

Syme, Lorna; Rink, Anke

Article

Nutzung von Geokoordinaten zur Verknüpfung des statistischen Unternehmensregisters mit dem Bundesarztverzeichnis

WISTA – Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Syme, Lorna; Rink, Anke (2023) : Nutzung von Geokoordinaten zur Verknüpfung des statistischen Unternehmensregisters mit dem Bundesarztverzeichnis, WISTA – Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 75, Iss. 5, pp. 38-46

This Version is available at:

<https://hdl.handle.net/10419/279661>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Dr. Lorna Syme

ist promovierte Wirtschaftswissenschaftlerin und seit 2021 im Referat „Unternehmensregister, -demografie, Verwaltungsdatenspeicher, Handwerk“ des Statistischen Bundesamtes tätig. Schwerpunkt ihrer Arbeit ist die Erschließung neuer Quellen für das statistische Unternehmensregister.



Anke Rink

ist Diplom-Geografin und Referentin im Referat „Unternehmensregister, -demografie, Verwaltungsdatenspeicher, Handwerk“ des Statistischen Bundesamtes. Neben dem Fachthema Unternehmensdemografie und Handwerk ist sie auch für die Georeferenzierung im statistischen Unternehmensregister zuständig. Sie erstellt methodische Konzepte zur Anbindung des statistischen Unternehmensregisters an den Geokodierungsdienst und betreut die Datenlieferungen.

NUTZUNG VON GEOKOORDINATEN ZUR VERKNÜPFUNG DES STATISTISCHEN UNTERNEHMENSREGISTERS MIT DEM BUNDESARZTVERZEICHNIS

Lorna Syme, Anke Rink

📌 **Schlüsselwörter:** Geokodierung – geografische Gitterzellen – Arztpraxis – Namensabgleich – räumliche Lage

ZUSAMMENFASSUNG

Das statistische Unternehmensregister ist seit 2021 an den Geokodierungsdienst des Bundesamtes für Kartographie und Geodäsie angebunden, der allen Niederlassungen eine räumliche Lagebeschreibung in Form einer Koordinate zuweist. Neben der Möglichkeit für georäumliche Analysen bietet diese Erweiterung auch die Chance, das statistische Unternehmensregister über Geokoordinaten mit anderen Quellen zu verknüpfen. Der Beitrag erläutert die Methodik dieses Verknüpfungsprozesses anhand des Beispiels des Bundesarztverzeichnisses.

📌 **Keywords:** geocoding – geographic grid cells – medical practice – name comparison – spatial location

ABSTRACT

The Statistical Business Register has been linked to the geocoding service of the Federal Agency for Cartography and Geodesy since 2021. This geocoding service assigns all local units in the Business Register a spatial location description in the form of a coordinate. As well as allowing for geospatial analyses, this new feature opens up the possibility of linking the Statistical Business Register to other data sources via geographic coordinates. This article describes the methodology of this linking process using the example of the Federal Directory of Doctors.

1

Einleitung

Das statistische Unternehmensregister ist eine zentrale Datenbank aller in Deutschland ansässigen Unternehmen mit Angaben zu Umsatz und Beschäftigten, zum Wirtschaftszweig und der Rechtsform der darin enthaltenen Einheiten. Es stellt die Grundgesamtheit für Unternehmensstatistiken dar und ist ein wichtiges Instrument für Auswertungen.

Die Informationen des statistischen Unternehmensregisters werden über mehrere Quellen regelmäßig aktualisiert, im Wesentlichen über Meldungen zur Sozialversicherung bei der Bundesagentur für Arbeit und Umsatzsteuer-Voranmeldungen aus der Finanzverwaltung. Die Daten des Bundesarztverzeichnisses dürfen seit 2021 aufgrund einer Gesetzesnovellierung¹ ebenfalls in das Unternehmensregister eingepflegt werden. Diese neue administrative Quelle wird jährliche Informationen liefern und dabei auch die Untererfassung im Bereich Gesundheitswesen beheben.

Der Mehrwert der neuen Quelle liegt darin, sowohl neue Einheiten für das statistische Unternehmensregister zu erfassen als auch zuverlässige Informationen über bereits im Register geführte Einheiten zu liefern. Bei im statistischen Unternehmensregister bereits geführten Einheiten sind die Informationen aus dem Bundesarztverzeichnis bei der Erstverarbeitung dem Pendant im statistischen Unternehmensregister zuzuordnen. Die Verknüpfung der beiden Datenbestände erfolgt über Namen und Adressen, da keine gemeinsamen Identifikatoren wie beispielsweise ein Handelsregistereintrag, die Umsatzsteuer-Identifikationsnummer oder eine Betriebsnummer in beiden Quellen vorliegen.

