DOI: 10.2478/wd-2023-0208

sciendo

Wirtschaftsdienst, 2023, 103(11), 750-753

JEL: C80, D22

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

Unternehmensdaten: Nutzbarkeit verbessern

Welche Maßnahmen der Regional- und Wirtschaftsförderung wirken und warum? Wie reagieren Firmen auf den zunehmenden Fachkräftemangel, Lieferkettenprobleme oder die steigende Inflation? Zuverlässige Antworten auf wirtschaftspolitisch relevante Fragen zur Lage von Unternehmen und damit auch der deutschen Volkswirtschaft sind fundamental wichtig für private Haushalte, die Unternehmen selbst und nicht zuletzt die Politik, die geeignete Rahmenbedingungen setzen muss. Wirtschaftsforschende können solche Fragen allerdings nur dann evidenzbasiert analysieren und zukünftige Entwicklungen prognostizieren, wenn ihnen qualitativ hochwertige Unternehmensdaten zeitnah zur Verfügung stehen. Gute Politikberatung und der Transfer in die Öffentlichkeit erfordert daher einen guten Datenzugang für die (wirtschafts-) wissenschaftliche Spitzenforschung und ist fundamental für den Wissenschaftsstandort Deutschland. Allerdings ist die Ausgangslage in Deutschland beim Forschungsdatenzugang im internationalen Vergleich eher schlecht. Auch die Bundesregierung hat die Dringlichkeit eines verbesserten Zugangs zu Forschungsdaten erkannt und in ihren Datenstrategien zahlreiche Gesetze und Maßnahmen angekündigt.

Status quo

Unternehmensdaten sind für viele wirtschaftswissenschaftliche Teildisziplinen fundamental. Oftmals werden möglichst zeitnah viele, detaillierte Informationen über Firmen, deren Geschäftstätigkeit, Handelspartner, Kund:innen und/oder Mitarbeitende benötigt. Sie ermöglichen es Treiber von makroökonomischen Größen, wie Inflation oder Arbeitslosigkeit, aber auch mikroökonomisches Verhalten von Arbeitnehmenden oder Unternehmen genauer zu untersuchen. Gerade die Verknüpfung verschiedener Datensätze mit unterschiedlichen Informationen auf Unternehmensebene trägt dazu bei. Zwar werden dabei Informationen über einzelne Unternehmen für die statistische Analyse benötigt, aber das Einzelunternehmen und dessen zum Teil sensitive Unternehmensinterna sind nicht der Untersuchungsgegenstand per se.

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

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

* Dies ist eine Kurzversion der Stellungnahme der Unterarbeitsgruppe Unternehmensdaten des Vereins für Socialpolitik. Die Langfassung findet sich hier: https://www.socialpolitik.de/de/vfs_stellungnahmen_daten_unternehmensdaten.

Die Wissenschaft strebt daher einen Datenzugang nach den FAIR-Prinzipien (Findable, Accessible, Interoperable, Re-usable) an, die für auffindbare, zugängliche, interoperable und nachnutzbare Forschungsdaten stehen. Neben guter Dokumentation und dauerhafter Verfügbarkeit der Daten bedeutet dies auch, dass die Zugangswege und Nutzungsbedingungen für alle Forschenden transparent sind. Im Falle von Firmendaten kann dies bedeuten, dass detaillierte (Mikro-)Daten nur für wissenschaftliche, nichtkommerzielle Forschung unter strengen vertraglichen, organisatorischen und technischen Sicherheitsmaßnahmen, wie z. B. in einem sicheren, überwachten Umfeld eines Forschungsdatenzentrums (FDZ) verfügbar sind, sodass hohes Analysepotenzial und der Schutz von personenbezogenen Daten und Unternehmensgeheimnissen gewährleistet ist. Grundsätzlich lassen sich Unternehmensdaten in zwei Kategorien einteilen:

