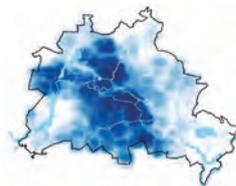


Räumliche Daten können aufschlussreiche Informationen über ein Merkmal liefern. Dabei dürfen keine Aufdeckungsrisiken erzeugt werden, die die statistische Geheimhaltung verletzen.

Gitterlänge (von oben): $d=500\text{ m}$, $d=1000\text{ m}$, $d=2000\text{ m}$, $d=2500\text{ m}$

Die Relevanz des Informationsverlusts durch Anonymisierung wird am Beispiel des Berliner Einwohnerregisters beleuchtet (S. 48).



Amt für Statistik Berlin-Brandenburg

Zeitschrift für amtliche Statistik
Berlin-Brandenburg
8. Jahrgang

Herausgeber
Amt für Statistik Berlin-Brandenburg
Behlertstraße 3a
14467 Potsdam
Tel.: 0331 8173-1777

Verantwortlicher Redakteur i. S. d. BbgPG
Hartmut Bömermann
Redaktion
Nicole Dombrowski, Steffi Kuß,
Dr. Holger Leerhoff,
Anja Malchin, Dr. Thomas Troegel,
Ramona Voshage (Leitung)

Preis
Einzelheft EUR 6,00
ISSN 1864-5356

Satz und Gestaltung
Amt für Statistik Berlin-Brandenburg

Druck
TASTOMAT GmbH,
15345 Petershagen/Eggersdorf

© Amt für Statistik Berlin-Brandenburg, 2014
Auszugsweise Vervielfältigung und
Verbreitung mit Quellenangabe gestattet.

Das Amt für Statistik Berlin-Brandenburg
hat seinen Sitz in Potsdam und weitere
Standorte in Berlin und Cottbus.

Auskunft und Beratung

Behlertstraße 3a
14467 Potsdam

Telefon: 0331 8173-1777
Fax: 030 9028-4091
info@statistik-bbb.de

Zeichenerklärung

- 0 weniger als die Hälfte von 1
in der letzten besetzten Stelle,
jedoch mehr als nichts
 - nichts vorhanden
 - ... Angabe fällt später an
 - () Aussagewert ist eingeschränkt
 - / Zahlenwert nicht sicher genug
 - Zahlenwert unbekannt oder
geheim zu halten
 - x Tabellenfach gesperrt, weil
Aussage nicht sinnvoll
 - p vorläufige Zahl
 - r berichtigte Zahl
 - s geschätzte Zahl
- Abweichungen in der Summe
können sich durch Schätzungen
ergeben

Liebe Leserinnen und Leser,

diese Ausgabe der Zeitschrift für amtliche Statistik Berlin Brandenburg setzt das im Heft 2/2014 begonnene Thema „Georeferenzierung“ mit sechs weiteren Beiträgen fort.

Die Anforderungen der nationalen, zwischenstaatlichen und internationalen Ebenen an eine Raumgliederung untersucht Dr. Susanne Schnorr-Bäcker vom Statistischen Bundesamt. Eine teilträumliche und dabei über nationale Grenzen hinweg vergleichbare Betrachtung gehört sowohl für die Europäische Kommission wie für die OECD zu den unverzichtbaren analytischen Zugängen zur räumlichen Struktur und zum Wandel von Ländern und Regionen. Die dafür entwickelten

Raumtypisierungen und Indikatoren-systeme werden in dem Beitrag zur **internationalen Perspektive** des Raumbezugs ebenfalls vorgestellt.

Tobias Kirchner, Franziska Pflanz, Dr. Andreas Techen und Lars Wagenknecht vom Amt für Statistik Berlin-Brandenburg berichten über die Erfahrung mit kleinräumigen Gliederungen, Georeferenzierung und der Rasterdarstellung im Zensus. Sehr ausführlich wird die Georefe-



Kurzberichte

- ▮ Statistische Woche 2014 in Hannover **3**
- ▮ Ausländische Delegationen informieren sich im AFS **5**

Entwicklungen in der Amtlichen Statistik

- ▮ Der Online-Atlas Agrarstatistik **6**
- ▮ Meldungen zur amtlichen Statistik jetzt online **10**

Historisches

- ▮ 100 Jahre Erster Weltkrieg – Berliner Statistik im Jahr 1914 **62**

Statistik erklärt

- ▮ „Geheimnisvolle“ Begriffe in statistischen Tabellen – Was bedeuten davon, darunter, und zwar sowie nachrichtlich? **65**

Neuerscheinungen

- ▮ Zweiter Kernindikatorenbericht zur nachhaltigen Entwicklung Berlins **66**

Fachbeiträge

Internationale Perspektive

- ▮ **Regionalstatistische Daten für Deutschland, Europa und weltweit** **12**
Susanne Schnorr-Bäcker

Zensus 2011

- ▮ **Kleinräumige Gliederung, Georeferenzierung und Rasterdarstellung im Zensus** **28**
Tobias Kirchner, Franziska Pflanz, Andreas Techen, Lars Wagenknecht

Fachgespräch mit Dr. Andreas Techen

- ▮ **Die rasterbasierten Auswertungen stehen erst am Anfang** **33**

Fachstatistische Anwendungen

- ▮ **Die kleinräumige Gliederung Grundlage eines Raumb Beobachtungssystems für ein strategisches Controlling in der Landeshauptstadt Potsdam** **34**
Lutz Rittershaus

- ▮ **Datenschutzkonzept zur Nutzung von SOEPgeo im Forschungsdatenzentrum SOEP am DIW Berlin** **42**
Jan Goebel, Bernd Pauer

Geheimhaltung

- ▮ **Informationsverlust durch Anonymisierung am Beispiel der Berliner Einwohnerregisterdaten** **48**
Anpassungsgüte nichtparametrischer Dichteschätzer bei diskretisierten Daten
Sebastian Schmon

- ▮ **Statistische Geheimhaltung bei der Auswertung georeferenzierter Daten** **54**
Jörg Höhne, Julia Höninger

renzung der Adressangaben im **Zensus 2011** behandelt, die erstmals in der amtlichen Statistik in einem so großen Umfang bewältigt werden musste. Als ein vergleichsweise neues Thema wird die Rasterdarstellung von Ergebnissen unter Wahrung der statistischen Geheimhaltung erörtert.

Über große und langjährige Erfahrung mit lokalen kleinräumigen Systemen verfügt die Kommunalstatistik. Lutz Rittershaus aus dem Bereich Statistik und Wahlen der Brandenburgischen Landeshauptstadt Potsdam stellt die kleinräumige Gliederung als Grundlage eines Raubeobachtungssystems für ein strategisches Controlling in Potsdam vor. Ergänzt wird der Beitrag durch ausgewählte **fachstatistische Anwendungen**.

Dr. Jan Goebel und Bernd Pauer vom DIW Berlin berichten über eine Erweiterung des Analysepotenzials des Sozio-oekonomischen Panels (SOEP) durch die Nutzung der Geokoordinaten der befragten Stichprobenhaushalte. Der Forschung eröffnet sich dadurch die Möglichkeit, beliebige geografische Abgrenzungen der Untersuchungsgebiete vorzunehmen oder die Befragungsdaten um Kontextmerkmale anzureichern. SOEPgeo ermöglicht es Wissenschaftlern, geokodierte Daten innerhalb des DIW zu nutzen,

ohne die hoch-sensitiven Geokoordinaten der Haushalte zugänglich zu machen. Die erforderlichen technisch-organisatorischen Vorkehrungen werden detailliert erläutert.

Mit dem Thema Informationsverlust durch Anonymisierung beschäftigt sich am Beispiel der Berliner Einwohnerregisterdaten Sebastian Schmon, Absolvent der Freien Universität Berlin. Verwendet wird ein räumlicher Dichteschätzer für punktbezogene Daten und unterschiedlich gewählte Granularität, um den Effekt von Anonymisierungsschritten abschätzen zu können. Das Thema ist für die raumbezogene Statistik von großer Bedeutung, da die Herausarbeitung räumlicher Strukturen keine Aufdeckungsrisiken erzeugen darf, die die Verpflichtung zur statistischen Geheimhaltung verletzen würden.

Der letzte Heftbeitrag behandelt das bereits angesprochene zentrale Thema der amtlichen Statistik, die Datenanonymität. Dr. Jörg Höhne und Julia Höninger vom Amt für Statistik Berlin-Brandenburg erörtern die statistische **Geheimhaltung** bei der Auswertung von georeferenzierten Daten. Diskutiert wird die Übertragung von Verfahren aus der Tabellengeheimhaltung auf die kartografische Ergebnisdarstellung. Die Besonderheiten von Verhältniszahlen in Karten und das Problem der geografischen Differenzbildung, das bei zusätzlichen Rasterkarten in noch stärkerem Maße auftritt, werden umrissen. Die Autoren betonen, dass das Thema noch nicht erschöpfend untersucht und eine weitere Forschung notwendig ist. Eingedenk der wachsenden Verbreitung von Geodaten und deren Nutzung ist dies kein Nischenthema, sondern sollte in einer breiteren Öffentlichkeit diskutiert werden.

Eine anregende Lektüre wünscht Ihnen



Hartmut Bömermann
verantwortlicher Redakteur

Kurzbericht

Statistische Woche 2014 in Hannover



Zensus 2011: Results and Outlook
(Session 3): Dr. Jörg Höhne
„Die statistische Geheimhaltung beim Zensus“

Die jährliche gemeinsame Tagung der Deutschen Statistischen Gesellschaft (DStatG), des Verbandes Deutscher Städtestatistiker (VDSt) und der Deutschen Gesellschaft für Demographie (DGD), die *Statistische Woche*, fand in diesem Jahr vom 16. bis zum 19. September auf dem Conti-Campus der Leibniz-Universität Hannover statt. Die Statistische Woche stellt regelmäßig eine hervorragende Gelegenheit zum Erfahrungsaustausch mit Kolleginnen und Kollegen aus der (nicht nur amtlichen) Statistik, weiteren Datenanbietern sowie Datennutzerinnen und -nutzern verschiedener Institute und Verbände, aus Politik und Wissenschaft dar. Zudem ermöglicht die Veranstaltung, bei den mathematisch-statistischen Methoden auf dem Laufenden zu bleiben, politische und soziale Entwicklungen zu diskutieren und im eher informellen Rahmen Kontakte für künftige Kooperationen zu knüpfen. Mit ihrer inzwischen 86-jährigen Geschichte erfüllt die Tagung zweifellos eine Schlüsselfunktion für den Zusammenhalt und die Kooperation der statistischen Gemeinschaft im deutschsprachigen Raum.

Wie auch in den letzten Jahren war das Amt für Statistik Berlin-Brandenburg (AfS) aktiv an der Veranstaltung beteiligt. So leitete Ramona Voshage (Leiterin des Referats *Mikrodaten, Analysen, Forschungsdatenzentrum* im AfS) eine Session zum Thema „Amtliche Firmendaten für Deutschland“ (AFiD), in der wissenschaftliche Arbeiten präsentiert wurden, die mit den AFiD-Mikrodaten des Forschungsdatenzentrums der Statistischen Ämter des Bundes und der Länder durchgeführt wurden.

Eine weitere Session wurde moderiert durch Prof. Dr. Ralf Münnich und Hartmut Bömermann (Leiter der Abteilung *Bevölkerung und Regionalstatistik* im AfS). Hier wurden Schätzverfahren und Imputationsmethoden zur Verbesserung regionalisierter Auswertungen vorgestellt und mit dem Plenum diskutiert.

Zu den diesjährigen programmatischen Schwerpunkten zählten der Zensus 2011 und Big Data. Die Vorträge machten deutlich, dass die Zensusdaten

ein enormes Analysepotenzial aufweisen – dies betrifft insbesondere regionalisierte Auswertungen – und von der Wissenschaft bereits in großem Umfang genutzt werden. In weiteren Vorträgen wurden Aspekte der beim Zensus 2011 angewendeten Methodik detailliert beschrieben und die Vergleichbarkeit mit anderen Daten der amtlichen Statistik teils kritisch hinterfragt. Auch hier brachte das AfS seine Expertise ein. So referierte Dr. Jörg Höhne (Leiter der Abteilung *Gesamtwirtschaft* im AfS) zu den von der amtlichen Statistik eingesetzten Methoden zur statistischen Geheimhaltung und den Erfahrungen des Einsatzes

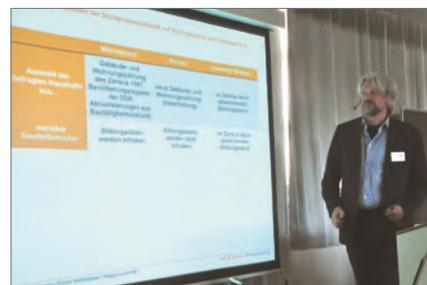
beim Zensus 2011. Auf Grundlage der Zensus- und Mikrozensusergebnisse für Berlin und Brandenburg legte Dr. Holger Leerhoff gemeinsam mit Prof. Dr. Klaus Rehkämper (Leiter des Referates *Schule Berlin, Bildungsanalysen* im AfS) und Prof. Dr. Ulrike Rockmann die unterschiedlichen Resultate zum Bildungsstand in Berlin und Brandenburg und die Konsequenzen für die Bildungsberichterstattung dar.

Das Thema Big Data ist seit einigen Jahren beinahe omnipräsent auf Fachtagungen. Zwei generelle Entwicklungsstränge können in diesem Bereich unterschieden werden: Auf der einen Seite werden die statistischen Analyseverfahren durch die Wissenschaft immer weiter entwickelt, sodass eine sinnvolle Analyse massiver, heterogen strukturierter Datenmengen möglich ist. Auf der anderen Seite werden zunehmend Szenarien thematisiert, in denen solche Daten überhaupt anfallen und bei denen entsprechende Analysen zu aussagekräftigen Ergebnissen führen könnten. Ob und inwieweit die amtliche Statistik auf diesem Gebiet künftig tätig werden kann oder gar muss, ist bisher kaum abzusehen und Gegenstand der Diskussion.

Auf der ebenfalls im Rahmen der Statistischen Woche stattfindenden VDSt-Mitgliederversammlung wurde Hartmut Bömermann zum stellvertretenden Vorsitzenden des Verbandes deutscher Städtestatistiker gewählt. Weiterhin gehört er der Schriftleitung der Zeitschrift „Stadtforchung und Statistik“ an, die vom VDSt herausgegeben wird. Die Zeitschrift erscheint zweimal jährlich und dient der fachlichen Diskussion und Kommunikation in der Städtestatistik.

