

Neue digitale Daten und Methoden in der amtlichen Statistik

Ein Überblick

Kevin Heinen

Die zunehmende Digitalisierung des gesellschaftlichen und wirtschaftlichen Lebens eröffnet der amtlichen Statistik neue Perspektiven in den Bereichen Erhebung, Aufbereitung, Analyse und Verbreitung von Daten. Vor dem Hintergrund der gesetzlichen Aufgabe der amtlichen Statistik, qualitativ hochwertige Daten nach den Grundsätzen der Neutralität, Objektivität und fachlichen Unabhängigkeit bereitzustellen, stellen diese Entwicklungen jedoch auch eine große Herausforderung dar. Im Folgenden soll ein Einblick in die Thematik „Neue digitale Daten in der amtlichen Statistik“ gegeben werden. Hierzu werden exemplarisch einige neue digitale Datenquellen bzw. Methoden vorgestellt.

Big Data und die amtlichen Statistik

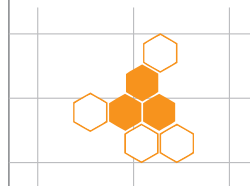
Die Art und Weise wie wir arbeiten, leben und kommunizieren befindet sich im stetigen Wandel. Aktuell spielt insbesondere der digitale Wandel eine entscheidende Rolle. So lassen sich in den letzten Jahren einige Veränderungen in der Form unserer Kommunikation feststellen. Auf der einen Seite steht die sogenannte Mensch-zu-Mensch-Kommunikation, deren Veränderung sich durch den Wandel der Kommunikationsmittel ausdrücken lässt. E-Mails, mobile Endgeräte sowie soziale Medien bzw. Netzwerke sind hierfür einige Beispiele. Des Weiteren hat die sogenannte Mensch-zu-Maschine-Kommunikation an Bedeutung gewonnen. Hier entwickelten sich im geschäftlichen Alltag diverse Substitute. Für den Bankschalter der Geldautomat oder für das Reisebüro verschiedene Internetportale. Darüber hinaus gewinnt die Maschine-zu-Maschine-Kommunikation immer mehr an Bedeutung.¹ Mittels solcher Kommunikationsformen werden weltweit immer mehr Daten generiert.

Die so entstehenden Daten, die sich neben ihrer Breite ebenfalls durch die Frequenz ihrer Datenproduktion auszeichnen, werden dabei allgemein unter dem Begriff Big Data zusammengefasst.² Solche neuen digitalen Daten bieten der amtlichen Statistik als Informationsdienstleister für die Öffentlichkeit, Politik und Wirtschaft neue Möglichkeiten in

den Bereichen der Datenerhebung, -analyse und -veröffentlichung. Für die amtliche Statistik sind die Daten von besonderem Interesse, die ihr Informationsangebot sinnvoll ergänzen bzw. erweitern können. Aktuelle Projekte der amtlichen Statistik beschäftigen sich unter anderem mit den Nutzungsmöglichkeiten von Mobilfunkdaten für Mobilitätsanalysen und Schätzungen der Bevölkerungszahlen, mit der Nutzung von Satellitendaten, zum Beispiel im Rahmen der Landwirtschaftsstatistiken und der Konjunkturschätzungen, mit Scannerdaten sowie verschiedenen Daten, die mittels Web Scraping von Internetseiten erhoben werden können. Im Folgenden werden die Chancen und Herausforderungen der Implementierung neuer digitaler Daten und Methoden exemplarisch an den Scanner- und Mobilfunkdaten sowie des Web Scrapings gezeigt.