Außer mit ihren Adressen werden seit dem Berichtsjahr 2021 alle Niederlassungen im statistischen Unternehmensregister mit Geokoordinaten geführt². Diese Geokoordinaten stammen aus dem Datenbestand des

Bundesamtes für Kartographie und Geodäsie (BKG), welches über Geokoordinaten aller Anschriften in Deutschland verfügt. Nach einem Abgleich der Anschriften im statistischen Unternehmensregister mit dem BKG-Datenbestand werden die entsprechenden Geokoordinatenpaare übernommen. Ein Geokoordinatenpaar entspricht der punktgenauen räumlichen Lagebeschreibung einer Anschrift, das ist oft die Gebäudemitte. Somit können auch zwei Anschriften – wenn zum Beispiel ein Gebäude zwei Eingänge in unterschiedlichen Straßen hat – demselben Geokoordinatenpaar zugewiesen werden. Dem Geokoordinatenpaar können auch geografische Gitterzellen verschiedener Auflösungen zugeordnet werden, um die Lagebeschreibung einer Anschrift nicht nur punktgenau, sondern auch innerhalb einer bestimmten georäumlichen Fläche zu definieren.

Um mehrere Datenbestände zu verknüpfen, sind Geokoordinaten besser geeignet als reine Adressangaben. Durch Geokoordinaten können Datenbestände verknüpft werden, die unterschiedliche Adressen für die gleiche Verortung führen. Außerdem lassen sich die Gitterzellen nutzen, um Einheiten nicht nur am punktgenauen Ort, sondern auch innerhalb eines bestimmten Radius zu suchen. Diese Art von Blocking beschränkt die Anzahl potenzieller Treffer auf Einheiten in derselben räumlichen Lage. Ein solcher Prozess ist für Einheiten gut geeignet, deren räumliche Lage sie eindeutiger identifiziert als ihr Name, zum Beispiel für Arztpraxen³.

Der folgende Beitrag beschreibt, wie Geokoordinaten dazu genutzt werden, um die Daten aus den beiden unterschiedlichen Quellen statistisches Unternehmensregister und Bundesarztverzeichnis miteinander zu verknüpfen. Kapitel 2 stellt die Datengrundlage vor. Kapitel 3 erläutert die Aufbereitung der Daten und die Methodik des Verknüpfungsprozesses. Die Ergebnisse der Verknüpfungen werden in Kapitel 4 zusammengefasst und bewertet. Kapitel 5 schließt den Bericht mit einem Fazit und Ausblick über die weiteren Anwendungsmöglichkeiten des Verknüpfungsprozesses.

1 Änderung des Sozialgesetzbuchs (SGB) Fünftes Buch (V), die durch das Gesetz zur Umsetzung der Verordnung des Europäischen Parlaments und des Rates über europäische Unternehmensstatistiken (EBS-Umsetzungsgesetz) am 3. März 2021 in Kraft getreten ist.

2 Die Speicherung von Geokoordinaten in elektronischen Registern ist durch das Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz – EGovG) geregelt. Für eine ausführliche Diskussion siehe Brenzel/Gebers (2020).

3 Ein Pendant für die „Arztpraxis Dr. Müller“ in München sucht man effizienter über den Standort als über alle Einheiten mit Namen „Müller“.

2

Datengrundlage

2.1 Das statistische Unternehmensregister

Das statistische Unternehmensregister enthält Kernangaben über alle in Deutschland ansässigen Niederlassungen, Rechtlichen Einheiten, Unternehmen und Unternehmensgruppen (Statistische Ämter des Bundes und der Länder, 2022). Eine Arztpraxis entspricht einer Rechtlichen Einheit, die definiert ist als eine wirtschaftlich tätige natürliche Person, eine juristische Person oder eine Personenvereinigung (von Eschwege, 2021). Im statistischen Unternehmensregister werden über 6,3 Millionen Rechtliche Einheiten geführt, davon etwa 350 000 im Wirtschaftsabschnitt „Gesundheits- und Sozialwesen“¹⁴.

2.2 Das Bundesarztverzeichnis

Das Bundesarztverzeichnis enthält Angaben über alle Ärztinnen und Ärzte sowie alle Psychotherapeutinnen und Psychotherapeuten (im Folgenden auch kurz: Ärztinnen und Ärzte), die Versicherte der gesetzlichen Krankenkassen ambulant behandeln dürfen (Kassenärztliche Bundesvereinigung, 2020).