Weitere Informationen finden Sie unter:

www.statistische-woche.de



Zensus 2011: Results and Outlook
(Session 4): Dr. Holger Leerhoff
„Bildungsstand in Berlin und Brandenburg aus Zensus und Mikrozensus“

Kurzbericht

┌ Ausländische Delegationen informieren sich im AfS

Ein fachlicher Austausch, der über Landesgrenzen hinausgeht, ist für das Amt für Statistik Berlin-Brandenburg (AfS) immer wieder eine besondere Gelegenheit, um über die eigene Arbeit zu informieren, und stellt für die beteiligten Mitarbeiterinnen und Mitarbeiter eine wertvolle Bereicherung dar. In den Monaten August bis Oktober 2014 durfte das AfS vier ausländische Delegationen begrüßen.

Volksrepublik China

Am 28. August besuchte eine 25-köpfige Delegation des chinesischen Nationalen Statistischen Amtes das AfS am Standort Berlin. Die Delegation befand sich auf einer Studienreise durch Deutschland zum Thema „Qualitätsanpassung in Preisstatistiken“ und hatte den Wunsch nach einem Besuchstermin zum Thema Verbraucherpreisstatistik. Vorbereitet und koordiniert wurde die Reise von der deutschen Firma RKWC GmbH. Den Teilnehmerinnen und Teilnehmern war es sehr wichtig, neben Informationen vom Statistischen Bundesamt auch Informationen über die konkrete Durchführung der Verbraucherpreisstatistik und die Ermittlung des Verbraucherpreisindex in einem statistischen Landesamt zu erhalten. Elke Zimmer, Leiterin des Referats *Preise, Verdienste, Arbeitskosten* im AfS, erläuterte den chinesischen Gästen die Verbraucherpreisstatistik in Berlin und Brandenburg.



Elke Zimmer, Leiterin des Referats *Preise, Verdienste, Arbeitskosten* im AfS, erläutert den chinesischen Gästen die Verbraucherpreisstatistik in Berlin und Brandenburg.

Verdienste, Arbeitskosten im AfS, informierte die chinesischen Gäste u. a. über die Bildung der Stichprobe, die Umsetzung des Erhebungskatalogs („Warenkorb“), die Erstellung der Erhebungsunterlagen, den Ablauf der Erhebung vor Ort sowie den Einsatz von Preiserhebenden und deren Schulung. Auch der Informationsaustausch zwischen den statistischen Ämtern sowie die Schulung der Mitarbeiterinnen und Mitarbeiter in den statistischen Landesämtern wurden näher hinterfragt. Abschließend äußerte sich die Delegation sehr zufrieden über die dargelegten Informationen und sprach dem Haus und allen Beteiligten ihren Dank aus. Nach der Rückkehr in die Heimat und einer Sichtung sowie Auswertung aller Informationen wurde ein weiterer Besuch in Aussicht gestellt.

Königreich Marokko

Im Vorfeld der Landtagswahl 2014 in Brandenburg besuchten am 11. September fünf Mitglieder der sozialdemokratischen marokkanischen Partei USFP (Union socialiste des forces populaires) das AfS, um sich über die Organisation von Wahlen auf Landesebene zu informieren. Vorbereitet und koordiniert wurde der Besuch durch die Friedrich-Ebert-Stiftung. Im AfS empfing Geert Baasen, Leiter der Geschäftsstelle der Berliner Landeswahlleiterin, die jungen Parlamentarierinnen und Parlamentarier, die sich besonders für die Einteilung der Wahlkreise und die Rolle des AfS bei der Ergebnisermittlung und -veröffentlichung interessierten.

Junge Parlamentarierinnen und Parlamentarier aus Marokko informieren sich im AfS über die Organisation von Landtagswahlen.



Nigeria

Um das Thema Wahlen ging es auch bei den Besuchen zweier nigerianischer Delegationen im Oktober, die wiederum durch Geert Baasen betreut wurden. Am 7. Oktober hatten zwei Mitglieder der nationalen Wahlkommission Nigerias und zwei Mitglieder der zivilgesellschaftlichen Organisation Centre for Social Justice die Gelegenheit zu einem Gespräch über die gesetzlichen Rahmenbedingungen von Wahlen in Deutschland mit besonderem Schwerpunkt auf gesetzlichen Regelungen und finanzierungsrelevanten Aspekten. Die Reise wurde ebenfalls von der Friedrich-Ebert-Stiftung organisiert und stand unter der Überschrift „Gesetzeslage und Monitoring – Wahlkampffinanzierung als wesentlicher Aspekt eines fairen Wahlprozesses“.

Die Informationsreise der zweiten nigerianischen Delegation organisierte das Goethe-Institut Besucherprogramm. Die Gäste wurden im Rahmen des Besucherprogramms der Bundesrepublik Deutschland vom Auswärtigen Amt nach Deutschland eingeladen. Vier nigerianische Journalisten tauschten sich während ihres siebentägigen Aufenthalts mit hiesigen Experten über Organisation und Durchführung sowie Kommunikation und Auswertung von Wahlen aus. Auf dem Programm standen neben dem Besuch der Geschäftsstelle der Berliner Landeswahlleiterin am 14. Oktober u. a. Termine beim Bundesministerium des Innern, bei infratest dimap sowie Gespräche mit Pressevertretern und ein Abendessen mit Berliner Wahlhelferinnen und -helfern.

Geert Baasen,
Leiter der Geschäftsstelle der Berliner
Landeswahlleiterin,
mit Journalistinnen
und Journalisten aus Nigeria.



Entwicklungen in der amtlichen Statistik

Der Online-Atlas Agrarstatistik

von Thomas Troegel

Zur Entstehung des Online-Atlases Agrarstatistik

Im November 2010 verabschiedete die Amtsleiterkonferenz in ihrer Sitzung ein Arbeitsprogramm für die amtliche Statistik, welches die Umsetzung der INSPIRE-Richtlinie (*Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft – Abl. L 108 vom 14. März 2007, S. 1*) gewährleisten soll. Dieses Arbeitsprogramm gliederte sich in unterschiedliche Arbeitspakete. Im Rahmen des Arbeitspaketes 3 *Nicht administrative Aggregate* wurde der Bereich *Agrarstatistik* beauftragt, sich mit Konzepten im Bereich der Datenanalyse, Geheimhaltung und Veröffentlichung als Grundlage für die Erarbeitung statistikübergreifender Konzepte zum Umgang mit georeferenzierten Daten zu befassen.

Aus diesem Grund wählte das Statistische Bundesamt in Zusammenarbeit mit den Statistischen Ämtern der Länder verschiedene Karten aus und diskutierte diese. Dabei standen neben den fachlichen Inhalten auch Fragen von praktikablen und möglichst innovativen Darstellungsformen im Fokus. Die intensiven Diskussionen führten zum Resultat, dass anhand ausgewählter Merkmale aus der Landwirtschaftszählung 2010 Kartenprototypen zu erstellen sind. In Verbindung mit der Weiterentwicklung des Veröffentlichungskonzepts für georeferenzierte Daten wurde beabsichtigt, die vorgesehenen Prototypen als Entscheidungshilfe für die Festlegung der Merkmale und der räumlichen Einheiten (Rasterzellen, Naturräume u. Ä.) zu verwenden. Diese Prototypen stellten dann die Basis dar, welche Karten zukünftig veröffentlicht werden.

Auch im Ergebnis dieser Vorleistungen durch das Statistische Bundesamt und die Referentenbesprechung *Agrarstatistiken* verabschiedete das Abteilungsleitergremium *Fachstatistiken* im Oktober 2012 unter anderem zwei Beschlüsse zur Veröffentlichung von georeferenzierten Ergebnissen, welche den Bereich der Agrarstatistik unmittelbar betrafen.

Einerseits sollten demnach Veröffentlichungen von georeferenzierten Ergebnissen „grundsätzlich elektronisch in kartografischer Form als Gemeinschaftsveröffentlichung“ erfolgen. Andererseits wurde die Zielstellung ausgegeben, im Jahr 2013 eine bundesweite kartografische Gemeinschaftsveröffentlichung für die georeferenzierten Daten aus der Landwirtschaftszählung 2010 zu realisieren. Diese bundesweite kartografische Gemeinschafts-

veröffentlichung stellt dabei eine Online-Veröffentlichung des Kartenprogramms (*ausschließlich Karten*) dar und war somit kein Konkurrenzprodukt zur erfolgreichen Gemeinschaftsveröffentlichung *Agrarstrukturen in Deutschland – Einheit in Vielfalt – Regionale Ergebnisse der Landwirtschaftszählung 2010*, die 2011 editiert wurde.

Inhaltliche Entscheidungen

Umfangreiche Diskussionen führten im Ergebnis von Referentenbesprechungen dazu, dass der größte Teil der vorgesehenen Karten auf einheitlichen Rasterweiten beruhen sollte. Die Untersuchungen des Statistischen Bundesamtes zur optimalen Rasterweite für Rasterkarten ergaben, dass es keine für alle Merkmale allgemein gültige Rasterweite gibt und dass Rasterweiten von mehr als 10 km nicht sinnvoll sind. Größere Raster würden in einem Teil der Fälle keinen inhaltlichen Erkenntnisgewinn gegenüber einer Kreiskarte bringen. Gleichzeitig war aber auch festzustellen, dass selbst eine Rasterweite von 10 km nicht immer die Zahl an Geheimhaltungsfällen auf ein vertretbares Niveau reduziert.

Des Weiteren wurde zum Beispiel geprüft,

- ob alternativ eine Darstellung mit variierender Rasterweite in der jeweiligen Karte erfolgen sollte oder
- ob auf die kartografische Darstellung des entsprechenden Merkmals komplett verzichtet werden müsste.

Weiterhin konnte herausgearbeitet werden, dass in dem gemeinsamen Kartenprogramm generell je Merkmal nur eine Rasterkarte vorzusehen ist, um die Aufdeckung von Einzelfällen durch die Überlagerung von Karten mit unterschiedlichen Rasterweiten zu vermeiden.

Das Ziel der Online-Karten liegt darin, den Nutzerinnen und Nutzern statistische Daten raumbezogen zu veranschaulichen. Mit Hilfe der Karten können sich die interessierten Anwenderinnen und Anwender recht unkompliziert einen Überblick über die kleinräumige Verteilung der Daten verschaffen. Eine solche Betrachtungsweise ist über die direkte Interpretation von Veröffentlichungstabellen nicht so einfach möglich.

Um den Nutzerinnen und Nutzern soweit wie möglich entgegenzukommen, beschränkt sich die Online-Veröffentlichung des Kartenprogramms nicht auf einen reinen Darstellungsdienst, sondern gestattet auch das Herunterladen von Karten.

Bei der Betrachtung und Bewertung der in den 16 Karten dargestellten Daten ist nachfolgender Fakt unbedingt zu beachten: Alle Angaben werden am Betriebsitz der landwirtschaftlichen Betriebe nachgewiesen. Dies ist in der Regel das Grundstück, auf dem sich die wichtigsten Wirtschaftsgebäude des Betriebes befinden. Dadurch kann die lagerichtige Darstellung der Merkmale nur eingeschränkt erfolgen. Dies gilt ganz besonders für Flächen- und Tiermerkmale.

Außerdem können aus den Karten keine Einzeldaten für die einzelnen Raster oder auch Flussgebietseinheiten entnommen und damit Geheimhaltungsfälle aufgedeckt werden. Um derartige Darstellungen möglich zu machen, wurden in vielen Fällen Größenklassen gebildet, in denen sich zum Beispiel die Betriebe in einem Raster einordnen lassen.

Für ein Bundesland wie Brandenburg stellt die kleinräumige Darstellung ein gewisses Problem dar, da bisher im Bereich der Agrarstatistiken noch nie Ergebnisse unterhalb der Kreisgrenzen veröffentlicht wurden. Dass dieser Atlas Begehrlichkeiten wecken könnte, lässt sich nicht generell ausschließen.

Zum Kartenaufbau des Online-Atlases Agrarstatistik

Der größte Teil der Karten basiert auf Rasterzellen mit 5 km Gitterweite. Mit diesem Vorgehen werden zum ersten Mal Karten mit einer räumlichen Auflösung angeboten, die keine Verwaltungsgrenzen abbilden und in einem nicht unerheblichen Teil sogar unterhalb der Gemeindeebene liegen.

Die Kartenanwendung selbst gliedert sich in drei Bereiche:

- Selektion der Karteninhalte,
- Karte mit Navigationsfunktionalität und Menüleiste,
- Legende.

Im Auswahlbereich stehen wiederum zwei Blöcke zur Verfügung. Die Anwenderinnen und Anwender haben die Möglichkeit der Selektion eines Themas (Auswahlfelder: Karteninhalte und Erhebungsjahr). Aufgrund dieser eindeutigen Themenwahl erfolgt das Laden der jeweiligen Karte.

Standardmäßig erfolgt die Anzeige der Grenzen der Bundesländer sowie der Bundeshauptstadt und der Landeshauptstädte.

Angenehme Boni sind die Möglichkeiten des Zuschaltens von Kreis- und Gemeindegrenzen, des Überlagerns von Naturräumen (z. B. Mecklenburger Seenplatte, Odertal, Westerwald) sowie des Einfügens großer Flüsse u. Ä.

Für die Navigationsfunktionen steht im Kartenbereich ein leicht zu bedienender Satz an Werkzeugen zur Verfügung:

- Verschieben der Karte,
- Gesamtansicht anzeigen,
- Vergrößern bzw. Verkleinern.

In Ergänzung wurden in diesem Atlas umfassende Metadaten als Hintergrundinformationen zu den Karten hinterlegt. Dabei wurde aber auf die Begrifflichkeit „Metadaten“ verzichtet und diese als „Erläuterungen zu den Statistiken“ bezeichnet.