1. *Daten von Unternehmen* sind z. B. Nutzungsdaten von sozialen Medien, Transaktionsdaten von Finanzdienstleistern oder Geschäftsprozessen sowie Mobilitäts-, Telekommunikations- oder Personaldaten. Der Datenzugang beruht auf projektspezifischen Forschungskooperationen, die in Deutschland aufgrund der rechtlichen Unsicherheit noch relativ selten sind.[1] Zudem werden diese Daten (in anonymisierter oder aufbereiteter Form) von Firmen selbst oder Datenhändlern angeboten – oftmals für hohe Summen und mit restriktiven Lizenzbedingungen. Um einen besseren Zugang für die Wissenschaft für solche Daten von Unternehmen auszuloten, wurden die NFDI-Sektion „Industry Engagement", die Stifterverbandsinitiative „Datagroup Business 2 Science" und das Projekt „Zugang zu Firmendaten für die Wirtschaftswissenschaft" des NFDI-Konsortiums „KonsortSWD" ins Leben gerufen. Da Kooperationen zwischen Wissenschaft und Unternehmen auf Freiwilligkeit beruhen, ist eine gesetzliche Regulierung kaum möglich.

1 Der neue EU Data Act ist keine Grundlage für einen verpflichtenden Forschungszugang zu diesen Unternehmensdaten außer in sehr eng definierten Notlagen. Positive Ausnahme in Deutschland sind die RWI-GEO-RED-Daten, die von der Immobilienplattform ImmobilienScout24 kommen und der Forschung kostenfrei bereit gestellt werden (https://www.rwi-essen.de/forschung-beratung/weitere/forschungsdatenzentrum-ruhr/datenangebot/rwi-geo-red-real-estate-data). Internationale Vorbilder für solche Kooperationen sind der Opportunity Insights Econmic 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 von Facebook.

2. *Daten über Unternehmen* werden in der Forschung bisher am häufigsten verwendet. Sie stammen aus verschiedenen Quellen:

In der amtlichen Statistik werden Daten erhoben und von staatlichen FDZ bereitgestellt, wie z. B. das AFiD-Panel Unternehmensregister. Das neue Business Tax Panel verknüpft Daten aus dem Unternehmensregister mit verschiedenen Steuerstatistiken. Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) bietet Informationen zu sozialversicherungspflichtig Beschäftigten und Betrieben. Diese Daten sind nur vereinzelt miteinander verknüpfbar.

Private Anbieter bieten Unternehmensdaten aus verschiedenen Quellen zu teils hohen Kosten an. So können z. B. aufbereitete Varianten des deutschen Handelsregisters käuflich erworben werden, da das offizielle Handelsregister keinen praktikablen Zugang für die Forschung bietet. Internationale Firmendatenbanken und Anbieter von Firmenadressdaten zum Versand von Umfragen oder zur Lokalisierung von Unternehmensfilialen oder Betriebsstätten sind ebenfalls relevant. Anbieter von Umwelt- und Nachhaltigkeitsdaten drängen aktuell ebenfalls auf den Markt. Zahlreiche Forschungsdaten stammen zudem aus Unternehmensbefragungen, die von öffentlichen Forschungseinrichtungen durchgeführt werden. Diese werden zumeist (in anonymisierter Form und unter zahlreichen Auflagen) über die vom RatSWD akkreditierten und qualitätsgesicherten FDZ kostenfrei zur Verfügung gestellt.[2]

Trotz der Vielzahl an Datenquellen könnte und sollte das Analysepotenzial dieser Daten erheblich erweitert werden, indem Datensätze verknüpft werden (dürfen), wie es in vielen europäischen Ländern bereits langjährig und erfolgreich praktiziert wird. Angesichts der bevorstehenden wirtschaftlichen Transformation durch Digitalisierung, KI, Klimawandel, Umweltschutz, steigende Energiekosten, Fachkräftemangel und Lieferkettenprobleme ist es entscheidend, Informationen aus verschiedenen Quellen zusammenführen zu können. Nur so kann hochwertige Forschung betrieben werden, die evidenzbasierte Politikberatung ermöglicht. Die Verknüpfung von amtlichen Statistikdaten (z. B. AFiD, BHP) mit Umfragedaten zu diesen Themen wäre äußerst wertvoll. Dies würde auch den Aufwand für Unternehmen verringern, da sie nicht mehrfach die gleichen Informationen für verschiedene Erhebungen bereitstellen müssten.

2 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 Moodys vertrieben werden. Durch diese Verknüpfung können Firmendaten mit Betriebsdaten kombiniert werden.

Dr. Sandra Gottschalk ist wissenschaftliche Mitarbeiterin im Forschungsbereich Innovationsökonomik und Unternehmensdynamik am ZEW Mannheim.