Eine Änderung im SGB V ermöglicht, das am 3. März 2021 in Kraft getretene Gesetz zur Umsetzung der Verordnung des Europäischen Parlaments und des Rates über europäische Unternehmensstatistiken (EBS-Umsetzungsgesetz) zu erfüllen. Diese Änderung regelt, dass die Kassenärztliche Bundesvereinigung dem Statistischen Bundesamt jährlich auf Anforderung bestimmte Daten aus dem Bundesarztverzeichnis übermittelt. Neben Namen, Arztnummer, Teilnahmestatus und Fachrichtung der Ärztinnen und Ärzte werden die Identifikationsnummer und Adressen der Praxis übermittelt, in der die Ärztin oder der Arzt tätig ist. Für das Berichtsjahr 2021 enthält das Bundesarztverzeichnis Angaben über insgesamt 105 966 Praxen.

¹⁴ Wirtschaftsabschnitt Q der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (Statistisches Bundesamt, 2008).

2.3 Geokoordinaten des Bundesamtes für Kartographie und Geodäsie

Der Geokodierungsdienst für Adressen und Geonamen ist ein Gemeinschaftsprojekt von Bund und Ländern für amtliche Zwecke. Das Bundesamt für Kartographie und Geodäsie bereitet die Daten auf und betreibt den an seinem Dienstleistungszentrum entwickelten Webdienst. Datengrundlage des Dienstes sind die Datenbestände „Amtliche Hauskoordinaten Deutschlands (HK-DE)“ der deutschen Landesvermessung und „Geographische Namen Deutschlands (GN250)“ des Bundesamtes für Kartographie und Geodäsie.

Der Geokodierungsdienst ist ein Webdienst, der Objekten (in diesem Falle Adressen) eine räumliche Lagebeschreibung in Form einer Koordinate zuweist. Der Dienst arbeitet fehlertolerant und belegt die gefundenen Objekte mit einem Score. Dieser Score gibt die Qualität und die Art der Zuordnung (hausgenau, straßengenau, ortsgenau) wieder und bewertet so die Genauigkeit ihrer Übereinstimmung mit der Suchanfrage. Die Objekte können dabei eindeutig und sicher einer entsprechenden Anschrift und den dazugehörigen Koordinaten zugeordnet werden. Ebenso weist der Geokodierungsdienst auf mehrdeutige Angaben hin, etwa bei Hausnummernspannen. Bei falschen oder fehlenden Adressteilen vergibt das System immer die nächstbeste Geokodierung, zum Beispiel bei fehlender oder falscher Hausnummer die Mitte der Straße. Diese Koordinate wird dann mit der Kennzeichnung „straßengenau“ versehen.

3

Methodik der Verknüpfung

3.1 Aufbereitung der Daten mit Geokoordinaten und geografischen Gitterzellen

Das Bundesarztverzeichnis wurde mit der Bundeskopie des statistischen Unternehmensregisters für das Berichtsjahr 2021⁵ verknüpft. Das statistische Unternehmensregister führt für jede geokodierte Anschrift sowohl die metergenauen x- und y-Koordinaten als auch die 100 m x 100 m große geografische Gitterzelle, in der das Koordinatenpaar liegt. Abgeleitet von den x- und y-Koordinaten wurde die Bundeskopie zudem mit der 1 000 m x 1 000 m großen geografischen Gitterzelle erweitert.

Die Adressen im Bundesarztverzeichnis für das Berichtsjahr 2021 wurden ebenfalls über den Geokodierungs-

⁵ Die Bundeskopie gibt den zum Gültigkeitsstand 31. Dezember eines Berichtsjahres t „eingefrorenen“ Stand des statistischen Unternehmensregisters wieder. Sie wird zum Bearbeitungsstand 30. September $t+1$ (für Niederlassungen und Rechtliche Einheiten) beziehungsweise 30. April $t+2$ (für Unternehmen) des Unternehmensregisters erstellt.

dienst des Bundesamtes für Kartographie und Geodäsie mit ihren entsprechenden x- und y-Koordinaten erweitert. Dabei konnten 95% der Adressen einer gebäudescharfen Geokodierung zugeordnet werden, die restlichen Adressen konnten zumindest ortsgenau georeferenziert werden. Aus diesen Koordinaten wurde das Bundesarztverzeichnis analog zu der Bundeskopie des statistischen Unternehmensregisters mit den 100 m x 100 m und 1 000 m x 1 000 m großen Gitterzellen erweitert.