Karteninhalte

Den interessierten Nutzerinnen und Nutzern stehen folgende 16 Karten zur Verfügung

- **Durchschnittliche Größe** landwirtschaftlicher Betriebe 2010,
- **Landwirtschaftliche Betriebe** 2010,
- **Anzahl Schweine** je 100 Hektar landwirtschaftlich genutzter Fläche 2010,
- **Anzahl Schweine** je 100 Hektar landwirtschaftlich genutzter Fläche 2010 in Flussgebietseinheiten,
- **Anzahl Rinder** je 100 Hektar landwirtschaftlich genutzter Fläche 2010,
- **Anzahl Rinder** je 100 Hektar landwirtschaftlich genutzter Fläche 2010 in Flussgebietseinheiten,
- **Großvieheinheiten** je 100 Hektar landwirtschaftlich genutzter Fläche 2010,
- **Großvieheinheiten** je 100 Hektar landwirtschaftlich genutzter Fläche 2010 in Flussgebietseinheiten,
- **Anteil der Maisfläche** an der landwirtschaftlich genutzten Fläche 2010,
- **Anteil der Dauergrünlandfläche** an der landwirtschaftlich genutzten Fläche 2010,
- **Anteil der Weizenfläche** an der landwirtschaftlich genutzten Fläche 2010,
- **Anteil der Rapsfläche** an der landwirtschaftlich genutzten Fläche 2010,
- **Anteil der ökologisch bewirtschafteten Fläche** an der landwirtschaftlich genutzten Fläche 2010,
- **Arbeitskräfte-Einheiten** je 100 Hektar landwirtschaftlich genutzter Fläche 2010,
- **Saisonarbeitskräfte-Einheiten** je 100 Hektar landwirtschaftlich genutzter Fläche 2010,
- **Vorherrschende betriebswirtschaftliche Ausrichtung** landwirtschaftlicher Betriebe 2010.

Mit Hilfe dieser Karten lassen sich verschiedenste Fragestellungen beantworten.

Zu diesen gehören beispielsweise:

- In welchen Regionen hat der ökologische Landbau größere Bedeutung?
- Wo gibt es besonders viele Schweine oder Rinder?
- Wo ist der Einsatz von Saisonarbeitskräften besonders hoch?
- Wo ist der intensivste Rapsanbau zu beobachten?

Der Online-Atlas zur Agrarstatistik ermöglicht faszinierende Einblicke zu ganz unterschiedlichen Themen, auch wenn er natürlich nicht jedes Teilgebiet der Landwirtschaft abdecken kann.

So erhält der Betrachter der Karten detaillierte bildliche Informationen zur Schweine- und Rinderdichte. Auffallend ist zum Beispiel, dass im Nordwesten Deutschlands eine bemerkenswert hohe Schweinedichte vorhanden ist. Tatsächlich sind Niedersachsen und Nordrhein-Westfalen die beiden Bundesländer mit der jeweils höchsten Anzahl an gehaltenen Schweinen. Im Osten der Republik sind hier viel geringere Besatzzahlen zu erkennen. Auch bei den Rindern ist ein deutliches „Gedränge“ im Nordwesten zu beobachten.

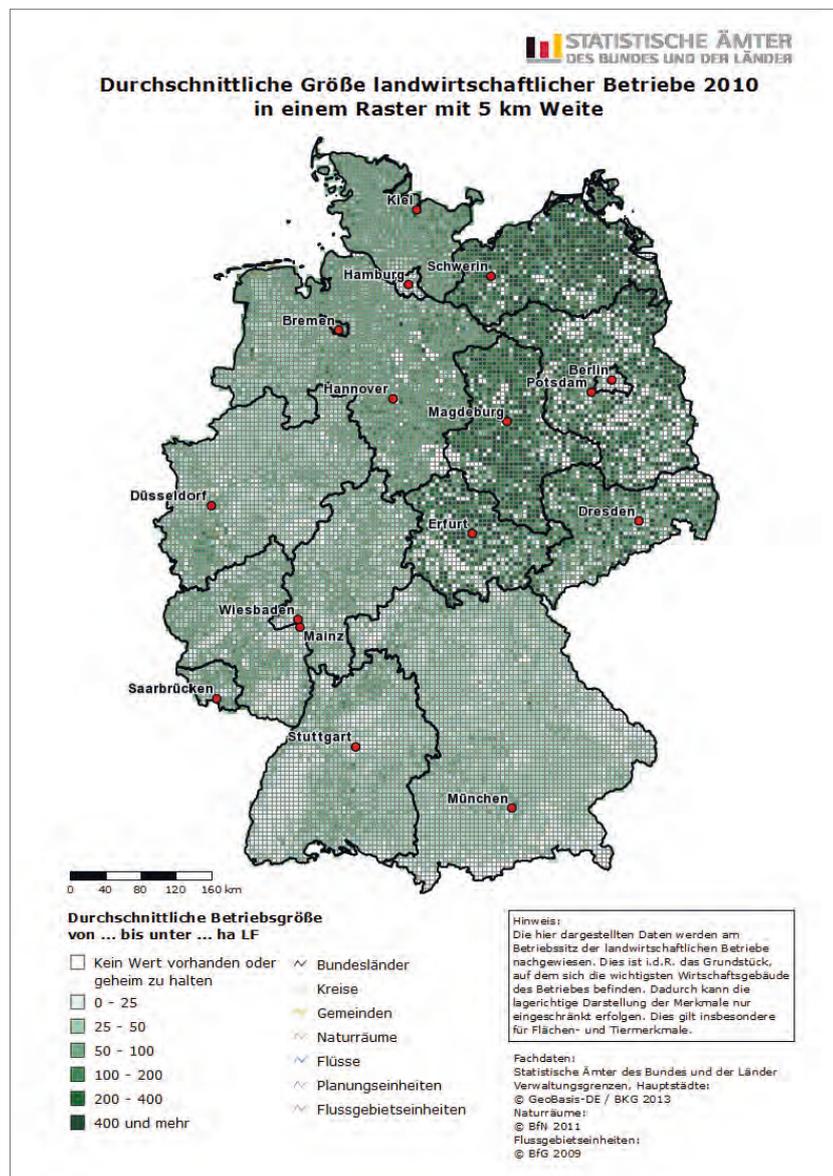
Dagegen lassen sich auch aus Geheimhaltungsgründen keine nutzerfreundlichen kartografischen Darstellungen zu Schafen oder Gänsen, Enten und anderem Geflügel produzieren.

Mit Hilfe der Karten wird nicht nur die geografische Verteilung von Schwerpunkten der deutschen Landwirtschaft aufgezeigt. Die Rasterkarten ermöglichen auch Einblicke in meist schwer fassbare Zusammenhänge.

Mehr als 20 Jahre nach der Wiedervereinigung lassen sich durchaus historische „Grenzen“ erkennen. Dies ist sehr gut beim Vergleich der Betriebsgrößen zu sehen. In gewisser Weise sind hier die unterschiedlichen Ansätze der jeweiligen Agrarpolitiken immer noch in einem bestimmten Umfang zu erkennen (Abbildung a).

Spannend ist auch die Übersicht zum Einsatz von Saisonarbeiterinnen und Saisonarbeitern. Hier wird die Aufmerksamkeit des Betrachters ganz beson-

a | Online-Atlas Agrarstatistik, ausgewählte Karte



Quelle: <http://www.atlas-agrarstatistik.nrw.de/>

ders auf die relevanten deutschen Anbaugelände von Spargel, Obst und Wein gelenkt. Für Brandenburg zeigt sich eine kleine Konzentration auf das berühmte Spargelanbaugebiet in der Region Beelitz. Doch auch der Baumobstanbau in Frankfurt (Oder) lässt sich relativ leicht erschließen.

Die schon angesprochene hohe Tierkonzentration im Nordwesten Deutschlands korrespondiert sehr deutlich mit einem intensiven Maisanbau. Gleichzeitig wird deutlich, dass in vielen Teilen der Bundesrepublik nicht unbedingt von einer „Vermaisung“ gesprochen werden sollte.

Anhand der Karte zum deutschen Weizenanbau sind prinzipiell Rückschlüsse auf die Bodengüte zu erkennen. Ein Bundesland wie Brandenburg, welches durch eher weniger gute Bodenqualitäten gekennzeichnet ist, verfügt über eine geringere Weizenproduktion als viele andere Bundesländer.

Bei der Betrachtung des Ökolandbaus ist die starke Konzentration auf den nordöstlichen Teil Deutschlands und auf größere Teile der Mittelgebirgsregionen wie den Taunus und den Westerwald, aber auch auf den Voralpenraum, zu erkennen. Die regional erheblichen Unterschiede beruhen unter anderem auf abweichenden Förderbedingungen sowie auf klimatischen Besonderheiten, Bodengüte und Topografie, aber auch auf fehlenden agrarischen Alternativen.

Die Karte zur betriebswirtschaftlichen Ausrichtung der landwirtschaftlichen Betriebe gibt Auskunft, in welchen Regionen zum Beispiel der Futterbau dominiert und wo sich der Anbau von Dauerkulturen konzentriert. Bestimmte Weinanbaugebiete lassen sich für Interessierte schnell ausfindig machen.

Fazit

Der Anfang 2014 freigeschaltete Online-Atlas Agrarstatistik stellt eine bemerkenswerte Erweiterung des Datenangebots der Statistischen Ämter des Bundes und der Länder um kleinräumige digitale Karten dar. Dieser Atlas enthält insgesamt 16 Karten zum Thema Landwirtschaft. Die statistische Datengrundlage bildete die Landwirtschaftszählung 2010, in welcher erstmals in der amtlichen Statistik Daten georeferenziert erhoben wurden.

Der Online-Atlas steht kostenfrei im gemeinsamen Statistik-Portal der Statistischen Ämter des Bundes und der Länder zur Verfügung (www.statistik-portal.de). Die einfach zu bedienende Web-Anwendung bietet neben der Ausgabe der Karten als Bilder und als PDF-Dateien auch einen WMS-Dienst (Web Map Service) an. Dieser Kartendienst gestattet die Nutzung der Karten in anderen Geografischen Informationssystemen.

In jedem Fall stellt der Online-Atlas Agrarstatistik ein interessantes und lohnendes Kartenwerk dar, das Nutzerinnen und Nutzern einen schnellen Überblick zu ausgewählten Themen der deutschen Landwirtschaft gibt. Eine Aktualisierung des vorliegenden Materials ist nach 2016 für ausgewählte Merkmale geplant. Basierend auf den Daten der Agrarstrukturerhebung 2016 wird dann ein neuer Kartensatz bereitgestellt, der jedoch aufgrund der Methodik der Erhebungsdurchführung nicht die Erneuerung aller bisher vorliegenden Karten zulässt.

Dr. Thomas Troegel leitet das Referat *Flächennutzung, Land- und Forstwirtschaft, Fischerei* des Amtes für Statistik Berlin-Brandenburg.

Entwicklungen in der amtlichen Statistik

└ Meldungen zur amtlichen Statistik jetzt online

von **Astrid Langer**

Seit nunmehr fast zehn Jahren besteht für Betriebe, Unternehmen und öffentliche Stellen in allen Bundesländern die Möglichkeit, ihrer Meldeverpflichtung zu amtlichen Statistiken auch elektronisch via Internet nachzukommen. Als eine Variante ist dabei das Online-Meldeverfahren IDEV (Internet Datenerhebung im Verbund) zu nennen, das die formularbasierte direkte Eingabe der Daten je Statistik vorsieht. Eine weitere Möglichkeit ist das Meldesystem eSTATISTIK.core (Common Online Rawdata Entry), bei dem die Daten über eine Schnittstelle vorzugsweise direkt aus der betriebswirtschaftlichen Software des Unternehmens oder der öffentlichen Stelle über das Internet übertragen werden. Die Übermittlung der statistischen Daten mit IDEV und eSTATISTIK.core erfolgt immer in verschlüsselter Form unter Verwendung des technischen Verfahrens HTTPS. Der elektronische Dateneingang der statistischen Ämter unterliegt den Empfehlungen für Sicherheitsanforderungen gemäß der Standards des Bundesamtes für Sicherheit in der Informationstechnik (BSI).

Im Zuge der Einführung elektronischer Verwaltungsdienste (E-Government) wurde seit längerer Zeit auch ein weiterer Schritt zur verbindlichen Nutzung von Online-Verfahren in der amtlichen Statistik vorbereitet. Am 1. August 2013 trat das Gesetz zur Förderung der elektronischen Verwaltung sowie zur Änderung weiterer Vorschriften vom 25. Juli 2013 (BGBl. I S. 2749) in Kraft. Dabei wurde unter anderem das Bundesstatistikgesetz vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 13 des

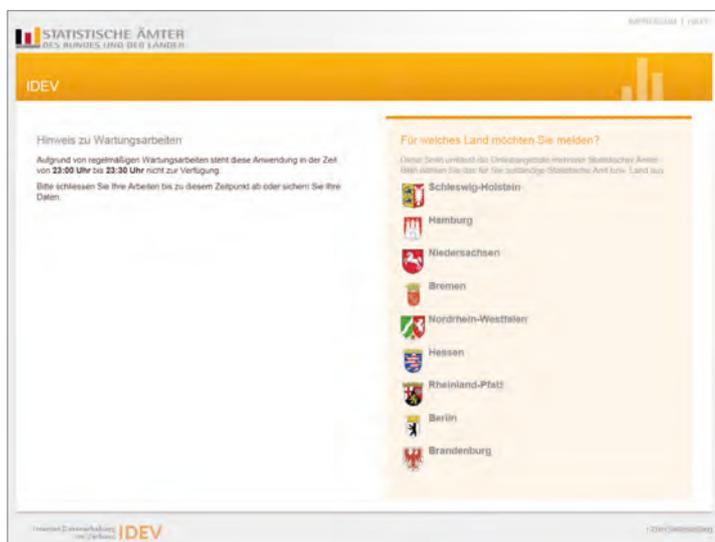
Gesetzes vom 25. Juli 2013 (BGBl. I S. 2745), u. a. um den § 11 a erweitert:

„§ 11 a: Elektronische Datenübermittlung

- (1) Soweit Stellen, die Aufgaben der öffentlichen Verwaltung wahrnehmen, Daten mittels standardisierter elektronischer Datenaustauschformate übermitteln, sind diese auch bei der Übermittlung der für eine Bundesstatistik zu erhebenden Daten zu verwenden. Ansonsten sind elektronische Verfahren nach Absprache der statistischen Ämter mit den betroffenen Stellen zu verwenden.
- (2) Werden Betrieben und Unternehmen für die Übermittlung der für eine Bundesstatistik zu erhebenden Daten elektronische Verfahren zur Verfügung gestellt, sind sie verpflichtet, diese Verfahren zu nutzen. Zur Vermeidung unbilliger Härten kann die zuständige Stelle auf Antrag eine Ausnahme zulassen.“

Diese gesetzliche Neuregelung erforderte auch im Amt für Statistik Berlin-Brandenburg (AfS) eine zeitnahe Umsetzung bei gleichzeitig konzentriertem und abgestimmtem Vorgehen in allen Statistikreferaten. Es galt, die Arbeitsaufgaben und -abläufe den neuen Bedingungen anzupassen und diese ggf. völlig neu zu gestalten. Die dafür notwendigen Voraussetzungen lieferte eine kurzfristig eingesetzte Arbeitsgruppe unter Beteiligung der Fachstatistikreferate und dem Referat Recht. Erarbeitete Handlungsempfehlungen sowie zahlreiche Fallbeispiele bilden seit Inkraft-Treten des Gesetzes die Voraussetzung dafür, den nunmehr gesetzlich festgelegten elektronischen Meldeweg einheitlich und effizient für die amtlichen Statistiken der Länder Berlin und Brandenburg umzusetzen. Bereits in den zurückliegenden Jahren hatte das AfS immer wieder berichtspflichtige Betriebe und Unternehmen in Berlin und Brandenburg mit gezielten, vor allem schriftlichen Informationen zu den elektronischen Meldeformen versorgt. Die Werbung zeigte jedoch bei den Auskunftgebenden in den einzelnen Statistiken eine sehr unterschiedliche Wirkung. Wurden bei einigen Statistiken durchaus bereits mehr als 50 % der Meldungen online vorgenommen, so stagnierte bei anderen Statistiken dieser Anteil bei 10 %.