Scannerdaten für die amtliche Statistik

Eine vielversprechende Datenquelle für die amtliche Preisstatistik sind die sogenannten Scannerdaten. Hierbei handelt es sich um digital verfügbare Kassendaten des Einzelhandels. Diese Transaktionsdaten beinhalten Angaben über Absatz, Umsatz und Art der verkauften Artikel sowie weitere Produktangaben. Dadurch eignen sich Scannerdaten für eine Vielzahl an statistischen Nutzungsmöglichkeiten. Die wichtigste hierbei ist die Verwendung innerhalb der Verbraucherpreisstatistik zur Ermittlung des Verbraucherpreisindex (VPI), denn im Vergleich zur traditionellen Preiserfassung, die auf einer repräsentativen Stichprobe basiert, bieten Scannerdaten eine umfangreichere zeitliche Perspektive.³ Dies ist im Anbetracht der stetigen Veränderungen im Umfeld der Verbraucherpreisstatistik von enormer Bedeutung. Der Vorteil der Scannerdaten besteht darin, dass sich die Zahl der Preisbeobachtungen über einem längeren Zeitraum deutlich erhöhen lässt. Die sich daraus ergebenden Durchschnittspreise ermöglichen es dann, Preisschwankungen einzelner Produkte detailliert zu erfassen und zu berücksichtigen. Weitere mögliche Analysen mit



Kevin Heinen M. Sc. ist Referent im Referat „Unternehmensregister, Gewerbeanzeigen, Verkehr“ des Statistischen Landesamtes Baden-Württemberg.

- 1 Wiengarten, Lara/Zwick, Markus (2017): Neue digitale Daten in der amtlichen Statistik, in: WISTA Wirtschaft und Statistik. (5), S.19–30.
- 2 Daas, Piet/Puts, Marco/Buelens, Bart/van den Hurk, Paul (2013): Big Data and Official Statistics, https://ec.europa.eu/eurostat/cros/content/big-data-and-official-statistic-piet-daas-marco-puts-and-bart-buelens-paul-van-den-hurk_en (Abruf: 10.09.2022).
- 3 Koch, Julia/Erdemsiz, Baran (2020): Einsatz von Scannerdaten während der Covid-19-Pandemie, in: WISTA Wirtschaft und Statistik. (4), S.96–06.

tels Scannerdaten, wären beispielsweise die Analyse der Preisentwicklung einzelner Produkttypen oder des Kaufverhaltens. Darüber hinaus sind tiefgegliederte regionale Auswertungen von Preisen mit Scannerdaten denkbar.⁴ Zusammenfassend bieten Scannerdaten die Möglichkeit, den VPI digital und zukunftsorientiert zu gestalten, wodurch sich die Genauigkeit sowie Qualität des Indexes auf Dauer sichern lässt. Dennoch sind bei der Implementierung von Scannerdaten in die amtliche Statistik Herausforderungen methodischer, organisatorischer und technischer Art zu beachten. Hierzu zählt unter anderem die zeitintensive Zuordnung einzelner Artikel zur jeweiligen VPI-Klassifikation oder die aktuell noch eingeschränkte Datenbasis.⁵