↳ Geografische Gitterzellen

Eine geografische Gitterzelle ist eine Gebietseinheit, die bezogen auf eine vorgegebene Kartenprojektion quadratisch und mindestens 1 Hektar groß ist (Gebers/Graze, 2019). Der Geokodierungsdienst des Bundesamtes für Kartographie und Geodäsie nutzt das metrische Koordinatenreferenzsystem ETRS89-LAEA (Lambert Azimuthal Equal Area), eine flächentreue azimutale Projektion, die die Abbildung von Gitterzellen unterschiedlicher Auflösungen ermöglicht. Das System weist jeder Anschrift metergenaue x- und y-Koordinaten zu. Aus diesen Koordinaten können Gitterzellen in verschiedenen Zellgrößen definiert werden, in denen das Koordinatenpaar liegt. Die Gitterzelle ist mit einer Gitter-ID gekennzeichnet, die aus einer Beschreibung der Weite der Zelle (zum Beispiel 100 m, 1 000 m) gefolgt von den x- und y-Koordinaten der südwestlichen Ecke der Gitterzelle besteht. ↳ Grafik 1

Grafik 1
Geokoordinatenpaar und 100 m x 100 m geografische Gitterzelle



Quelle: Bundesamt für Kartographie und Geodäsie 2023

2023 - 122

Tabelle 1

Verknüpfung des Bundesarztverzeichnisses mit dem statistischen Unternehmensregister für das Berichtsjahr 2021 nach der Art der Verknüpfung

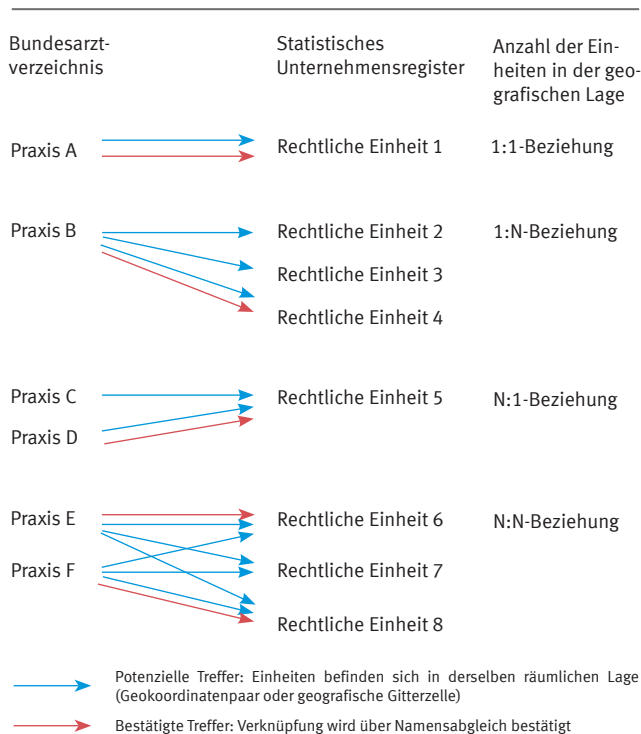
	Geokoordinatenpaar	100 m x 100 m geografische Gitterzelle	1 000 m x 1 000 m geografische Gitterzelle
Noch nicht verknüpfte Einheiten	105 966	26 655	24 480
Mögliche Treffer über Geokoordinaten identifiziert	100 605	22 848	24 466
Treffer über Namensabgleich bestätigt	79 311	2 175	2 337
darunter:			
1:N-Verknüpfungen	4 756	86	89
N:1-Verknüpfungen	3 398	57	34
Treffer über Namensabgleich nicht bestätigt	21 294	23 575	22 129
Kein möglicher Treffer über Geokoordinaten identifiziert	5 361	905	14

3.2 Verknüpfung über Geokoordinatenpaare

Im ersten Schritt wurden die beiden Datenbestände über die Geokoordinatenpaare verknüpft. Für 100 605 der 105 966 Praxen im Bundesarztverzeichnis konnte in der Bundeskopie mindestens eine Rechtliche Einheit über das Geokoordinatenpaar zugeordnet werden; das heißt, an der hausgenauen Verortung existiert im statistischen Unternehmensregister mindestens eine Rechtliche Einheit. Für 5 361 Praxen gibt es an den entsprechenden Geokoordinaten keine Rechtliche Einheit. [↘ Tabelle 1](#)