Nachdem die Betriebe und Unternehmen schriftlich über die nun gesetzliche Verpflichtung zur Online-Meldung in Kenntnis gesetzt wurden, veränderte sich bereits im ersten Berichtsmonat 2013 nach Inkrafttreten des § 11 a BStatG das bisherige Meldeverhalten in den einzelnen Statistiken deutlich. Hatte beispielsweise in der monatlichen Tourismusstatistik für Berlin und Brandenburg zusammen der Anteil elektronischer Meldungen mittels IDEV bisher bei 45 % gelegen, stieg dieser



für die Meldung im September auf 80 % an und betrug zum Jahresende 97 %. Ähnlich deutliche Veränderungen in den Meldeformen zeigten auch andere Statistiken. So hatten bei der vierteljährlichen Verdiensterhebung vor Inkrafttreten der elektronischen Meldepflicht fast 20 % der Betriebe das Meldesystem eSTATISTIK.core (also die direkte Weiterleitung von Daten aus der betrieblichen Software) genutzt sowie weitere 35 % das IDEV-Verfahren. Diese Meldeform erreichte nunmehr ca. 79 %, bei einem nahezu unveränderten Anteil der eSTATISTIK.core-Meldungen.

Betrachtet man nach über einem halben Jahr die Veränderung der Meldeformen, so zeichnet sich für die im AFS durchzuführenden Statistiken folgendes erstes Bild ab: Monats- und weitere unterjährige Statistiken wurden bis zum Inkrafttreten der neuen gesetzlichen Regelungen zu durchschnittlich 20 % von den Unternehmen elektronisch gemeldet und anschließend in die jeweiligen Plausibilisierungsprogramme eingespielt. Zurzeit liegt dieser Anteil bei 91 %.

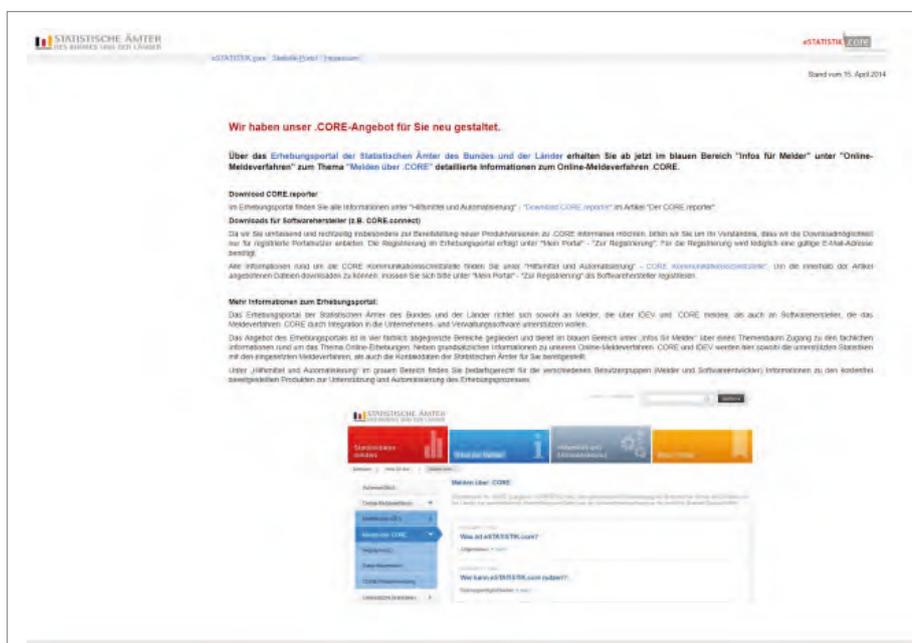
Hatten jährliche und auch mehrjährige Statistiken zusammen einen Anteil elektronischer Meldungen von etwa 15 % erreicht, stieg dieser auf nunmehr bereits 85 % an. Gerade bei den jährlichen und mehrjährigen Statistiken fällt allerdings auf, dass die Größe der Unternehmen bzw. Betriebe durchaus bei der Nutzung des Internets von Bedeutung sein kann, zumal eine deutlich höhere Zahl kleiner und mittlerer Unternehmen bzw. Betriebe im Vergleich zu den monatlichen Meldungen auskunftspflichtig sind. Verfügten laut Statistischem Bundesamt im Jahr 2013 rd. 87 % aller Unternehmen in Deutschland über einen Internetzugang [1], kommen neben der bereits erwähnten Unternehmensgröße auch der jeweiligen Branche, in der das Unternehmen tätig ist, der aufgabenbezogenen Nutzung des Internets (z. B. Social Media) sowie weiterer Kriterien eine

entscheidende Bedeutung zu. Ein wichtiger Faktor für die Internetnutzung durch das Unternehmen oder den Betrieb ist dabei das Vorhandensein einer stabilen und leistungsstarken Internetverbindung. Diese ist derzeit im Land Brandenburg noch nicht in allen Regionen gegeben, wie u. a. der Breitbandatlas des Bundesministeriums für Verkehr und digitale Infrastruktur verdeutlicht [2]. Dieser Umstand spiegelt sich möglicherweise auch darin wider, dass die Unternehmen mit Internetzugang in Brandenburg zu 44 % eine eigene Website betreiben, in Berlin zu 48 %, dagegen bundesweit bereits zu 66 %. Unternehmen mit einem Internetzugang nutzten aber das Netz zur Kommunikation mit der öffentlichen Verwaltung im Jahr 2013 in Berlin bereits zu 88 % und in Brandenburg zu 83 %, was weit über dem Durchschnitt aller Bundesländer lag (77 %) [1].

Die seit dem 1. August 2013 gesetzlich festgelegte elektronische Meldepflicht in der amtlichen Statistik trägt den hier beispielhaft aufgeführten Kriterien und damit auch dem derzeitigen technischen Stand Rechnung. Es sind deshalb ausdrücklich befristete Ausnahmeregelungen zugelassen, um sogenannte unbillige Härten für die Unternehmen zu vermeiden. Im Falle beispielsweise fehlender technischer Voraussetzungen ist auf Antragstellung durch den Auskunftgebenden weiterhin die Datenmeldung in Papierform möglich.

In den Unternehmen und Betrieben dürfte allerdings der Interneteinsatz und dessen Nutzungsspektrum in den kommenden Jahren einen deutlichen Anstieg erfahren, der sicherlich die weitere elektronische Abwicklung von Geschäftsprozessen zwischen den Unternehmen und der öffentlichen Verwaltung einschließen wird und damit auch die Wahrnehmung elektronischer Meldeverpflichtungen in der amtlichen Statistik.

Astrid Langer leitet das Referat *Handel, Gastgewerbe, Tourismus, Außenhandel* im Amt für Statistik Berlin-Brandenburg.



Literatur/Links

- [1] Unternehmen und Arbeitsstätten, Nutzung von Informations- und Kommunikationstechnologien in Unternehmen 2013; Statistisches Bundesamt, Wiesbaden 2013, sowie Pressemitteilungen vom Thüringer Landesamt für Statistik, Nummer 019/2014, 091/2014 bis 094/2014 http://www.tls.thueringen.de/presse/pr_kurzinh.asp
- [2] Zukunft Breitband, Informationsplattform zum Netzausbau; Bundesministerium für Verkehr und digitale Infrastruktur, 2014, S. 11 ff. http://www.zukunft-breitband.de/Breitband/DE/Breitbandatlas/breitbandatlas_node.html

Internationale Perspektive

Regionalstatistische Daten für Deutschland, Europa und weltweit

von Susanne Schnorr-Bäcker

Vergleichbare regionalstatistische Daten für Deutschland gewinnen zunehmend an Bedeutung. Grundlegend dafür sind regionale Klassifikationen und Standards. Besonders die Europäische Union, aber auch die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) haben wichtige Vorarbeiten geleistet, wie z. B. die Schaffung einer hierarchischen und überschneidungsfreien Gebietsgliederung oder die Festlegung von Kriterien zur Typisierung von Regionen. Diese sind unverzichtbar für harmonisierte grenzüberschreitende kleinräumige Betrachtungen, wie sie z. B. im Rahmen von „open government“ oder für die europäische Förder-, Umwelt- und Nachhaltigkeitspolitik gefordert werden. Ein Überblick über das derzeitige regionalstatistische Datenangebot aus der Bundesstatistik für Deutschland schließt die Betrachtungen ab.

1. Vorbemerkungen

In diesem Beitrag soll aufgezeigt werden, welchen Bedarf es an flächendeckenden und vergleichbaren regionalstatistischen Daten für Deutschland gibt und wie das Datenangebot dazu derzeit aussieht. Im Fokus stehen dabei die Anforderungen auf supra- und internationaler Ebene, vor allem seitens der Europäischen Union, der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) und der Vereinten Nationen. Nach einer kurzen Beschreibung der allgemeinen Rahmenbedingungen und der Darstellung der Gebietseinheiten soll im Einzelnen aufgezeigt werden, welcher Bedarf an regionalstatistischen Daten auf supra- und internationaler Ebene besteht. In diesem Zusammenhang wird auch dargestellt, welche Maßnahmen zur Harmonisierung und Standardisierung der nationalen Datenangebote bereits ergriffen wurden bzw. geplant sind. Auf dieser Grundlage sollen besonders europaweite Initiativen sowie ausgewählte inter- und supranationale Strategien und Projekte mit regionalstatistischem Bezug vorgestellt werden. Wie die damit verbundenen Anforderungen an flächendeckende, regionalstatistische Datenangebote für Deutschland aus dem Programm der Bundesstatistik, vor allem aus dem Gemeinschaftsprogramm der Statistischen Ämter des Bundes und der Länder, derzeit bedient werden können, soll zum Abschluss aufgezeigt werden.

2. Rahmenbedingungen

Vergleichbare und qualitativ hochwertige Daten der amtlichen Statistik sind ein unverzichtbares öffentliches Gut für Gesellschaft, Politik, Wirtschaft und Wissenschaft. Amtliche regionalstatistische Daten bieten eine wichtige Grundlage für die kleinräumige Planung und Entscheidungsvorbereitung sowie die Evaluation von politischen Maßnahmen. Dabei geht es vor allem um statistische Daten für Teilräume (wie z. B. Kreise oder Gemeinden, Arbeitsmarktregionen,

Wasserschutzgebiete u. Ä. m.). Welche Elemente als jeweils kleinste regionale Bausteine anzusehen sind, hängt im Wesentlichen von der Datenverfügbarkeit bzw. ihren Zugangsmöglichkeiten sowie dem jeweiligen Verwendungszweck ab.

Auch wenn politische Maßnahmen auf kleinräumiger Ebene vor allem zur Daseinsvorsorge, häufig – wie in Deutschland – gesetzlich geregelte Aufgaben eines jeweiligen Bundeslandes oder seiner Kommunen sind (einschließlich der damit verbundenen Datenhoheit), werden im Programm der amtlichen Statistik für ausgewählte statistische Merkmale und Regionen eine Vielzahl von weitgehend harmonisierten statistischen Nachweisungen angeboten.

Bereits seit Langem ist für einige wenige Statistiken eine kleinräumige Darstellung möglich, beispielsweise für ausgewählte Umweltstatistiken oder die Straßenverkehrsunfallstatistik.¹ Auch hat die Bundesstatistik bereits Erfahrungen mit der Georeferenzierung von Statistiken zu landwirtschaftlichen Betrieben gewonnen und dies erstmals auf einzelgesetzlicher Grundlage im Rahmen des Agrarzensus 2010 erproben können.²

Mit Inkrafttreten des sogenannten E-Government-Gesetzes³ am 1. August 2013 wurde das Bundesstatistikgesetz dahingehend geändert, dass nun generell eine Georeferenzierung statistischer Daten zulässig ist. Die bereits im Bundesstatistikgesetz (BStatG) bestehenden Regelungen für statistische Daten für

- Personen- und Haushaltsstatistiken (gemäß § 10 BStatG) und
- Wirtschafts- und Umweltstatistiken (§ 13 BStatG) wurden entsprechend modifiziert.⁴ So dürfen für Personen- und Haushaltsstatistiken die Merkmale in geografischen Gitterzellen von mindestens 100 m × 100 m unter Wahrung der bestehenden

1 bis 75

Information: Die Fußnoten des gesamten Beitrages befinden sich unter Anmerkungen auf den Seiten 26 und 27.

Löschungsvorschriften gespeichert werden. Für die Wirtschafts- und Umweltstatistiken dürfen exakte Geokoordinaten verwendet werden. Darüber hinaus gibt es im E-Government-Gesetz eine Regelung zur Georeferenzierung von öffentlichen und nicht-öffentlichen Registern für Datenbestände auf Grundlage einer Rechtsvorschrift des Bundes.⁵

3. Regionalstatistische Gebietseinheiten

3.1 Arten und Abgrenzung

Statistische Daten haben grundsätzlich eine zeitliche und eine räumliche Dimension. Das heißt, sie gelten für einen bestimmten Zeitpunkt oder Zeitraum und ein bestimmtes Gebiet. Wesensbestimmend für regionalstatistische Daten ist, dass sie für nationale Teilräume unterhalb einer bestimmten Größenordnung gelten. So besteht für Deutschland eine lange Tradition, erst statistische Daten unterhalb der Länderebene, das sind in der Regel Teilräume bzw. kleinere regionale Einheiten, als regionalstatistische Informationen zu bezeichnen.

