten stark in Zweifel zu ziehen, wenn sogar tagesaktuelle, sehr detaillierte Unternehmensdaten mit Klarnamen von kommerziellen Anbietern veröffentlicht werden dürfen.

5.2 Verbesserung der Nutzbarkeit von (bisher nur) gewerblichen Unternehmensdaten

Derzeit haben einzelne Forschende insbesondere an wenig finanzkräftigen, kleineren Hochschulen nur begrenzt Möglichkeiten, Zugriff auf Unternehmensdaten gewerblicher Anbieter zu erhalten. Einzelpersonen an Universitäten und Forschungseinrichtungen haben nur begrenzt die Chance, Verhandlungen mit großen Datenanbietern zu führen und ebenfalls meist keine ausreichenden Finanzmittel. Dies führt zu ungleichen Chancen innerhalb der Forschung. Vor diesem Hintergrund sind mehrere Szenarien zur verbesserten Nutzung von Unternehmensdaten denkbar.

a. Forschungstauglicher Massendatenzugang zu amtlichen Registern

Viele privatwirtschaftliche Datenanbieter verkaufen oftmals amtliche Datenbanken, die allgemein zugängliche Informationen beinhalten, wie z.B. das Handelsregister oder den Bundesanzeiger. Diese öffentlichen Register erlauben oftmals nur Einzelabfragen und verlangen manchmal für bestimmte Informationen eine kleine Gebühr. Für Forscher*innen ist dieser (halfwegs) freie Zugang zu diesen Registern aber oftmals nutzlos, da ein automatisiertes Herunterladen von großen Mengen von Unternehmensinformationen, die für die volkswirtschaftliche Forschung nun mal notwendig sind, entweder technisch verhindert wird oder sogar explizit unter Strafe steht. Diese technischen Barrieren und Strafregelungen sollten abgeschafft werden und die Forschung sollte Zugang zu diesen Registern über eine Programmierschnittstelle (API) halten bzw. ein Abzug der Datenbanken sollte zumindest mehrmals im Jahr zum Download zur Verfügung gestellt werden.

b. Vereinfachte Nutzung bestehender Unternehmensdaten durch gemeinsame Verhandlungsfüh-

rung: Damit alle Forschenden bessere Möglichkeiten der Datennutzung kommerzieller Daten haben, sollten die Wissenschaftsinstitutionen - ähnlich wie bei wissenschaftlichen Zeitschriften über die Bibliotheken - gemeinsame Verhandlungen über Datenzugänge führen, um so die Kosten für den Einzelnen zu verringern, indem Marktmacht aufgebaut wird. Über entsprechende Nationallizenzen könnten auch Forschende kleinerer Institutionen Unternehmensdaten für ihre Forschung nutzen. Dafür sind entsprechend staatliche Mittel für Universitäten und Forschungseinrichtungen vorzusehen.

c. Stärkere Berücksichtigung von Datenkosten in Förderprogrammen:

Auch bei gemeinsamen Datenverhandlungen werden Kosten für Daten von und über Unternehmen zukünftig steigen, da bestimmte Fragestellungen aufgrund von Lücken in der amtlichen Statistik nur mit privatwirtschaftlichen Daten beantwortet werden können. Dies erfordert, dass entsprechende Budgets bei der Forschungsförderung wie der DFG entsprechend gegeben sein müssen, dass Kosten der Datennutzung in Anträgen stärker berücksichtigt werden können und/oder dass Kooperationen mit Unternehmen auch im Rahmen von Förderrichtlinien berücksichtigt werden können.

d. Aufbau einer institutsübergreifenden Unternehmensdaten-Infrastruktur:

Zur Stärkung der Wissenschaft im Datenzugangsprozess sollte eine Infrastruktur auf- und ausgebaut werden, die die Verknüpfung und Nutzung von Daten aus unterschiedlichen Quellen ermöglicht (s. 4.1). Hier würden sich die FDZ des statistischen Verbundes anbieten (s. auch 4.3). Auch das geplante Dateninstitut, Datentreuhänder und NFDI-Konsortien könnten hierbei ihr Wissen einbringen. Vergleichbar zu den Beispielen aus anderen Ländern sollten die Forschenden persönlich besonders auf Geheimhaltung verpflichtet werden und damit aber auch entsprechende Nutzungsrechte bekommen. Dies könnte in Form eines Wissenschaftsgeheimnisses (analog zum Bankgeheimnis) in einem Forschungsdatengesetz geregelt werden.

5.3 Kapazitätsausbau der Forschungsdatenzentren und zeitnahe Bereitstellung von Daten

a. Ausbau der Personalkapazitäten: Derzeit sind mehrere FDZ der statistischen Ämter und weiterer Behörden stark unterbesetzt. Dies führt dazu, dass Forschungsprojekte teils nur verspätet begonnen werden können, sich Datenschutzprüfungen zeitintensiv gestalten und vor allem nur wenige Termine an Gastwissenschaftlerarbeitsplätzen zur Verfügung stehen. Dies verzögert den wissenschaftlichen Prozess und kann insbesondere bei jungen Wissenschaftler*innen mit befristeten Verträgen negative Auswirkungen auf ihre Karriere haben.

b. Ausbau der IT-Infrastruktur: Durch eine Modernisie-

rung der IT-Infrastruktur in den Forschungsdatenzentren könnten Forschungsprozesse deutlich beschleunigt werden, beispielsweise durch eine schnellere Datenverarbeitung am Gastwissenschaftlerarbeitsplatz oder durch einen vor Zugriff gesicherten Remote Desktop Zugriff von außerhalb auf die Daten eines FDZ (siehe 4.4).