Für die Teilmenge der Einheiten, die über die Geokoordinatenpaare verknüpft werden, befinden sich sowohl 1:1-Beziehungen (an diesen Geokoordinaten existiert seitens des Bundesarztverzeichnisses und seitens der Bundeskopie nur eine Einheit) als auch Mehrfachbeziehungen (an diesen Geokoordinaten existieren mehrere Einheiten). Im ersten Fall muss die Richtigkeit des Treffers bestätigt werden, im zweiten Fall muss für jede Einheit im Bundesarztverzeichnis aus der Menge potenzieller Treffer das richtige Pendant im statistischen Unternehmensregister identifiziert werden. Dies erfolgt über einen Namensabgleich. [↘ Grafik 2](#)

Grafik 2
Der Verknüpfungsprozess



2023 - 123

3.3 Namensabgleich

Für den Namensabgleich wurden im Bundesarztverzeichnis alle Namen der Inhaberinnen und Inhaber der Arztpraxen⁶ standardisiert. Dies bedeutet beispielsweise, in Großbuchstaben zu schreiben, Umlaute und ß zu ersetzen und Titel zu entfernen. Das statistische

⁶ Diese Personen sind über ihren Teilnahmestatus „Zulassung nach §§ 19/24 SGB V“ identifizierbar. Angestellte Ärztinnen und Ärzte werden im Namensabgleich nicht berücksichtigt.

Unternehmensregister enthält die Bezeichnungen der rechtlichen Träger, die neben den Namen der Inhaberrinnen und Inhaber oft zusätzliche Informationen über die Art (zum Beispiel Hausarztpraxis), Fachrichtung (zum Beispiel Augenarzt) und Rechtsform (zum Beispiel Gesellschaft bürgerlichen Rechts) der Praxis aufweisen. Diese Begriffe wurden entfernt und die Bezeichnung des rechtlichen Trägers analog zum Bundesarztverzeichnis standardisiert.

Der Namensabgleich erfolgt in mehreren Schritten über einen Algorithmus. Zuerst werden die standardisierten Namen verglichen. Gibt es keine Übereinstimmung der standardisierten Namen, werden weitere Namensbestandteile betrachtet sowie die Reihenfolge (Vor- und Nachnamen) variiert. Ein bestätigter Treffer liegt vor, wenn mindestens ein Name des Bundesarztverzeichnisses in der Bezeichnung der Rechtlichen Einheit gefunden wurde.

Der Algorithmus erlaubt hierbei nur eine exakte Übereinstimmung der Zeichenketten. Damit das Verfahren auch mit kleinen Rechtschreibunterschieden wie „Mustermann“ statt „Musterman“ in den Namen umgehen kann, wird für jede Namenskombination die Levenshtein-Distanz berechnet. Diese Distanz ermittelt die Ähnlichkeit zweier Zeichenketten durch die Mindestanzahl der Änderungsoperationen, die notwendig sind, um die erste Zeichenkette in die zweite umzuwandeln. Bei einer Levenshtein-Distanz von 2 oder kleiner wird der Name als übereinstimmend angesehen und als bestätigter Treffer gekennzeichnet.¹⁷

Über den Namensabgleich konnten 79 311 Praxen des Bundesarztverzeichnisses einer Rechtlichen Einheit des statistischen Unternehmensregisters zugeordnet werden. Für 21 294 Praxen war es nicht möglich, über den Namensabgleich-Algorithmus ein Pendant zuzuordnen. Für diese Praxen und auch die 5 361 Praxen ohne Treffer beim Abgleich über die Geokoordinaten wurde im nächsten Schritt über geografische Gitterzellen nach einer Zuordnung gesucht.

¹⁷ Mit einer Levenshtein-Distanz von 2 sind typische Rechtschreibunterschiede abgedeckt, ohne unterschiedliche Namen wie „Max Musterman“ und „Uwe Mustermann“ (berechtigte Levenshtein-Distanz von 3) fälschlicherweise als Übereinstimmung zu kennzeichnen.

3.4 Verknüpfung über geografische Gitterzellen

Im zweiten Schritt wurden die 26 655 noch nicht verknüpften Praxen über die 100 m x 100 m große geografische Gitterzelle zusammengeführt. Hierbei wurde für 22 848 Praxen mindestens eine Rechtliche Einheit innerhalb der entsprechenden geografischen Gitterzelle gefunden. Bei 2 175 Praxen bestätigt der Namensabgleich die Zuordnung zur Rechtlichen Einheit.