Prof. Dr. Carla Krolage ist Professorin an der Universität Regensburg und Forschungsdirektorin am ifo Zentrum für Makroökonomik und Befragungen in München.

Dr. Georg Licht leitete den Forschungsbereich Innovationsökonomik und Unternehmensdynamik am ZEW Mannheim.

Prof. Dr. Andreas Peichl ist Leiter des ifo Zentrums für Makroökonomik und Befragungen und Professor für Volkswirtschaftslehre, insbesondere Makroökonomie und Finanzwissenschaft, an der Ludwig-Maximilians-Universität München.

Dr. Sandra Schaffner ist Leiterin des Forschungsdatenzentrum (FDZ Ruhr) am RWI Essen.

Dr. Sebastian Wichert ist Leiter des LMU-ifo Economics & Business Data Center (EBDC, Forschungsdatenzentrum) in München.

In Deutschland stoßen die Nutzung und Verknüpfung von Unternehmensdaten für Forschungszwecke derzeit auf große Probleme, die in anderen europäischen Ländern entweder bereits gelöst sind oder überhaupt nicht auftreten.

- Ein zentrales Problem beim Zugang zu amtlichen Daten stellen die zu restriktiven datenschutzrechtlichen und sonstigen Nutzungsbestimmungen dar. Dies betrifft sowohl die Vielzahl an nicht harmonisierten Regelungen, aber auch deren (über-)strenge Auslegung hierzulande, die je nach involvierter Datenschutzstelle von Bund, Land, Behörden und Einrichtungen auch divergiert.

- Verknüpfungen von Daten der amtlichen Statistik mit Daten aus wissenschaftlichen Unternehmensbefragungen sind in Deutschland – im Gegensatz zu Verknüpfungen mit öffentlich verfügbaren Unternehmensdaten kommerzieller Datenanbieter – nicht bzw. nur unter unrealistischen Bedingungen zulässig.

- Die verschiedenen Datensätze verfügen über keinen einheitlichen Identifikator für Unternehmen. D. h. auch

wenn die Daten verknüpft werden dürfen, ist dies sehr aufwändig. Zudem müssen manche Identifikatoren nach einer gesetzlichen Frist gelöscht werden.

- Daten, die im Rahmen von staatlichen (Förder-/Kontroll-)Maßnahmen erhoben werden, stehen, abseits von einer gesetzlich vorgeschriebenen Evaluation, oftmals nicht für die weitere Forschung zur Verfügung.

- Die Nutzung von Daten kommerzieller Anbieter, oftmals von eigentlichen öffentlichen Daten, birgt mehrere Probleme. Die Nachvollziehbarkeit/Replizierbarkeit kann nicht sichergestellt werden, da eine Versionierung, Dokumentation etc. nach den FAIR-Prinzipien nicht wie in einem FDZ, sichergestellt ist. Eine kostenfreie Nutzung zu Replikationszwecken wird in der Regel nicht eingeräumt. Die Preisgestaltung kann dazu führen, dass insbesondere Nachwuchsforschende mit geringeren finanziellen Mitteln keinen Zugang zu den Daten erhalten.

Empfehlungen

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

§13a BStatG erlaubt die Verknüpfung einiger Datensätze zu Unternehmen, weist jedoch Limitationen für die Wissenschaft auf. Folgende Punkte sollten in einem Forschungsdatengesetz wissenschaftsfreundlicher gestaltet werden.

Wissenschaftliche Forschung als Verknüpfungszweck: Wissenschaftliche Forschung sollte im §13a BStatG als Verknüpfungszweck explizit erwähnt werden, um Ablehnungen von Datenverknüpfungen zu vermeiden.

Einführung einer Beweislastumkehr: Eine restriktive Auslegung der aktuellen Gesetzeslage ermöglicht nur die Verknüpfung von amtlichen Datensätzen untereinander, bei denen dies explizit erlaubt ist (z. B. durch das Verwaltungsdatenverwendungsgesetz). Eine Art „Beweislastumkehr" sollte bestehen, um Verknüpfungen zu ermöglichen, die nicht explizit untersagt sind, anstatt auf ausdrückliche Genehmigungen angewiesen zu sein.