Mobilfunkdaten

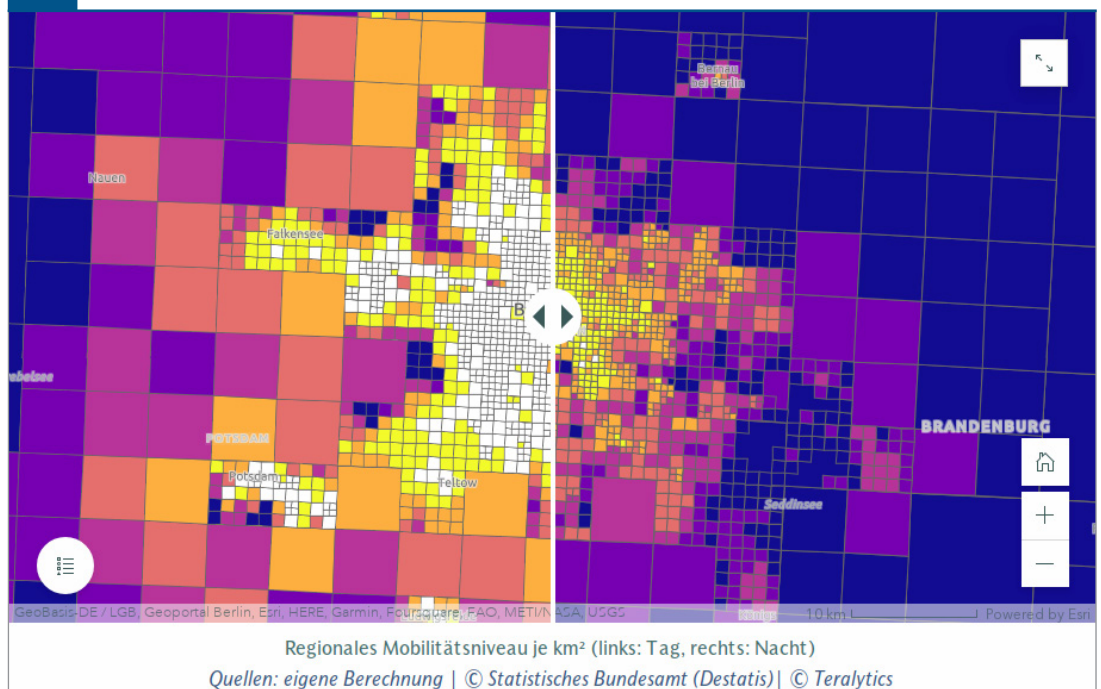
Eine weitere, für die amtliche Statistik interessante Datenquelle ergibt sich aus Daten, die bei der Nutzung von Mobilfunkgeräten entstehen. Allgemein lassen sich, durch die Verbindung zwischen mobilen Endgeräten und Funkmasten, der Aufenthaltsort sowie die Aufenthaltsdauer der mobilen Endgeräte, und damit ihrer Nutzerinnen und Nutzer, ermitteln. Das Potenzial von Mobilfunkdaten für die amtliche Statistik wird bei der Betrachtung der Verbreitung von Handys in Deutschland deutlich. So besaßen 2021 ungefähr 97,6 % der privaten Haushalte in Deutschland mindestens ein Mobiltelefon, wovon mehr als die Hälfte Smart-

phones waren.⁶ Die sich durch die Nutzung solcher Endgeräte entstehenden Bewegungs- und Aufenthaltsdaten lassen sich hierbei generell in statische und dynamische Daten unterteilen, die unter anderem zur Darstellung der Mobilität der Bevölkerung genutzt werden können.⁷ Im Rahmen der Corona-Pandemie kam ein zusätzliches Interesse an Daten auf, die die Mobilität der Bevölkerung aktuell und möglichst exakt abbilden konnte. Hierbei boten aggregierte und anonymisierte Mobilfunkdaten die Möglichkeit, etwaige Kontaktbeschränkungsmaßnahmen kleinräumig sowie tagesaktuell abzubilden.⁸ Das Statistische Bundesamt veröffentlichte hierzu unter der Rubrik „Experimentelle Daten“ verschiedene Mobilitätsindikatoren.⁹ Zum Beispiel werden kleinräumige Mobilfunkdaten kartografisch in Form von interaktiven Karten visualisiert, wo unter anderem das unterschiedliche Mobilitätsverhalten am Tag und in der Nacht nebeneinander ersichtlich wird (siehe Abbildung)¹⁰.

Eine weitere denkbare Nutzung dieser Daten ist eine dynamischere und aktuellere Abbildung der Bevölkerungszahlen in Deutschland. Erste Fallstudien hierzu zeigen auf, dass Mobilfunkdaten grundsätzlich geeignet sind, die Bevölkerung abzubilden. Die bisherigen Anwendungsbeispiele von Mobilfunkdaten in der amtlichen Statistik zeigen zum einen das breite Spektrum an Verwendungsmöglichkeiten dieser Daten auf. Andererseits müssen etwaige Saison- bzw. Kalendereffekte in der Analyse mitberücksichtigt werden und metho-