Im letzten Schritt wurde für die 24 480 noch nicht verknüpften Praxen über die 1 000 m x 1 000 m große geografische Gitterzelle nach einem Pendant im statistischen Unternehmensregister gesucht. Der Namensabgleich ergab für weitere 2 337 Praxen einen Treffer in der Bundeskopie des statistischen Unternehmensregisters.

4

Ergebnisse

4.1 Analyse der Verknüpfungen

Insgesamt konnte mithilfe von Geokoordinaten und Namensabgleich für 83 823 Praxen (79%) des Bundesarztverzeichnisses eine Rechtliche Einheit im statistischen Unternehmensregister gefunden werden. Für die Mehrheit der verknüpften Praxen (95%) erfolgte die Zuordnung über das Geokoordinatenpaar. Für 5,3% der Praxen wurde der Suchradius erweitert, um einen Treffer zu finden. Grundsätzlich wären zusätzliche Verknüpfungen über noch größere geografische Gitterzellen denkbar. Jedoch steigt die Anzahl der Einheiten, die in den Namensabgleich kommen, wodurch die Genauigkeit des Verfahrens sinkt (siehe Kapitel 5).

In der Teilmenge der über Namensabgleich bestätigten Treffer befinden sich 1:N-Verknüpfungen (einer Praxis im Bundesarztverzeichnis wurden zwei oder mehr Rechtliche Einheiten im statistischen Unternehmensregister zugeordnet) und N:1-Verknüpfungen (zwei oder mehr Praxen im Bundesarztverzeichnis wurden der gleichen Rechtlichen Einheit im statistischen Unternehmensregister zugeordnet). Ursache für diese unscharfen Zuordnungen ist zum Teil die unterschiedliche Abbil-

derung von Praxisgemeinschaften⁸ im Bundesarztverzeichnis und im statistischen Unternehmensregister. Durch die Daten des Bundesarztverzeichnisses und durch weitere fachliche Prüfungen ist es möglich, Praxisgemeinschaften im statistischen Unternehmensregister zu identifizieren und statistisch abzubilden.

Eine weitere Kooperationsform, das medizinische Versorgungszentrum, stellt ebenfalls einen Sonderfall dar. Hierbei handelt es sich um eine ärztlich geleitete Einrichtung, die eine fachübergreifende Versorgung aus einer Hand ermöglicht. Im medizinischen Versorgungszentrum können zugelassene und/oder angestellte Ärztinnen und Ärzte tätig werden. Manchmal kommen die Namen der zugelassenen Ärztinnen oder Ärzte im Namen des medizinischen Versorgungszentrums vor, dann ist eine Bestätigung des Treffers über den Namensabgleich möglich. Viele medizinische Versorgungszentren sind jedoch allgemeiner benannt (beispielsweise „MVZ Musterort GmbH“), sodass der Namensabgleich-Algorithmus keinen Treffer bestätigen kann. Von den 4 096 im Berichtsjahr 2021 im Bundesarztverzeichnis geführten medizinischen Versorgungszentren war für nur 37 % über Geokoordinaten und Namensabgleich ein Pendant im statistischen Unternehmensregister zu finden.

Nach dem Verknüpfungsprozess bleiben insgesamt 22 143 Praxen im Bundesarztverzeichnis ohne ein entsprechendes Pendant in der Bundeskopie des statistischen Unternehmensregisters. Davon sind 2 579 medizinische Versorgungszentren ohne Arztnamen in der Bezeichnung, sodass der Namensabgleich-Algorithmus nicht greifen kann. Bei den restlichen Einheiten handelt es sich entweder um Praxen, die im statistischen Unternehmensregister zwar vorhanden sind, aber über den Verknüpfungsprozess nicht gefunden wurden, oder um Praxen, die tatsächlich (noch) nicht im statistischen Unternehmensregister geführt werden. Bei Letzteren sollte eine Aufnahme geprüft werden, um so die Qualität des statistischen Unternehmensregisters zu steigern.

⁸ Zwei oder mehr Einzelpraxen oder Berufsausübungsgemeinschaften, die Räume, Geräte und Personal teilen.

4.2 Qualitätsprüfung des Verknüpfungsprozesses

Um die Qualität des Verknüpfungsprozesses einzuschätzen, wurden die Verknüpfungen in einer 1-%-Stichprobe manuell überprüft. [↘ Tabelle 2](#) stellt die Ergebnisse dieser Prüfungen als Konfusionsmatrix dar.