Schaffung einer Rechtsgrundlage für die Verknüpfung von Unternehmens- mit Personendaten: Bisher besteht im Rahmen des BStatG keine Rechtsgrundlage, um Unternehmensdaten mit Personendaten zu verknüpfen. Dadurch ist es z. B. nicht möglich, Auswirkungen von makroökonomischen Schocks auf Firmen mit den Reaktionen der Beschäftigten in der Einkommensteuerstatistik zu verknüpfen.

Einführung einer einheitlichen Unternehmens-ID zur Verknüpfung von Datensätzen: Eine einheitliche Unternehmens-ID sollte in allen Unternehmensstatistiken zur Erleichterung der Verknüpfung von Verwaltungsdaten untereinander und mit externen Daten eingeführt werden. Ein positives Beispiel ist hier das neu geschaffene Business Tax Panel.

Schaffung einer Rechtsgrundlage für die Verknüpfung von Unternehmens- mit Befragungsdaten: Für die Verknüpfung von Unternehmensdaten mit externen Umfragedaten gibt es bisher keine Rechtsgrundlage. Diese Forschungsdaten sind nicht öffentlich und eine konkrete Einwilligung zum spezifischen Verknüpfungsprojekt liegt nicht vor. (Befragungs-)Daten, die nur Unternehmen und deren Geschäftstätigkeit beschreiben, sollten nicht den gleichen strengen datenschutzrechtlichen Maßstäben wie personenbezogene Daten unterliegen. Stattdessen könnte klargestellt werden, dass die DSGVO nicht für diese Art von Unternehmensdaten gilt. Mindestens könnte jedoch klargestellt werden, dass eine „breite Einwilligung" zur Forschung generell ausreichend für eine Verknüpfung ist.

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

Derzeit haben einzelne Forschende, insbesondere an weniger finanzkräftigen, kleineren Hochschulen nur begrenzt Möglichkeiten, Zugang zu Firmendaten gewerblicher Anbieter zu erhalten, bzw. kaum die Chance, mit großen Datenanbietern zu verhandeln. Dies führt zu ungleichen Chancen innerhalb der Forschung, denen durch mehrere Maßnahmen entgegengewirkt werden könnte:

Forschungstauglicher Massendatenzugang zu amtlichen Registern: Öffentliche Register, wie das Handelsregister oder der Bundesanzeiger, erlauben nur Einzelabfragen und verlangen für bestimmte Informationen eine geringe Gebühr. Für Forschende ist dieser Zugang nutzlos, da das automatisierte Herunterladen großer Mengen von Unternehmensinformationen, die für die Forschung notwendig wären, technisch verhindert oder sogar explizit unter Strafe gestellt wird. Diese technischen Hürden und Strafregelungen sollten für wissenschaftliche Zwecke aufgehoben werden. Die Forschung sollte über eine Programmierschnittstelle (API) Zugang erhalten.

Vereinfachte Nutzung bestehender Unternehmensdaten durch gemeinsame Verhandlungsführung: Um die Nutzungsmöglichkeiten kommerzieller Daten für alle Forschenden zu verbessern, sollten die wissenschaftlichen Einrichtungen, ähnlich wie bei den wissenschaftlichen Zeitschriften über die Bibliotheken, den Datenzugang gemeinsam aushandeln, um die Kosten für einzelne zu sen-

ken und Ungleichheiten zu beheben. Hierfür sollten entsprechende öffentliche Mittel bereitgestellt werden.

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 entsprechende Budgets bei Forschungsförderern, wie der DFG, sodass Kosten der Datennutzung in Anträgen stärker berücksichtigt werden und/oder dass Kooperationen mit Unternehmen auch im Rahmen von Förderrichtlinien berücksichtigt werden können.

Aufbau einer institutsübergreifenden Unternehmensdateninfrastruktur: Zur Stärkung der Wissenschaft im Datenzugangsprozess sollte eine Infrastruktur auf- und ausgebaut werden, die die Verknüpfung und Nutzung von Daten aus unterschiedlichen Quellen ermöglicht. Hierfür bieten sich die FDZ des Statistischen Verbundes an. Auch das geplante Dateninstitut, Datentreuhänder und NFDI-Konsortien könnten hier ihr Wissen einbringen. Analog zu anderen Ländern sollten die Forschenden persönlich besonders auf Geheimhaltung verpflichtet werden und damit entsprechende Nutzungsrechte erhalten. Dies könnte in Form eines Wissenschaftsgeheimnisses (analog zum Bankgeheimnis) im Forschungsdatengesetz geregelt werden.