- 4 Siehe Statistisches Bundesamt <https://www.destatis.de/DE/Service/EXDAT/Datensaetze/auswertung-scannerdaten.html;jsessionid=96E43E79C949AC793B2ED37167613A2B.live712> (Abruf: 07.10.2022).
- 5 Bieg, Matthias (2019): Nutzung von Scannerdaten in der Preisstatistik – Eine Untersuchung anhand von Marktforschungsdaten, in: WISTA Wirtschaft und Statistik. (2), S.25–37.
- 6 Siehe Laufende Wirtschaftsrechnung <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Ausstattung-Gebrauchsgueter/Tabellen/a-infotechnik-d-lwr.html> (Abruf: 25.10.2022).
- 7 Siehe Statistisches Bundesamt <https://www.destatis.de/DE/Service/EXDAT/Datensaetze/mobilfunkdaten.html> (Abruf: 13.09.2022).
- 8 Bohnensteffen, Sarah/ Mühlhan, Jannek/ Saidani, Younes (2021): Mobilität während der Corona-Pandemie – Ausgewählte Analysen auf Basis von Mobilfunkdaten, in: WISTA Wirtschaft und Statistik. (3), S.89–103.
- 9 Siehe Statistisches Bundesamt <https://www.destatis.de/DE/Service/EXDAT/Datensaetze/mobilitaetsindikatoren-mobilfunkdaten.html;jsessionid=CB460A0276223E23EF7EB7DFA22C2E09.live711> (Abruf: 05.10.2022).
- 10 Quelle: Mobilitätsindikatoren auf Basis von Mobilfunkdaten, Statistisches Bundesamt (Destatis), 2022. <https://storymaps.arcgis.com/stories/58b7b89db42749e69b611a4764870546> (Abruf: 02.12.2022).

A Beispiel einer interaktiven Visualisierung von Mobilfunkdaten



dische Annahmen getroffen werden. Hinzu kommt die derzeit eingeschränkte Datengrundlage, da nur jeweils Informationen von einzelnen Mobilfunkanbietern bisher genutzt werden konnten. Des Weiteren gilt es, notwendige datenschutzrechtliche Regelungen zu beachten, sodass keine Rückschlüsse auf ein einzelnes Gerät bzw. eine einzelne Person möglich sind.

Web Scraping als neue Methode in der amtlichen Statistik

Neben neuen digitalen Daten ergeben sich aus der Digitalisierung auch neue Methoden für die Bereitstellung und Nutzung dieser Daten. Das Internet bietet eine Vielzahl an neuen Datenquellen. So sammelt sich ein breites Spektrum an Informationen in Internetportalen oder auf Webseiten, welches als ergänzendes oder zusätzliches Datenmaterial für die amtliche Statistik denkbar ist. Herausfordernd hierbei ist, diese Informationen mit einem angemessenen Aufwand zu erheben und zu verarbeiten. Neben manuellen Erhebungen im Internet, die zeitlich aufwendig sind, existieren automatisierte Methoden, die generell unter dem Begriff Web Mining zusammengefasst werden. Sie bieten alternative sowie effiziente Ansätze der Datenerhebungen im Internet.¹¹

Eine dieser Methoden, namens Web Scraping, steht hierbei für die amtliche Statistik im besonderen Fokus. Dabei handelt es sich um eine Methode der automatisierten Extraktion und Speicherung von Daten aus dem Internet. Allgemein beschrieben, umfasst dieses Verfahren Prozesse, die es ermöglichen, unstrukturierte Informationen von Internetseiten zu extrahieren und diese für die Weiterverarbeitung verständlich zu strukturieren.¹² Für die Implementierung solcher Verfahren müssen rechtliche Rahmenbedingungen sowie die nötige IT-Infrastruktur geschaffen werden. Ebenso gilt es, die technische bzw. fachliche Wartung zu berücksichtigen, die mit einem gewissen zeitlichen Aufwand verbunden sind.