In der Bundesstatistik haben sich vor allem zwei Arten von Regionalgliederungen herausgebildet:

- administrative Gebietseinheiten und
- nicht-administrative Gebietseinheiten.

Als administrative Gebiete werden vor allem solche Teilräume bezeichnet, deren Grenzen politisch vorgegeben sind und die das allgemeine Hoheitsgebiet einer öffentlichen Verwaltung bzw. Gebietskörperschaft darstellen. Die administrative Gliederung Deutschlands auf den verschiedenen föderalen Ebenen setzt sich aus Bund, Ländern, Regierungsbezirken, Landkreisen/kreisfreien Städten und Gemeinden/Gemeindeverbänden zusammen. Diese Untergliederung ist klar definiert durch rechtsverbindliche Grenzen, überschneidungsfrei und hierarchisch strukturiert.

Alle anderen Gebietsabgrenzungen werden als nicht-administrative Gebietseinheiten bezeichnet. Diese lassen sich weiter nach den Kriterien Homogenität und Funktionalität unterscheiden. Maßgeblich für ersteres ist die Abgrenzung hinsichtlich der Ähnlichkeit bezüglich bestimmter Eigenschaften bzw. Merkmale (z. B. Siedlungsdichte, Einkommensniveau, bestimmte geo- oder topografische Eigenschaften wie Küstenregionen oder Berggebiete).

Bei einer Abgrenzung hinsichtlich der Funktionalität wird auf die Interdependenzen oder wechselseitigen Abhängigkeiten in einem bestimmten Gebiet abgestellt. Häufig geht es dabei um Verflechtungen von Orten im Raum in wirtschaftlicher, sozialer oder ökologischer Hinsicht (wie z. B. bei Arbeitsmarktregionen, ökologischen Systemen wie Naturschutzgebieten oder Wassereinzugsgebieten).

3.2 Regionale Bausteine

Um Regionen statistisch abbilden zu können, werden kleinste regionale Elemente oder Bausteine benötigt. Grundsätzlich lassen sich dabei vier verschiedene Formen unterscheiden⁶:

- Punkt,
- Linie,
- Vektor,
- Rasterzelle.

Bei statistischen Angaben zu Punkten handelt es sich um geografische Orte von geringer räumlicher Ausdehnung (z. B. Straßenverkehrsunfallsschwerpunkte oder Wassereinleitstellen). Derartige Angaben sind in der amtlichen Statistik nur in Ausnahmefällen zulässig, wenn sie in einem entsprechenden Fachstatistikgesetz explizit genannt sind.⁷

Statistische Nachweisungen in Form von Linien kommen in der Bundesstatistik ebenfalls relativ selten vor. Beispiele dafür sind netzgebundene Verkehrswege (wie z. B. im Güterverkehr auf der Straße, der Schiene oder auf Wasserstraßen). Als Hintergrundinformationen für kartografische Darstellungen bieten sie Ansatzpunkte für mögliche Erklärungen oder weiterführende Analysen.

Bei Vektoren handelt es sich in der Regel um Teilräume mit einem unregelmäßigen Grenzverlauf, wie dies beispielsweise bei Gemeinden oder Gemeindeteilen der Fall ist. In der Bundesstatistik erfolgt ihre Abgrenzung vor allem anhand von Identifikatoren (z. B. dem amtlichen Gemeindegeschlüssel (AGS) oder Postleitzahlen).

Lange Tradition und weit verbreitet sind in der Geografie sogenannte Rasterzellen. Es handelt sich dabei üblicherweise um quadratische Raumeinheiten in metrischer Form in einer Größe von 100 m x 100 m oder 1 km². In der Bundesstatistik waren derartige Darstellungen lange nicht möglich.⁸

Mit Inkrafttreten der Änderungen im Bundesstatistikgesetz auf der Grundlage des E-Government-Gesetzes sind nunmehr sämtliche Statistiken entsprechend darstellbar, vorausgesetzt, sie sind fachlich dafür geeignet. Rasterzellen bieten gegenüber Vektoren sowohl von der Datenhaltung wie auch von der Darstellung her Vorteile. Sie erlauben eine bessere Vergleichbarkeit statistischer Daten sowohl in räumlicher wie auch in zeitlicher Hinsicht. Unter räumlichen Aspekten ermöglichen sie den Vergleich von Raumeinheiten gleichgroßer Fläche⁹. Auch in zeitlicher Hinsicht haben sie Vorteile, vor allem wenn sich die Gebietsgrenzen von Teilräumen im Zeitablauf ändern, wie dies für Deutschland beispielsweise bei Gebietsstandsänderungen von Gemeinden der Fall ist¹⁰.

Von besonderer Bedeutung waren bislang als kleinste regionalstatistische Einheiten in der Bundesstatistik die Gemeinden aus dem Gemeindeverzeichnis¹¹.

In Zukunft können Rasterzellen zur Datenerhebung und Darstellung für die Personen- und Haushaltsstatistiken verwendet werden; für die Wirtschafts- und Umweltstatistiken dürfen zur exakten Lokalisierung einer Anlage, eines Betriebes oder eines Unternehmens Geokoordinaten gespeichert werden. Auf der Grundlage dieser Bausteine lassen sich die entsprechenden realen Phänomene in der Bundesstatistik auch unterhalb der Ebene der Gemeinde erfassen und gegebenenfalls – unter Wahrung der bestehenden Geheimhaltungsregeln – genauer darstellen als auf der Grundlage von administrativen Einheiten.

Diese verschiedenen Möglichkeiten sollen anhand eines Beispiels zur durchschnittlichen Größe landwirtschaftlicher Betriebe aus der Landwirtschafts-

zählung 2010 kurz dargestellt werden. Die beiden Karten (siehe Abbildung a) zeigen ähnliche Sachverhalte. Hierbei geht es um die durchschnittliche landwirtschaftliche Betriebsgröße für das Gebiet im Bereich Wiesbaden und Mainz. Ungeachtet der Klassifizierung werden im Regionalatlas – mit administrativen Gebietseinheiten – die Werte für die durchschnittliche Größe von Agrarbetrieben wesentlich weniger differenziert und für eine größere Fläche angezeigt als im Agraratlas, der Angaben für Rasterzellen enthält: Während der Regionalatlas in der Abbildung für das ausgewählte Gebiet für die durchschnittliche Betriebsgröße nur einen Wert (10 bis 84 Hektar) anzeigt, enthält der Agraratlas auch Rasterzellen mit größeren Betriebseinheiten (mit z. B. 100 bis 200 Hektar bzw. 200 bis 400 Hektar). Die Verwendung von Rasterkarten bedeutet allerdings auch, dass Raster Verwaltungsgrenzen – z. B. zwischen Kreisen und sogar Bundesländern – überschreiten können, was u. a. besondere Anforderungen an die Geheimhaltung stellt.

Das bedeutet für Rasteranwendungen, dass ergänzende Informationen zu Verwaltungsgrenzen oder anderen topografischen Besonderheiten für die Interpretation eine noch stärkere Bedeutung haben als für weniger differenzierte räumliche Darstellungen auf der Grundlage von administrativen Einheiten. In beiden Online-Anwendungen lassen sich weitere Informationen, z. B. zu administrativen Gebietsgrenzen (in Schwarz die Grenzen der Bundesländer), Naturräumen (in Braun) oder Flüssen (in Blau), hinzufügen.

3.3 Regionalklassifikationen und Gebietstypisierungen

Da die politische Planung und Entscheidungsvorbereitung auf regionaler Ebene den räumlichen Besonderheiten in siedlungsstruktureller und/oder funktionsräumlicher Hinsicht Rechnung tragen muss, wurden von den entsprechenden Entscheidungssträ-

gern – teilweise auch in Abstimmung mit der amtlichen Statistik – in Deutschland und mehr noch auf der Ebene der Europäischen Union oder der OECD entsprechende Gebietsklassifikationen festgelegt.

Wichtige Abgrenzungskriterien sind dabei, neben Siedlungsdichte, d. h. der Einwohnerzahl je km², vor allem die folgenden funktionsräumlichen Kriterien zur Typisierung von Regionen oder verstärkten Gebieten: Die Erreichbarkeit eines Ortes (z. B. mit öffentlichen Nahverkehrsmitteln) sowie die Zugangsmöglichkeiten vor allem in ökonomischer Hinsicht (gemessen z. B. an den Kosten oder dem Zeitaufwand für den Transport); in der anglo-amerikanischen Literatur werden diese Kriterien auch als „accessibility“ und „affordability“ bezeichnet.

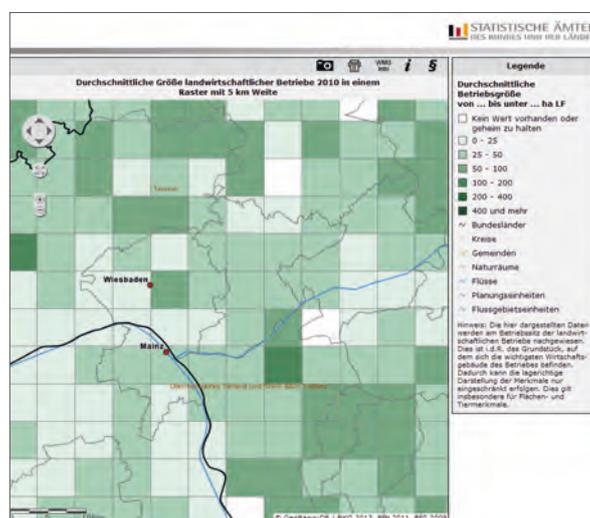
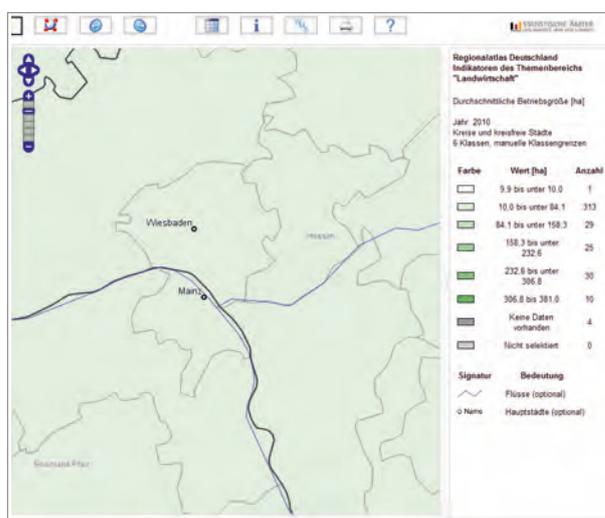
– Für Deutschland

Für Deutschland ist das System der sogenannten zentralen Orte wichtigstes raumordnerisches, flächendeckendes Instrument und im Raumordnungsgesetz geregelt¹². Es ist Grundlage für die überörtliche Leistungserbringung im Rahmen der Daseinsvorsorge in Form von Infrastruktureinrichtungen sowie für die Entwicklung des ländlichen Raums. Dieses hierarchische System besteht im Wesentlichen aus drei Funktionsstufen: sogenannten Ober-, Mittel- und Unter-/Grund-/Kleinzentren. Ihre Abgrenzung und Zuordnung erfolgt in Raumordnungsplänen für das Gebiet eines jeweiligen Bundeslandes. Dabei orientieren sich die landesplanerischen Ausweisungen an Ausstattungskatalogen der Länder.

Die Zuordnung einer Kommune zu einem bestimmten Zentralen Orte-Typ erfolgt auf der Grundlage des Angebots an Gütern und Dienstleistungen für Privathaushalte und der Erreichbarkeit dieser Orte mit öffentlichen Nahverkehrsmitteln^{13, 14}:

- Die Oberzentren (insgesamt mehr als 100) zeichnen sich durch das an einem Ort vorhandene Angebot für einen spezialisierten höheren Bedarf an

a | Kartografische Darstellungen anhand von administrativen Gebietseinheiten und Rasterzellen



Quelle: Regionalatlas

<https://www-genesis.destatis.de/gis/genView?GenMLURL=https://www-genesis.destatis.de/regatlas/ai009.xml&CONTEXT=REGATLAS01> und

Agraratlas <http://www.atlas-agrarstatistik.nrw.de/>

Gütern und Dienstleistungen für private Haushalte aus; zudem sollen sie mit öffentlichen Verkehrsmitteln in maximal 90 Minuten erreichbar sein.

- In Mittelzentren (insgesamt rund 1 000) findet man Güter des gehobenen Bedarfs; zu ihrem Einzugsbereich zählen rund 30 000 bis 35 000 Einwohner und diese Zentren sollen mit öffentlichen Verkehrsmitteln in maximal 45 Minuten erreichbar sein.
- Zentrale Orte der unteren Stufe sollen den Grundbedarf der Bevölkerung im Nahbereich decken, wobei das Einzugsgebiet mindestens 7 000 bis 10 000 Einwohner betragen soll.

– Für Europa

Auf europäischer Ebene bilden ebenfalls administrative Einheiten und Rasterdaten die kleinsten Bausteine für entsprechende Gebietstypisierungen bzw. Regionalklassifikationen. Die NUTS-Klassifikation (Nomenclature des unités territoriales statistiques) ist eine Systematik der Gebietseinheiten für eine hierarchische Untergliederung des EU-Wirtschaftsraums.

Die Abgrenzung der NUTS-Ebenen basiert auf geografisch festgelegten Verwaltungseinheiten, in der Regel die beiden obersten regionalen Hauptebenen 1 und 2; für die zusätzliche dritte Ebene werden kleinere Verwaltungseinheiten zusammengefasst (Abbildung b).

Für die Zuordnung der Gebietseinheiten zu den verschiedenen regionalen Ebenen sind Unter- und Obergrenzen für die Größe, gemessen an den Einwohnerzahlen, festgesetzt, die aber aufgrund der unterschiedlichen Größe der europäischen Mitgliedstaaten nicht immer eingehalten werden können.

Für NUTS-3-Einheiten – für Deutschland sind das die Kreise – sollte die Einwohnerzahl zwischen 150 000 und 800 000 Einwohnern liegen, für NUTS-2-Einheiten (für Deutschland sind das die Regierungsbezirke oder sogenannte Statistische Regionen) sollte die Bevölkerungszahl zwischen

800 000 und 3 Mill. liegen und für NUTS-1-Einheiten (in Deutschland sind das die Bundesländer) sollte die Bevölkerung zwischen 3 und 7 Mill. Personen aufweisen.