3. *Kapazitätsausbau der FDZ und zeitnahe Bereitstellung*

Ausbau der Personalkapazitäten: Einige FDZ der statistischen Ämter sind stark unterbesetzt, was Verzögerungen bei Forschungsprojekten, zeitaufwendige Datenschutzprüfungen und begrenzte Gastwissenschaftlerarbeitsplatz-Termine verursacht.

Ausbau der IT-Infrastruktur: Eine bessere IT-Infrastruktur der FDZ kann die Forschungsprozesse beschleunigen, insbesondere kann dies durch schnellere Datenverarbeitung am Gastwissenschaftlerarbeitsplatz und einen sicheren Remote-Desktop-Zugriff ermöglicht werden.

Zeitnahe Bereitstellung von Daten: Viele Datensätze in den FDZ erscheinen stark zeitverzögert. Daten aus der Gewerbesteuerstatistik reichen z. B. nur bis 2018, was die Evaluierung aktueller Politikmaßnahmen mit amtlichen Daten unmöglich macht. Eine schnellere Datenbereitstellung schafft Abhilfe.

Klare Zuständigkeiten, Harmonisierung und Straffung der Prozesse in den Statistischen Ämtern: Forschungsprojekte mit amtlichen Statistiken erfordern derzeit Genehmigungen von verschiedenen Statistischen Landesämtern, die unterschiedliche Landesdatenschutzregelungen anwenden. Dies führt zu Redundanzen, Personalaufwand und Datenzugangsverzögerungen. Einheitliche, bundeslandübergreifende Regelungen und eine Prozessoptimierung sind notwendig.

Eigener Forschungsauftrag für das Statistische Bundesamt und die Statistischen Landesämter: Das Statistische Bundesamt darf nur nach langwierigen Verordnungs- und Gesetzesverfahren wissenschaftlich tätig werden. Die Schaffung eines Forschungsinstituts basierend auf einem solchen Auftrag, ähnlich der Verbindung der Bundesagentur für Arbeit und dem Forschungsinstitut IAB, wäre sinnvoll. Die Statistischen Landesämter sollten den gesetzlichen Auftrag zum Betrieb eines FDZ erhalten.

4. *Ausweitung des Datenzugriffs per Remote-Access*

Die Statistischen Ämter stellen Daten über Scientific-Use-Files (SUF), sichere Gastwissenschaftsarbeitsplätze (GWAP) oder kontrollierte Datenfernverarbeitung (KDFV) zur Verfügung. Bei der Nutzung von SUFs müssen die Originaldaten zur Anonymisierung vergröbert werden, was das Analysepotenzial einschränkt. GWAP und KDFV bieten detailliertere Daten, erfordern jedoch Vor-Ort-Termine mit Wartezeiten und Anreisekosten. In der KDFV senden Forschende per Email Programmcodes an das FDZ, was zu Kommunikationsaufwand und Fehleranfälligkeit führt. Ein Remote-Access ermöglicht die direkte Fernverarbeitung von Daten auf FDZ-Servern und sollte als Standardzugang dienen. Dabei bleiben die Daten in der FDZ-Infrastruktur, ermöglichen jedoch einen direkten Datenzugriff und eine Programmentwicklung. Forschende arbeiten in einer sicheren Umgebung mit Originaldaten, ohne Terminprobleme oder Reisen. Dies verringert den Personalaufwand der FDZ, da SUF und Gastforscherbetreuung entfallen. Dies erfordert gesetzliche Regelungen, um die Datengranularität beizubehalten.

Title: *Company Data: Improve Usability*

Abstract: *High-quality company data is crucial for state-of-the-art empirical research and competition, as well as for evidence-based policy advice, reliable forecasts and policy evaluations. In Germany, however, there are striking deficits in the availability of and access to company data compared to other European countries. This leads to disadvantages in scientific competition, but above all to major welfare losses for the German economy and society as a whole. The planned Research Data Act is intended to remedy the situation. This article proposes four key measures in the area of company data to sustainably improve research and policy evaluation in Germany.*