Tabelle 2
Konfusionsmatrix zur Qualitätsprüfung des Verknüpfungsprozesses

Vorhersage \ Richtigkeit	Treffer	Kein Treffer
Treffer	908 (richtig Treffer)	30 (fälschlicherweise Treffer)
Kein Treffer	60 (fälschlicherweise kein Treffer)	162 (richtig kein Treffer)

Einfache 1-%-Zufallsstichprobe des Bundesarztverzeichnisses (1 160 Einheiten).

Eine Konfusionsmatrix zeigt, inwieweit die vorhergesagten Zuordnungen (in diesem Fall „Treffer“ oder „kein Treffer“ laut Verknüpfungsprozess) richtig oder falsch sind. Aus der Konfusionsmatrix können unterschiedliche Kennzahlen berechnet werden, um die Qualität der Zuordnungen zu beurteilen. Die Kennzahl „Genauigkeit“⁹ beschreibt die insgesamt richtigen Zuordnungen („richtig Treffer“ und „richtig kein Treffer“) im Verhältnis zu allen Zuordnungen. In der untersuchten 1-%-Stichprobe liegt sie bei 0,92, das heißt 92 % der Praxen wurden richtig zugeordnet.

Alle 30 Fälle der fälschlicherweise als Treffer identifizierten Praxen wurden über die Nachnamen zugeordnet. Aufgrund der Berufsausübungsgemeinschaften, bei denen oft nur Nachnamen in der Bezeichnung vorkommen, muss der Namensabgleich-Algorithmus diese Unschärfe zulassen.

⁹ Die „Sensitivität“ (im Englischen „Recall“) beschreibt, inwieweit positive Fälle richtig erkannt werden. Diese Kennzahl wird oft für Klassifikationsmodelle genutzt, bei denen es am wichtigsten ist, positive Fälle zu erkennen, auch wenn negative Fälle fälschlicherweise als positiv klassifiziert werden (zum Beispiel bei der Erkennung von Krankheiten). Bei den hier vorliegenden Verknüpfungsprozessen hingegen sollen auch falsche Verknüpfungen vermieden werden, deshalb wird nach Genauigkeit bewertet.

4.3 Bewertung

Insgesamt konnte über Geokoordinaten und anschließenden Namensabgleich für 79% der Praxen im Bundesarztverzeichnis ein Pendant in der Bundeskopie des statistischen Unternehmensregisters gefunden werden. Qualitätsprüfungen auf Basis einer 1%-Stichprobe weisen eine hohe Genauigkeit von 92% korrekten Verknüpfungen auf.


Der Prozess ist unter anderem deswegen effizient, weil die anfängliche geokodierte Verknüpfung die Anzahl potenzieller Treffer, für die der Namensabgleich nötig ist, minimiert. Durch mehrere, nacheinander geschaltete Abgleiche können der Namensabgleich-Algorithmus je nach gewünschter Treffsicherheit angepasst oder die einzelnen Bestätigungen unterschiedlich behandelt werden: So könnte zum Beispiel die Verknüpfung für 57 370 (68%) bestätigte Treffer, bei denen es sich um Einzelpraxen handelt, für die nur eine Rechtliche Einheit an den hausgenauen Geokoordinaten in der Bundeskopie gefunden wurde und bei denen der Name des einzelnen Inhabers oder der einzelnen Inhaberin in beiden Datenbeständen komplett übereinstimmt, automatisiert erfolgen. Dagegen könnten diejenigen für eine fachliche Sichtung aufbereitet werden, die noch über Nachnamen bestätigt werden müssen. Der Algorithmus kann auch angepasst werden, um beim Abgleich weitere Merkmale wie die Fachrichtung zu berücksichtigen. Bei Sonderfällen wie Praxisgemeinschaften oder medizinischen Versorgungszentren könnte man nach Kriterien filtern, die für eine Praxisart spezifisch sind. Praxisgemeinschaften werden beispielsweise meist in der Rechtsform Gesellschaft bürgerlichen Rechts (GbR) betrieben. Ein solches Vorgehen könnte die Trefferquote erhöhen.

Den Erfolg der Verknüpfung über die Geokoordinaten garantiert auch die Qualität der Anschriften im Bundesarztverzeichnis: Alle Adressen sind vollständig und postalisch korrekt abgebildet. Somit konnte der Geokodierungsdienst des Bundesamtes für Kartographie und Geodäsie für 95% der Adressen eine gebäudescharfe Geokodierung vornehmen. Bei schlechterer Qualität hätte der Geokodierungsdienst im besten Fall eine straßen- oder ortsgenaue, im schlimmsten Fall keine Geokoordinaten zuordnen können. Mit einer straßen- oder ortsgenaue Geokodierung ist eine Verknüpfung über Geokoordinaten noch möglich, aber aufgrund der

höheren Zahl der zu vergleichenden Fälle und der erforderlichen größeren Suchradien würden sowohl Effizienz als auch Genauigkeit sinken.