c. Zeitnahe Bereitstellung von Daten: Viele Datensätze in den FDZ sind nur mit erheblicher Zeitverzögerung verfügbar. Beispielsweise reichen aktuell die Daten aus der Gewerbesteuerstatistik lediglich bis zum Jahr 2018. Dadurch ist eine Evaluierung aktueller Politikmaßnahmen mit amtlichen Daten nicht möglich. Eine schnellere Datenbereitstellung und eine Schaffung der dafür erforderlichen Personalkapazitäten in den statistischen Ämtern würde hier Abhilfe schaffen.

d. Klare Zuständigkeiten, Harmonisierung und Straffung der Prozesse in den Statistischen Ämtern: Neue Forschungsprojekte mit amtlichen Statistiken müssen teils von allen Statistischen Landesämtern geprüft und genehmigt werden, die abweichenden Landesdatenschutzregelungen unterliegen oder diese unterschiedlich auslegen. Dieser Prozess birgt hohe Redundanzen, bindet Personalkapazitäten in den Ämtern und führt zu Verzögerungen beim Datenzugang. Hier sind einheitliche, bundeslandübergreifende Regelungen und eine Straffung der Prozesse nötig, mit der administrative Daten nur noch von einer einzelnen Instanz genehmigt werden müssen (One-for-all-Prinzip). Dies beschleunigt nicht nur den Beginn von Forschungsprojekten, sondern reduziert die Belastung des Personals in den statistischen Ämtern. Frei werdende Personalkapazitäten könnten stattdessen zur Erweiterung von Datenprodukten und Datenzugriffsmöglichkeiten für Forschende eingesetzt werden.

e. Eigener Forschungsauftrag für das Statistische Bundesamt und die Statistischen Landesämter: Aktuell dürfen das Statistische Bundesamt und die statistischen Landesämter nur nach langwierigen Verordnungs- und Gesetzesverfahren tätig werden. Einen wissenschaftlichen Betätigungszweck gibt es nicht. Nach dem Vorbild des Verbundes zwischen der Bundesagentur für Arbeit und dem IAB könnte, mit einem Forschungsauftrag ausgestattet, ein Forschungsinstitut am Statistischen Bundesamt eingerichtet werden. Für die Landesämter muss zudem der Betrieb eines Forschungsdatenzentrums gesetzlich verankert werden.

5.4 Ausweitung des Datenzugriffs per Remote Access

Die Bereitstellung von Daten durch die Statistischen Ämter erfolgt i.d.R. durch Scientific Use Files (SUF), per Arbeit an einem sicheren, überwachten Gastwissenschaftlarbeitsplatz (GWAP) oder per kontrollierter Datenfernverarbeitung (KDFV). Da Forschende ein SUF herunterladen und am eigenen Arbeitsplatz bearbeiten können, müssen die Originaldaten teils deutlich vergrößert werden, um eine ausreichende Anonymisierung sicherzustellen.

Dies reduziert das Analysepotential der Daten. Gleichzeitig haben SUFs den Vorteil, dass die Forschenden von ihrem Arbeitsplatz aus arbeiten können und nicht auf Öffnungszeiten der FDZs oder auf teils langwierige Datenschützprüfungen durch FDZ-Mitarbeitende angewiesen sind. Am GWAP oder per KDFV werden deutlich detaillierte Daten bereitgestellt. Jedoch sind die Forschenden am GWAP darauf angewiesen, einen Termin im FDZ vor Ort zu erhalten. Hier bestehen teils lange Wartezeiten bei der Terminbuchung und ggf. eine weite Anreise. Bei der KDFV programmieren die Forschenden an Dummy-Datensätzen und schicken daraufhin ihren Code ein, der von FDZ-Personal an den Originaldaten ausgeführt wird. Ergebnisse werden dann nach Prüfung den Forscher*innen zugeschickt. Das Programmieren, ohne selbst direkt auf die Daten Zugriff zu haben, führt zu einer stark erhöhten Fehleranfälligkeit und oftmals zu einer zeitintensiven Kommunikation zwischen dem FDZ-Mitarbeitenden und Forschenden, die signifikante Personalressourcen bindet.

Alle drei Zugangswege sind auch für die FDZs mit erheblichem Aufwand verbunden, sei es der Aufwand zur Erstellung und Anonymisierung der SUFs, der Bereitstellung und Überwachung eines Arbeitsplatzes oder der Prüfung und Freigabe von Ergebnissen.

Der in vielen anderen Ländern (siehe 3.) übliche Remote Access ermöglicht hingegen die direkte Fernverarbeitung von Daten auf den Servern der Forschungsdatenzentren durch Forschende. Hierbei verbleiben die Daten zwar in der FDZ-Infrastruktur, die Forschenden haben jedoch die Möglichkeit, direkt auf die Daten zuzugreifen und Programmcodes auf Basis der Daten zu entwickeln, ohne selbst die Daten exportieren zu können.

Dies birgt mehrere Vorteile gegenüber den herkömmlichen Datenzugangswegen: Forschende können direkt in einer sicheren und datenschutzkonformen Umgebung mit Originaldaten arbeiten und sind dabei nicht durch terminliche Engpässe seitens der FDZ und Reiseaufwände eingeschränkt. Gleichzeitig erfordert dies weniger Personalaufwand der FDZs, da die Erstellung von SUFs und die zeitintensive Betreuung von Gastforschenden entfällt.

Daher sollte Remote Access grundsätzlich als Zugangsweg etabliert werden. Dies erfordert entsprechende gesetzliche Regelungen, beispielsweise innerhalb des § 16 Abs. 6 BStatG. Wichtig ist dabei, dass die Granularität der Daten im Vergleich zur GWAP- oder KDFV-Nutzung nicht signifikant reduziert wird.