Die Bedeutung des Web Scrapings für die amtliche Statistik lässt sich beispielsweise anhand der Betrachtung der Preisstatistik aufzeigen, da die Preissetzung im Internet durch die Ausweitung des Online-Handels zunehmend an Bedeutung gewonnen hat. Zusätzlich sorgt zum Beispiel die dort eingesetzte dynamische Preissetzung für eine gestiegene Volatilität der Preise. Für die Ermittlung der Preisentwicklung seitens der amtlichen Statistik ergibt sich daraus die Aufgabe, solche Preisdynamiken im Internet adäquat zu erfassen. Grundlegend werden dafür

größere Datenmengen benötigt, die in kürzeren Abständen erhoben werden müssen. Das Web Scraping ermöglicht es, diesen Entwicklungen gerecht zu werden, indem sich damit Preisinformationen gezielt und in ihrer Frequenz nahezu beliebig oft erheben lassen. Langfristig ergibt sich dadurch eine effiziente Lösung im Vergleich zur manuellen Preiserhebung im Internet. Somit ist das Web Scraping eine Methode, die die Qualität der Verbraucherpreisstatistik langfristig gewährleisten kann.¹³ Ein weiterer Vorteil des Web Scrapings ist, dass diese Interneterhebungen mit überschaubarem Aufwand häufiger und in kürzeren Abständen zueinander durchgeführt werden können. Ebenfalls lässt sich die Stichprobe dabei flexibel anpassen.¹⁴

Darüber hinaus ist es denkbar, öffentlich zugängliche Daten von Auskunftspflichtigen nicht mehr im Rahmen der Berichtspflicht für die amtliche Statistik zu erheben, sondern mittels Web Scraping, wodurch sich eine Entlastung der Auskunftspflichtigen ergeben würde. Weitere Anwendungsmöglichkeiten des Web Scrapings lassen sich beispielsweise in der allgemeinen Berichtspflege, der Tourismusstatistik oder Unternehmensstatistik finden.

Ausblick

Die sich wandelnden gesellschaftlichen Bedingungen hinsichtlich der Digitalisierung bieten der amtlichen Statistik insgesamt Chancen, ihr bisheriges Datenangebot zu ergänzen bzw. zu erweitern. Neue digitale Daten besitzen darüber hinaus das Potenzial, die Aktualität und Genauigkeit amtlicher Statistiken zu verbessern und deren Qualität langfristig zu sichern. Die im Beitrag vorgestellten neuen digitalen Datenquellen und Methoden verdeutlichen dieses Potenzial neuer digitaler Daten. Die Implementierung neuer digitaler Daten in die amtliche Statistik geht dennoch mit großen Herausforderungen einher. So muss bei der Nutzung neuer digitaler Daten für die amtliche Statistik die gesetzliche Vorgabe, qualitativ hochwertige Daten unter den Grundsätzen der Neutralität, Objektivität und fachlichen Unabhängigkeit bereitzustellen, gewahrt werden. Auch andere rechtliche Aspekte sind zu beachten. Hier stehen beispielsweise der Datenschutz oder der geregelte Zugang zu neuen Datenquellen im Fokus. Darüber hinaus muss auch die Qualifikation der Mitarbeiterinnen und Mitarbeiter angepasst werden und eine geeigneten IT-Infrastruktur auf- bzw. ausgebaut werden. ■

11 Kühnemann, Heidi (2021): Anwendungen des Web Scraping in der amtlichen Statistik, in: Asta Wirtschafts- und Sozialstaatliches Archiv. (15), S. 5–25.

12 Siehe Statistik Hessen <https://statistik.hessen.de/unsere-zahlen/experimentelle-statistik/web-scraping-neue-digitale-daten-aus-dem-internet> (Abruf: 12.09.2022).

13 Bladow, Christian/ Seeger, Daniel (2019): Fortschritte beim Einsatz von Web Scraping in der amtlichen Verbraucherpreisstatistik – Ein Werkstattbericht, in: WISTA Wirtschaft und Statistik. (4), S.19–30.

14 Bladow, Christian/ Ostermann, Holger (2020): Entwicklung eines generischen Programms für die Nutzung von Web Scraping in der Verbraucherpreisstatistik, in: WISTA Wirtschaft und Statistik. (5), S.103–113.

Weitere Auskünfte erteilt Kevin Heinen, Telefon 0711/641-27 32, Kevin.Heinen@stala.bwl.de