Die NUTS-Klassifikation wurde bereits 1970 für die Regionalstatistik in der damaligen Europäischen Gemeinschaft konzipiert und mehrere Jahrzehnte auf der Grundlage von so genannten Gentlemen's Agreements zwischen den Mitgliedstaaten und Eurostat angewandt und aktualisiert. Seit 2003 basiert diese Gebietssystematik auf einer Rechtsgrundlage¹⁵, der Verordnung (EG) Nummer 1059/2003. Sie wird regelmäßig¹⁶ im Dreijahresturnus überprüft und modifiziert. Die Mitgliedstaaten sind verpflichtet, innerhalb von zwei Jahren ihre (historischen) Daten entsprechend zu aktualisieren, das heißt dem aktuellen Gebietsstand anzupassen.¹⁷

Für den Bedarf an lokalen europaweit vergleichbaren Statistiken wurde auf europäischer Ebene ein mit der NUTS-Klassifikation kompatibles System lokaler Verwaltungseinheiten (local administrative units bzw. LAU) festgelegt. Es besteht aus einer oberen LAU-Ebene (LAU 1: für Deutschland sind das Gemeindeverbände) sowie einer unteren LAU-Ebene (LAU 2: für Deutschland sind das Gemeinden). Gemäß den Bestimmungen der NUTS-Verordnung müssen die EU-Staaten dem Statistischen Amt der Europäischen Union (Eurostat) jährlich eine vollständige Liste ihrer lokalen Verwaltungseinheiten zuleiten.

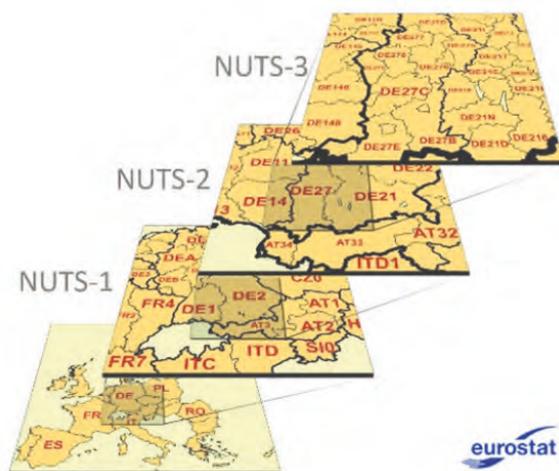
Alle gemeinschaftsweiten Regionalstatistiken beziehen sich auf die NUTS-Klassifikation. Für einige europäische Regionalstatistiken wurde auf der Grundlage von Postleitzahlen ein Referenzschlüssel zur geografischen Identifizierung entwickelt. Er ermöglicht eine Zuordnung von Postleitzahlen zur NUTS-3-Ebene und wird benötigt für Projekte, z.B. von Eurostat, anderen Kommissionsdienststellen oder der Europäischen Investitionsbank. Auch für öffentliche Einrichtungen außerhalb der Bundesstatistik hat die europäische Regionalklassifikation NUTS eine gewisse Bedeutung, wenn sie entsprechende Informationen für Deutschland an Europäische Dienststellen melden müssen.

Auch die OECD beschäftigt sich mit kleinräumigen Entwicklungen in ihren Mitgliedstaaten. Sie verwendet dafür eine eigene Gebietsgliederung „Territorial Levels“ (TL), mit drei verschiedenen Ebenen: TL1 „Nationale Staaten“, TL2 „Regionen“ sowie TL3 „kleinere Gebietseinheiten“. Sie ist ebenfalls hierarchisch, d.h. überschneidungsfrei, konzipiert und baut für Deutschland auf den Kreisen auf.

Neben dieser europaweiten Gebietsgliederung auf administrativer Ebene finden zunehmend auch Rasterdaten von 1 km² Verwendung. Da Rasterdaten bislang nur für die Minderzahl der europäischen Mitgliedstaaten zulässig sind (auch für Deutschland war eine generelle Darstellung in der Bundesstatistik auf der Grundlage von Rastern bis zum 31. Juli 2013 nicht möglich), hat Eurostat eine Pilotstudie initiiert, um diese Daten aus anderen, bereits vorliegenden Quellen abzuleiten¹⁸.

Diese beiden Bausteine, d.h. kleinste administrative Einheiten auf LAU-Ebene sowie Rasterzellen von

b | NUTS-Klassifikation



Quelle:

http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

1 km², bilden die Grundlage für siedlungsstrukturelle oder funktionsräumliche Gebietstypisierungen für die Mitgliedstaaten der Europäischen Union.

Im Folgenden werden die wichtigsten Typen, die von der Europäischen Kommission in enger Zusammenarbeit mit der OECD entwickelt wurden, näher beschrieben (Abbildung c). Dazu gehören eine

- 1) Abgrenzung ländlicher von städtischen Gebieten (DegUrba),
- 2) Typisierung von Städten und verstäderten Regionen,
- 3) Abgrenzung von Metropolregionen.

ad (1) Abgrenzung ländlicher von städtischen Gebieten (DegUrba)

Auf europäischer Ebene wurde ein integriertes System zur Gebietstypisierung auf der Grundlage von LAU-2-Einheiten entwickelt. Dieses diente zunächst zur Abgrenzung des städtischen Raums vom ländlichen Raum. Diese Typisierung „DegUrba (Degree of Urbanisation)“²⁰ wurde um eine Gebietstypisierung nach Verstädterungsgrad ergänzt und später noch für verschiedene Stadt(teil)typen wie Kernstadt, Agglomeration und Metropolregionen verfeinert.

Zur Abgrenzung dieser drei Gebietstypen werden folgende Merkmale herangezogen: Zum einen die Siedlungsdichte, d.h. die Wohnbevölkerung je km², und zum anderen die Gemeindegröße gemessen an der Einwohnerzahl (Abbildung d).

Bei DegUrba werden die folgenden drei Gebietstypen unterschieden:

- dünn besiedeltes Gebiet (ländlicher Raum), also Gebiete mit einer Siedlungsdichte von weniger als 100 Personen je km² und einer Gemeindegröße von weniger als 50 000 Einwohnern,
- Gebiete mit mittlerer Besiedlungsdichte (z. B. kleinere Städte und Vorstädte), also Gemeinden mit einer Siedlungsdichte zwischen 100 und unter 500 Personen je km² und einer Gemeindegröße von mindestens 50 000 Einwohnern,
- dicht besiedeltes Gebiet (Städte oder großstädtischer Raum), d.h. Rasterzellen mit einer Bevöl-

kerungsdichte von mindestens 500 Einwohnern je km² und eine Wohnbevölkerung von mehr als 50 000 Personen.

Die Gebietstypisierung DegUrba kommt bereits für einige Gemeinschaftsstatistiken zur Anwendung wie z.B. für ausgewählte Tourismus- oder Arbeitsmarktstatistiken. Im Gemeindeverzeichnis des Statistischen Bundesamtes ist die Zuordnung der Gemeinden zu den jeweiligen DegUrba-Typen²¹ enthalten.

ad (2) Typisierung von Städten und verstäderten Regionen

Diese Gebietstypisierung wurde 2013 gemeinsam von der Europäischen Kommission und der OECD erarbeitet und festgelegt.²² Sie gilt für verstäderte Gebiete mit einer Siedlungsdichte von mehr als 1 500 Einwohnern je Quadratkilometer und Städte mit mehr als 50 000 Einwohnern. Danach lassen sich folgende Gebietstypen unterscheiden:

- a) Urbane Center in verschiedenen Ausprägungen je nach Einwohnerzahl reichen von kleinen Urbanen Centern mit bis zu 100 000 Einwohnern über sogenannte XXL-Urbane Center mit einer Einwohnerzahl von 1 Mill. bis 5 Mill. Einwohnern bis hin zu globalen Urbanen Centern mit mehr als 5 Mill. Einwohnern; Urbane Center sind Kern einer urbanen Stadt.
- b) Urbane Stadt (d. h. die Stadt im Sinne der europäischen Städteerhebung Urban Audit²³).
- c) Verstäderte Region (larger urban zone).

Es handelt sich bei dieser Gebietstypisierung von den konkreten Abgrenzungskriterien her um eine hierarchische Gliederung, d.h. jede nachfolgende Stufe baut auf der vorherigen auf. So ist das Urbane Center die Voraussetzung für eine Urban-Audit-Stadt unter Berücksichtigung von zwei weiteren Kriterien: Erstens soll mindestens die Hälfte der Bevölkerung der Urban-Audit-Stadt im Urbanen Center leben und zweitens sollen mindestens 75% der Bevölkerung des Urbanen Centers in der Stadt leben (siehe dazu auch Abbildung e).

c | Gebietsgliederungen auf europäischer Ebene¹⁹



Quelle: http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/documents/Defining_urban_areas.pdf

d | Deutschland nach der europäischen Klassifikation DegUrba



Quelle: http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=D-SP_DEGURBA

Ausgehend von diesem Städtetyp wird die sogenannte verstädterte Region gebildet, d.h. Kern der verstädterten Region ist eine Urban-Audit-Stadt. Hinzu kommen die folgenden Kriterien: Weniger als 15% der Beschäftigten der Urban Audit-Stadt leben in einer anderen Stadt und mindestens 15% der Wohnbevölkerung der Urbanen Stadt arbeiten am Wohnort. Und schließlich gilt drittens für die verstädterte Region/Larger Urban Zone, dass sämtliche Kommunen dieses Gebiets zu einem räumlich zusammenhängenden funktionalen Gebiet gehören. Diese drei Gebietsgliederungen kommen u.a. in der Europäischen Städterhebung Urban Audit²⁴ zur Anwendung. Für Deutschland wird dieses Projekt gemeinschaftlich, auch zusammen mit dem Statistischen Bundesamt, abgewickelt: Die Daten werden von einem offiziellen externen Datenlieferanten der KOSIS-Gemeinschaft Urban Audit als eigene Angelegenheit ermittelt und zusammengestellt.²⁵

Im Weiteren werden für diese Gebietstypen noch Spezialfälle unterschieden, auf die hier nicht weiter eingegangen werden soll.²⁶ Die Städte in Europa, vor allem ihre räumliche Verteilung, lassen sich wie in Abbildung f darstellen:²⁷

ad (3) Abgrenzung von Metropolregionen

Ausgangspunkt sind wieder sogenannte Urbane Center, für die aufgrund der Siedlungsstruktur in Europa die unter (2)a) genannten Kriterien gelten.²⁸ Danach werden alle umgebenden Rasterzellen einbezogen, in denen mehr als 15% der Wohnbevölkerung gleichzeitig in dem Urbanen Center beschäftigt sind. In die Betrachtung können noch weitere soziodemografische, ökonomische und ökologische Merkmale einbezogen werden, die für die 275 Metropolregionen mit mehr als 500 000 Einwohnern der über 29 OECD Mitgliedstaaten in der Datenbank der OECD gespeichert sind.

Eine andere Abgrenzung hat die Europäische Kommission für Metropolregionen gewählt.²⁹ Danach sind alle NUTS-3-Einheiten (für Deutschland die Kreise) einzeln oder in unmittelbar räumlicher Nähe von Städten mit mehr als 250 000 Einwohnern sogenann-

te Urbane Agglomerationen. Dabei werden folgende Arten unterschieden: (1) Hauptstädte und in Abhängigkeit von der Bevölkerungszahl (2) sekundäre Metropolregionen. Diese lassen sich – je nach wirtschaftlicher Bedeutung, vor allem bezüglich Arbeitsmarkt, Transportverbindungen und/oder ausgewählten Bildungseinrichtungen wie z.B. Universitäten – weiter untergliedern, oder sie werden auf nationaler Ebene festgelegt.

Einen Überblick über die wichtigsten kleinräumigen Gebietstypisierungen und Abgrenzungskriterien für statistische Zwecke auf europäischer Ebene gibt Tabelle 1.

4. Bedarf an regionalstatistischen Daten auf supra- und internationaler Ebene

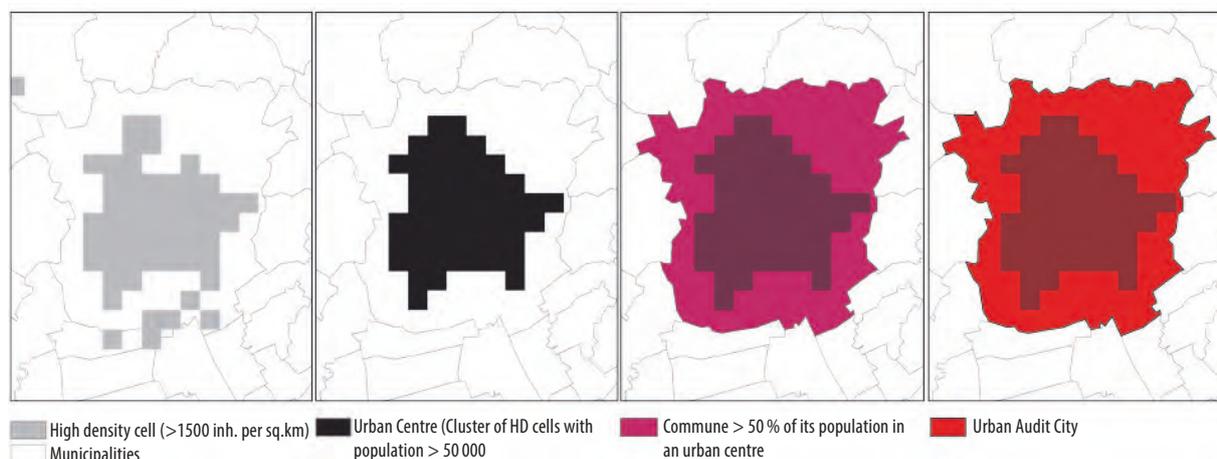
Die nachfolgenden Ausführungen konzentrieren sich auf die Anforderungen und Maßnahmen auf supra- und internationaler Ebene, da hiervon derzeit die größten Impulse auf eine Harmonisierung des gemeinschaftsweiten wie auch nationalen regionalstatistischen Programms ausgehen.

Vor allem auf der Ebene der Europäischen Union gibt es zahlreiche Initiativen, die von Bedeutung für die nationale Regionalstatistik in den Mitgliedstaaten sind, ungeachtet der jeweiligen Institutionen, die entsprechende Daten bereitstellen.