5

Fazit und Ausblick

Der Statistische Verbund¹⁰ nutzt geografische Informationen bereits bei Produkten wie dem Krankenhausatlas oder dem Unfallatlas (Brenzel/Gebers, 2020). Dieser Beitrag zeigt eine zusätzliche Verwendung im Bereich der Data Linkages. Besonders für Quellen, bei denen die Suche nach einem Pendant nur anhand von Namen und räumlicher Lage gesteuert werden kann, hat die geografische Information gegenüber der Nutzung reiner Adressangaben Vorteile und führt zu einer höheren Trefferquote. Dies konnte anhand einer Verknüpfung des statistischen Unternehmensregisters mit dem Bundesarztverzeichnis gezeigt werden. Die in diesem Beitrag beschriebene Methodik ist individuell gestaltbar und lässt sich auf andere Anwendungszwecke erweitern. 

10 Den Statistischen Verbund bilden die Statistischen Ämter des Bundes und der Länder.

LITERATURVERZEICHNIS

Brenzel, Hanna/Gebers, Kathrin. *Werkstattbericht: Georeferenzierung im Statistischen Verbund*. In: WISTA Wirtschaft und Statistik. Ausgabe 6/2020, Seite 48 ff.

Doll, Hendrik/Gábor-Tóth, Eniko/Schild, Christopher-Johannes. *Linking Deutsche Bundesbank Company Data, Technical Report 2021-05*. Version v2021-2-6. 2021. Deutsche Bundesbank, Research Data and Service Centre. [Zugriff am 20. Juli 2023]. Verfügbar unter: www.bundesbank.de

von Eschwege, Katja. *Neuerungen im statistischen Unternehmensregister: Auswertungskonzept, Relevanzschwellen und weitere Quellen*. In: WISTA Wirtschaft und Statistik. Ausgabe 5/2021, Seite 86 ff.

Gebers, Kathrin/Graze, Philip. *Statistische Datengewinnung durch die Nutzung geografischer Informationen*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2019, Seite 11 ff.

Kassenärztliche Bundesvereinigung. *Erläuterungen zu den Statistischen Informationen aus dem Bundesarztregister*. 2020. [Zugriff am 13. Juli 2023]. Verfügbar unter: www.kbv.de

Kassenärztliche Bundesvereinigung. *Medizinische Versorgungszentren aktuell*. 2022. [Zugriff am 18. Juli 2023]. Verfügbar unter: www.kbv.de

Kassenärztliche Bundesvereinigung. *Optionen & Kooperationsformen*. 2023. [Zugriff am 18. Juli 2023]. Verfügbar unter: www.kbv.de

Statistische Ämter des Bundes und der Länder. *Methodische Grundlagen, Definitionen und Qualität des statistischen Unternehmensregisters*. 2022. [Zugriff am 12. Juli 2023]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Gliederung der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)*. Wiesbaden 2008.

Statistisches Bundesamt. *Glossar zur Kostenstrukturstatistik im medizinischen Bereich*. 2023. [Zugriff am 12. Juli 2023]. Verfügbar unter: www.destatis.de

RECHTSGRUNDLAGEN

Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz – EGovG) vom 25. Juli 2013 (BGBl. I Seite 2749), das zuletzt durch Artikel 1 des Gesetzes vom 16. Juli 2021 (BGBl. I Seite 2941) geändert worden ist.

Gesetz zur Umsetzung der Verordnung des Europäischen Parlaments und des Rates über europäische Unternehmensstatistiken zur Aufhebung von zehn Rechtsakten im Bereich Unternehmensstatistiken und zur Änderung anderer Statistikgesetze (EBS-Umsetzungsgesetz) vom 22. Februar 2021 (BGBl. I Seite 266).

Sozialgesetzbuch (SGB) Fünftes Buch (V) – Gesetzliche Krankenversicherung – (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I Seite 2477), das zuletzt durch Artikel 9 des Gesetzes vom 16. August 2023 (BGBl. I Nr. 217) geändert worden ist.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Daniel Vorgrimler

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im Oktober 2023

Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-23005-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2023

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.