Arbeitsschwerpunkt bei Eurostat, auch in enger Zusammenarbeit mit der OECD, ist derzeit die Schaffung einer qualitativ hochwertigen und vergleichbaren Regionalstatistik zur Beurteilung der Lebens- und Arbeitsbedingungen in den verschiedenen Mitgliedstaaten auch auf subnationaler Ebene. Derartige Informationen werden nicht nur für verschiedene Politikbereiche benötigt, sondern sie dienen zur Information von Bürgerinnen und Bürgern, gesellschaftlichen, wirtschaftlichen, politischen, wissenschaftlichen Gruppierungen, Verbänden und sonstigen Einrichtungen. Besondere Einflüsse auf die europäische Regionalstatistik gehen aus von

- 1) einer weltweiten Open-Government-Politik,
- 2) der europäischen Kohäsions- und Regionalpolitik,
- 3) einer grenzüberschreitenden Umweltpolitik in einer globalisierten Welt.

e | Europäische Gebietstypisierung für Städte und verstädterte Regionen am Beispiel von Graz



Quelle: Cities in Europe, The new OECD-EC definition, http://ec.europa.eu/regional_policy/sources/docgener/focus/2012_01_city.pdf, hier S. 2.

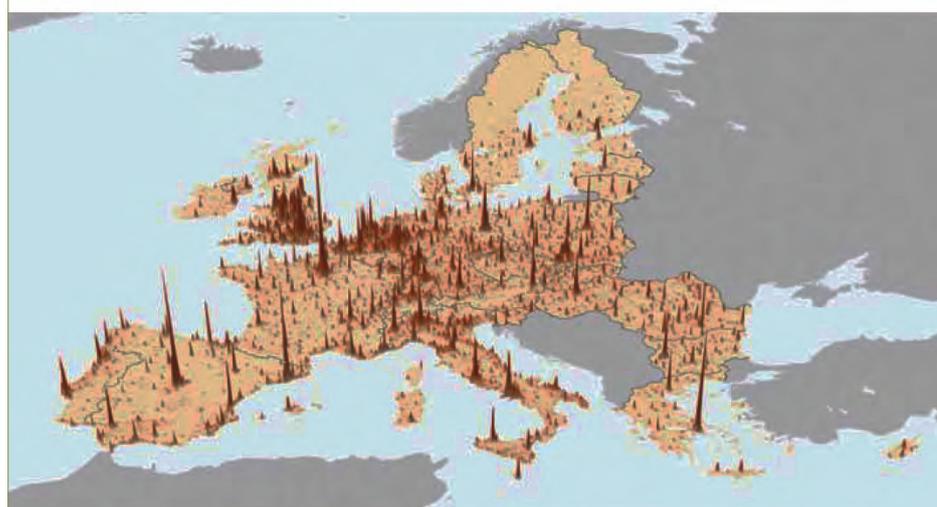
1 | Europäische Gebietstypisierungen

Name	Bezeichnung	Bausteine		Arten	Kriterien		
		Administrativ (LAU)	Raster (1 km)		Siedlungsdichte	Einwohnerzahl	Sonstige
DegUrba	Regional-typisierung	LAU 2	1 km	hoch mittel niedrig	> 500 100 - 500 < 100	> 50 000 > 50 000 < 50 000	
European City EU/ OECD	Urbane Center	LAU 2	1 km	Urban Center (UC): Arten nach Einwohnergröße S 50 000 - 100 000 M 100 000 - 250 000 L 250 000 - 500 000 XL 500 000 - 1 Mill. XXL 1 Mill. - 5 Mill. Global City > 5 Mill.	> 1 500	> 50 000	
	City	LAU 2	1 km	City (=Urban Audit City) Spezialfall: Urban center > city (=greater city, nicht für Deutschland)	> 1 500	> 50 000	1) > 50 % der Bevölkerung der Stadt (LAU 2) lebt im UC 2) 75 % des UC lebt in der City
	Larger Urban Zone	LAU 2	1 km	Larger Urban Zone	> 1 500	> 50 000	1) < 15 % der Beschäftigten einer Stadt wohnt in einer anderen Stadt 2) > 15 % der Wohnbevölkerung arbeitet in der Stadt 3) funktionale Verflechtungen mit weiteren LAU-2-Einheiten zu einem zusammenhängenden Gebiet
Metropol-region OECD	Metropol-regionen	LAU 2	1 km	Hauptstadt	> 1 500 für Europa > 1 000 für US etc.	> 50 000 für Europa > 100 000 für Mexiko, Japan	Pendlerverflechtungen (verschiedene Definitionen)

Quelle: Eigene Darstellung

f | Räumliche Verteilung der Städte in Europa

Map 1 Population density in Europe, 2001



Note: Colours and peak heights represent population density by 1km² raster cells. Sources: Eurostat, JCR, EFGS, REGIO-GIS

Quelle:
http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/citiesoftomorrow/citiesoftomorrow_final.pdf, hier S. 3.

Im Folgenden sollen kurz die wichtigsten Kernelemente skizziert und anhand von Beispielen für die amtliche Regionalstatistik veranschaulicht werden.

ad (1) Open Government

Daten und Informationen sind grundsätzlich wesentliche Ressourcen einer Informationsgesellschaft. Sie bieten nicht nur eine wichtige Grundlage für politische Planungen und Entscheidungen, sondern sie ermöglichen auch der Wirtschaft neue Geschäftsmodelle und Kooperationen. Für wissenschaftliche Vorhaben sind sie eine unverzichtbare Grundlage.³⁰

Im Zuge der zunehmenden Digitalisierung aller Bereiche von Politik, Wirtschaft, Wissenschaft und Gesellschaft haben sich verschiedene Bewegungen mit „open...“ wie z. B. „open access“ oder „open source“ herausgebildet. Für den Bereich der öffentlichen Verwaltung ist dies „Open Government“. Open Government bezeichnet ein offenes Regierungs- und Verwaltungshandeln. Dafür ist eine neue, transparente und auf Zusammenarbeit ausgelegte Verwaltungskultur notwendig. Open Government gilt grundsätzlich für alle Verwaltungsebenen und -einrichtungen in Deutschland³¹ sowie Europa³² und wird weltweit³³ empfohlen. Es stützt sich auf verschiedene Rechtsakte, besonders auf das Informationsfreiheitsgesetz (IFG)³⁴ sowie das Informationsweiterverwendungsgesetz (IWG)³⁵ auf Bundesebene. Open Government dient vor allem folgenden Zielen: Der Schaffung von

- Transparenz (transparency),
- Teilhabe (participation) sowie der
- Rechenschaftslegung (accountability).

Das bedeutet, Regierungshandeln und Verwaltungsvorgänge sollen leichter nachvollziehbar werden; Informationen sollen bereitgestellt werden, die es jedem Einzelnen in der Zivilgesellschaft ermöglichen, sich an gesellschaftlichen und politischen Prozessen aktiv zu beteiligen. Ein Beispiel dafür ist eine neue Form der Bürgerbeteiligung in der Europäischen Union. Sie wurde mit dem Vertrag von Lissabon

eingeführt. Es handelt sich um die sogenannte Europäische Bürgerinitiative (EBI).³⁶ Sie ermöglicht den EU-Bürgerinnen und -Bürgern ein Mitspracherecht bei der Gesetzgebung der EU.³⁷

Um die Verwaltungsdaten technisch an einer Stelle auffindbar und einfacher nutzbar zu machen, wurde für Deutschland ein Datenportal „GovData“ für das Internet entwickelt.³⁸ Es bietet einen einheitlichen und zentralen Zugang zu Verwaltungsdaten von Bund, Ländern und Kommunen und beinhaltet nicht nur offene Daten, sondern auch solche, die eingeschränkt nutzbar sind. Im Weiteren gibt es einen Überblick über Fachportale oder solche von Ländern und Kommunen (Abbildung g).

Vorreiter für Open Data und Open Government in Deutschland ist das Statistische Bundesamt mit seinem Angebot von über 390 Bundesstatistiken, von denen ca. 70% auf einer europäischen Rechtsgrundlage basieren und somit europaweit harmonisiert sind.³⁹

Open Government ist eine Plattform für die zunehmend evidenzbasierten und faktenorientierten politischen Strategien und Maßnahmen – sowohl bereichsübergreifende wie auch bereichsspezifische – auf nationaler, supra- und internationaler Ebene.

Für Deutschland haben derartige Strategien bereits eine lange Tradition. Das bedeutet, immer häufiger werden politische Strategien anhand von geeigneten statistischen Indikatoren, überwiegend aus der amtlichen Statistik, überprüft und ggf. modifiziert. Beispiele dafür sind:

- Die „Nachhaltige Entwicklung in Deutschland“ als bereichsübergreifende Strategie mit den Bereichen Generationengerechtigkeit, Lebensqualität, sozialer Zusammenhalt und Wahrung der internationalen Verantwortung; das Statistische Bundesamt zeigt für 38 ausgewählte Kennzahlen, welche Ziele erreicht wurden bzw. wo die Entwicklungen hinter den Erwartungen zurückbleiben.^{40, 41} Dazu gehören auch zahlreiche themen- oder bereichsspezifische politikbegleitende Indikatorensysteme zu Integra-

g | Datenangebot aus GovData

The screenshot shows the GovData portal interface. At the top, there is a navigation bar with links for 'Daten', 'Dokumente', 'Apps', 'Neues', 'Bibliothek', 'Das Portal', and 'Anmelden'. Below the navigation bar, a search bar contains the text '3099 Treffer'. The main content area is divided into several sections:

- Offenheit der Lizenz:** A sidebar on the left with radio buttons for 'Alle', 'Freie Nutzung (3059)', and 'Eingeschränkte Nutzung (40)'.
- Kategorien:** A sidebar on the left with checkboxes for various categories like 'Bevölkerung (2120)', 'Infrastruktur, Bauen und Wohnen (2000)', 'Wirtschaft und Arbeit (603)', etc.
- Search Results:** A list of search results with columns for 'Sortieren nach', 'Relevanz', 'Name', and 'Änderung'. The first result is 'Europawahl 2014, Ergebnisse der Bundesrepublik Deutschland' with a 'Freie Nutzung' button. The second result is 'Amtliche Ergebnisse der Wahl zum 8. Europäischen Parlament 2014 in Deutschland' with a 'Freie Nutzung' button. The third result is 'Veröffentlichende Stelle: Bundeswahlleiter' with a 'Freie Nutzung' button. The fourth result is 'Monatszahlen für Bevölkerung, Erwerb, Soziales, Bautätigkeit, Landwirtschaft, Gewerbe, Handel, Verkehr, ...' with a 'Freie Nutzung' button. The fifth result is 'Webanwendung mit 230 Merkmalen der amtlichen Statistik zu Daten der Amtlichen Statistik seit dem Jahr 2000. U. a. zur Bevölkerung, Wanderungen, Erwerbstätigkeit, Arbeitsmarkt, SGB II, Bau...' with a 'Freie Nutzung' button. The sixth result is 'Veröffentlichende Stelle: Statistisches Amt für Hamburg und Schleswig-Hol...' with a 'Freie Nutzung' button. The seventh result is 'Liste der Wahlkreise zur Bezirksversammlungswahl 2014 in Hamburg' with a 'Freie Nutzung' button.

Quelle:

<https://www.govdata.de/web/guest/daten>

tion⁴², Gleichstellung von Männern und Frauen⁴³, Gesundheit⁴⁴ oder zur schulischen Bildung.⁴⁵ Alle diese Indikatorensysteme sind zumindest auf nationaler Ebene, d. h. für Deutschland insgesamt verfügbar, für einige gibt es länderspezifische Modifikationen wie z. B. zur Integration.⁴⁶ Einige wenige liegen bis zur kommunalen Ebene vergleichbar vor wie z. B. die Bildungsberichterstattung.⁴⁷ Die Abbildung h enthält ein Beispiel zu Schulen in freier Trägerschaft und jeweiligen Schüleranteilen.

- Auf europäischer Ebene ist besonders die Europäische Wachstums- und Beschäftigungsstrategie „Europa 2020“ als Nachfolger der Lissabon-Strategie von 2000 bis 2010 zu nennen. Hinzu kommen weitere Indikatorensysteme für z. B. ausgewählte Leitinitiativen wie „Digitale Agenda“ oder Bildung⁴⁸. Überwiegend richten sich diese politikbegleitenden Indikatorensysteme für Deutschland an die nationale Ebene. Eine Ausnahme für eine engere Zusammenarbeit auf regionaler Ebene bildet die sogenannte Territoriale Agenda der Europäischen Union für die Lissabon-Strategie: Sie ist in der sogenannten Leipzig-Charta aus dem Jahr 2007 enthalten. Für „Europa 2020“ wurde sie als „Territoriale Agenda der Europäischen Union 2020“ vom 19. Mai 2011 in Gödöllő verabschiedet.⁴⁹ Ein Beispiel für die Europäische Raumentwicklung enthält Abbildung i für zwei verschiedene Zeiträume.
- Seit ca. 10 Jahren arbeitet die OECD intensiv und systematisch an einem wissenschaftlich fundierten⁵⁰ statistischen Indikatorensystem zur Fortschrittsmessung „Progress measurement“. Ein Bericht zu „How's life?“ wurde bereits zum zweiten Mal in einer aktualisierten Fassung 2013 veröffentlicht.⁵¹ Neben der Weiterentwicklung eines solchen Ansatzes für die nationale Ebene wurden die regionalen Implikationen auf OECD-Ebene bereits differenzierter diskutiert.⁵² Eine Veröffentlichung „How is life in your

region?“ ist für Oktober 2014 geplant. Für Deutschland wurde das Konzept der OECD für „How's life?“ für ausgewählte große und kleine kreisfreie Städte in Deutschland anhand ausgewählter regionalstatistischer Daten überwiegend aus der Regionaldatenbank Deutschland untersucht.⁵³

- Auf UN-Ebene steht derzeit die Überarbeitung der sogenannten Millennium Development-Ziele im sogenannten „post 2015“-Prozess an.⁵⁴ Auch diese politische Strategie soll von einem statistischen Monitoringsystem begleitet werden; eine internationale Arbeitsgruppe – die sogenannte Friends of the Chair Group (FOC) – mit hochrangigen Vertreterinnen und Vertretern aus der amtlichen Statistik, u. a. vom Statistischen Bundesamt⁵⁵, unterstützt die politischen Entscheidungsträger hinsichtlich der möglichen statistischen Indikatoren und deren Implikationen. Derzeit werden 29 Themenbereiche mit rund 100 Indikatoren diskutiert.⁵⁶ Da die „post 2015“-Strategie für alle Mitgliedstaaten der Vereinten Nationen gelten soll – unabhängig von ihrem wirtschaftlichen und gesellschaftlichen Entwicklungsstand sowie subnational empfohlen wird, werden allgemein wichtige Impulse für die Regionalstatistik weltweit erwartet.⁵⁷

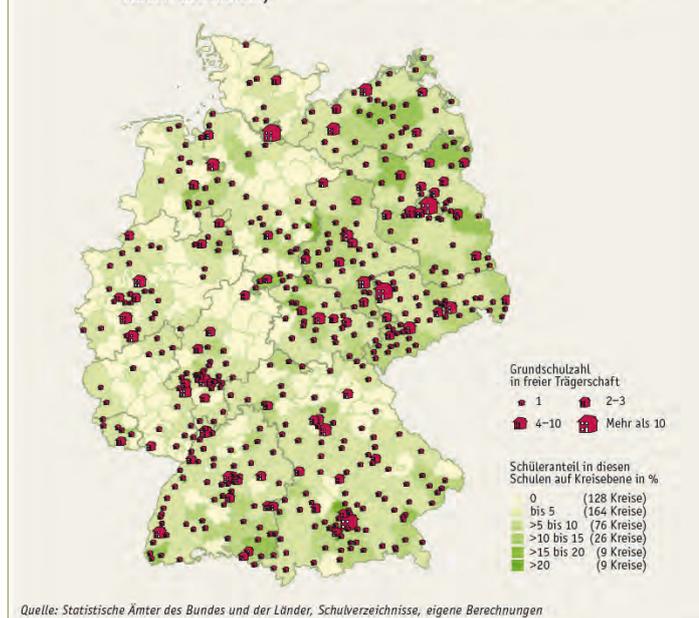
ad (2) Europäische Kohäsions- und Regionalpolitik

Ein wichtiges Ziel für die Europäische Union ist die Stärkung des wirtschaftlichen, sozialen und territorialen Zusammenhalts.⁵⁸ Ein Instrument dafür ist die europäische Regionalpolitik, die auf allen Ebenen (Gemeinschaftsebene, nationale Ebene, Europas Regionen und Städte) dazu beiträgt, eine harmonische Gesamtentwicklung in den Mitgliedstaaten und ihren Regionen zu fördern und zu unterstützen.⁵⁹ Die europäische Regionalpolitik ist Teil der sogenannten Struktur- und Kohäsionspolitik und wird für einen

h | Politikbegleitende regionalstatistische Daten am Beispiel des Bildungsmonitoring

Quelle: Autorengruppe Bildungsberichterstattung (Hrsg.): Bildung in Deutschland 2012, Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf, Bielefeld, 2012, hier S. 71.

Abb. D1-3: Grundschulen in freier Trägerschaft 2010 (Anzahl der Schulen je Gemeinde) und Schüleranteil in diesen Schulen (in % aller Grundschülerinnen und -schüler des Kreises)



Sieben-Jahreszeitraum – derzeit 2014 bis 2020 – festgelegt. Sie fördert durch eine Fülle von Maßnahmen vor allem Länder und Regionen, die in ihrer Entwicklung zurückliegen, mit dem Ziel, die wirtschaftlichen und sozialen und territorialen Unterschiede zu verringern.

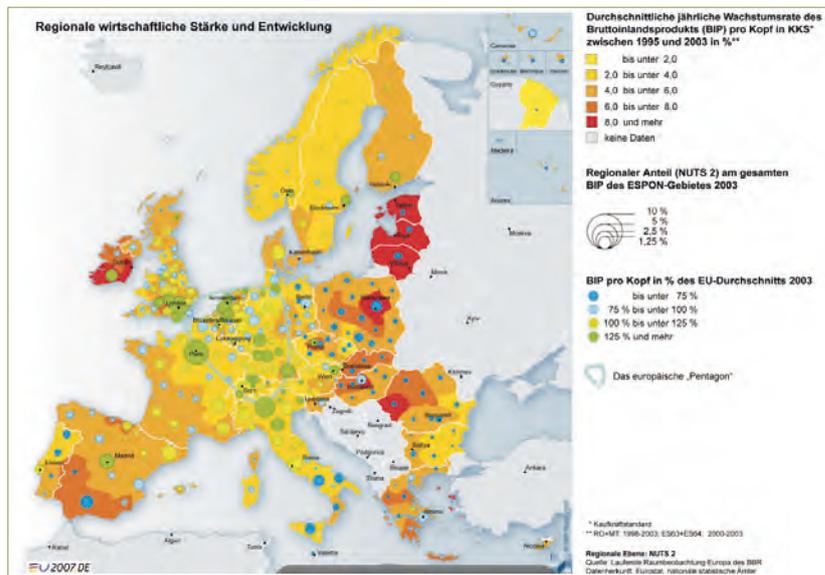
Ziele und Maßnahmen der jeweiligen Förderperiode werden in entsprechenden Rechtsakten festgelegt und anhand weiterer statistischer Indikatoren spezifiziert.⁶⁰ So wurden in der Förderperiode 2007 bis 2013 die zu fördernden Regionen bezüglich des Ziels „Konvergenz“ derart abgegrenzt, dass nur Regionen gefördert wurden, in denen das Bruttoinlands-

produkt pro Einwohner weniger als 75% des europäischen Durchschnitts betrug.

Die aktuelle, reformierte Kohäsionspolitik für den Zeitraum 2014 bis 2020 mit insgesamt 351,8 Mrd. EUR für Investitionen in Regionen und Städte für die Europäische Union dient vor allem zur Verwirklichung der Ziele von „Europa 2020“. Hierbei geht es um die Schaffung von Wachstum und Arbeitsplätzen, die Bekämpfung des Klimawandels und der Energieabhängigkeit sowie der Verringerung von Armut und sozialer Ausgrenzung.⁶¹

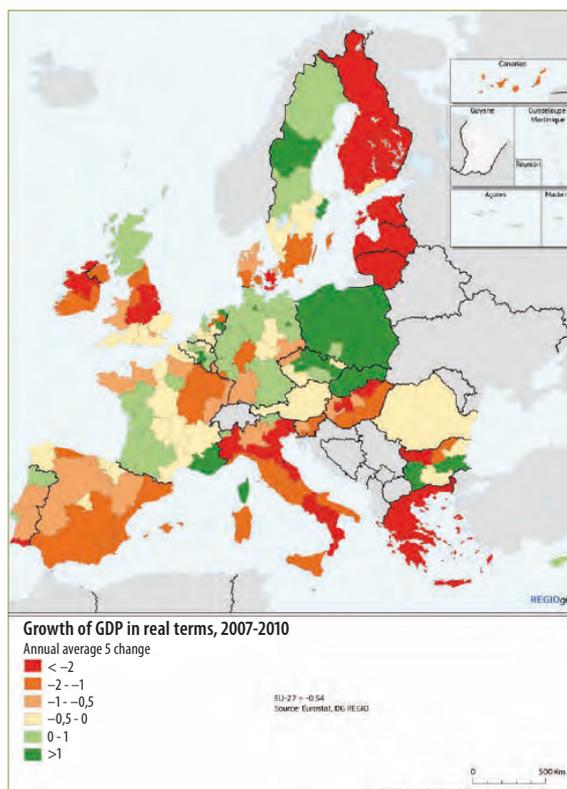
Die europäische Förderpolitik (siehe Abbildung j a) hat auch Auswirkungen auf die regionale Wirt-

i a | Regionale wirtschaftliche Stärke und Entwicklung 1995 bis 2003



Quelle:
Bundesministerium für Verkehr,
Bau und Stadtentwicklung
(Hrsg.): Karten zur
Europäischen Raumentwick-
lung, Leipzig Mai 2007, hier S. 9
bzw. http://www.bmvi.de/SharedDocs/DE/Anlage/StadtUndLand/Laendlicher-Raum/kartensammlung-zur-europaeischen-raumentwicklung.pdf?__blob=publicationFile

i b | Regionale wirtschaftliche Stärke und Entwicklung 2007 bis 2010



Quelle: Europäische Kommission:
Die regionale und urbane
Dimension der Krise. Achter
Zwischenbericht über den
wirtschaftlichen, sozialen und
territorialen Zusammenhalt,
Brüssel Juni 2013, Annex, S. 33.

schaftspolitik der Bundesregierung „Verbesserung der regionalen Wirtschaftsstruktur“ (GRW), insofern, als beispielweise für die Abgrenzung der Fördergebiete auch die Leitlinien für Regionalbeihilfen 2014 bis 2020 der Europäischen Union bis auf eine Ausnahme Berücksichtigung finden (siehe dazu auch Abbildung j b).

Die europäische Struktur- und Kohäsionspolitik mit ihren verschiedenen Komponenten und Fonds, besonders dem Europäischen Fonds für die regionale Entwicklung (EFRE) und dem Europäischen Sozialfonds (ESF), hat wesentlich zur Entwicklung des in Abschnitt 3 näher beschriebenen regionalstatistischen Programms für Europa beigetragen. Grundlage ist die europäische Gebietsklassifikation NUTS, die zur Abgrenzung der Fördergebiete verwendet wird: Sie erfolgt vor allem für Gebietseinheiten der Ebene NUTS 2 (für Deutschland Regierungsbezirke und Statistische Regionen).

Die europäische Förderpolitik wird in regelmäßigen Abständen überprüft und in einem Kohäsionsbericht überwiegend anhand von statistischen Indikatoren kritisch bezüglich ihrer Zielwirkungen analysiert und dokumentiert⁶². Eine umfassende

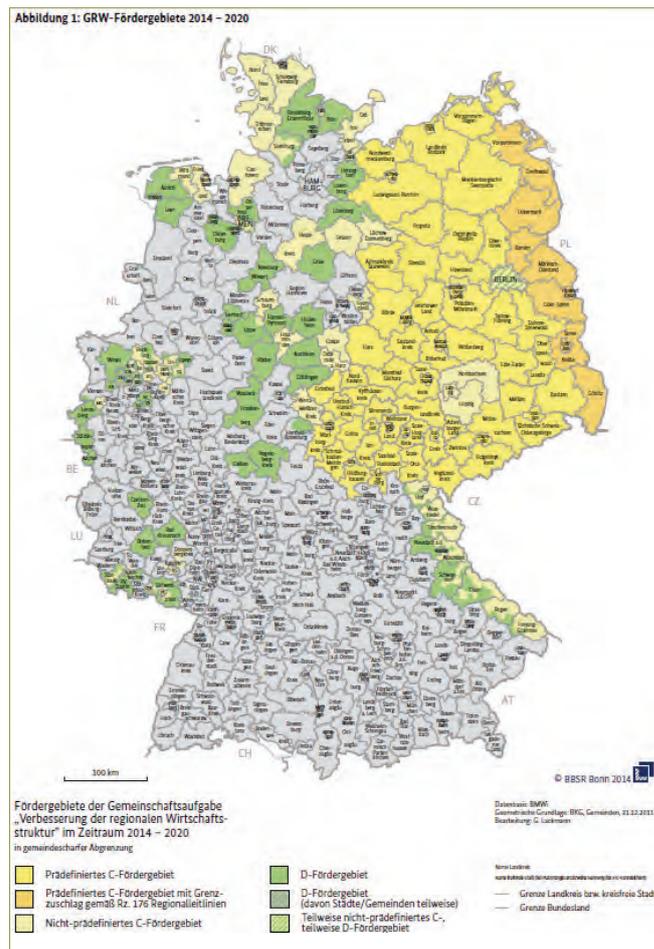
Darstellung enthält der sogenannte Kohäsionsbericht (zuletzt 2010)⁶³, der jeweils um sogenannte Interimsreporte ergänzt wird (zuletzt 2013)⁶⁴.

Eine weitere wichtige Informationsquelle für gemeinschaftsweite europäische Politiken ist die Europäische Erhebung Urban Audit mit derzeit insgesamt 885 Großstädten.⁶⁵ Für diese Erhebung werden seit 1999 statistische Daten für Deutschland durch die KOSIS-Gemeinschaft „Urban Audit“ u.a. mithilfe von EU-Fördermitteln bereitgestellt und über das Statistische Bundesamt vertraglich und finanziell abgewickelt. Diese Daten werden überwiegend von den in die Erhebung einbezogenen deutschen Städten bereitgestellt.⁶⁶ Aber zunehmend findet auch das regionalstatistische Datenangebot der Statistischen Ämter des Bundes und der Länder, besonders aus der Regionaldatenbank Deutschland, Verwendung. Für Deutschland sind derzeit 125 Städte mit mehr als 50 000 Einwohnern einbezogen, wobei für die Abgrenzung auch funktionsräumliche Kriterien (wie Oberzentrum für die umgebende Region) herangezogen wurden. Diese Erhebung ermöglicht Städtevergleiche in siedlungsstruktureller und funktionsräumlicher Hinsicht.

j a | Abgrenzung der Fördergebiete für den Zeitraum 2014 bis 2020



j b | Abgrenzung der GRW-Fördergebiete auf Gemeindeebene für den Zeitraum 2014 bis 2020



Europäische Kommission: EU-Kohäsionspolitik in Deutschland, http://ec.europa.eu/regional_policy/information/cohesion-policy-achievement-and-future-investment/factsheet/germany_de.pdf

Quelle: <http://www.bmwi.de/Dateien/BMWi/PDF/Monatsbericht/schlaglichter-der-wirtschaftspolitik-05-2014.property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf>, S. 27.

Das Statistische Bundesamt hat 2013 einen Flyer⁶⁷ veröffentlicht, in dem fünf deutsche Großstädte mit deren englischen und französischen Partnerstädten anhand ausgewählter Urban Audit-Indikatoren näher betrachtet werden. Dazu gehören beispielsweise die Größe der Städte gemessen an den Einwohnerzahlen und unter anderem das Wohneigentum. Dabei zeigt sich z.B., dass unabhängig von der Einwohnerzahl einer Stadt die Haushalte in den ausgewählten Städten in Deutschland weniger in eigenen Wohnungen leben als in ihren europäischen Partnerstädten (Abbildung k).

Für Urban Audit kommen die in Abschnitt 3 vorgestellten, überwiegend von der Europäischen Kommission DG Regio in enger Zusammenarbeit mit Eurostat und der für Regionalpolitik zuständigen Einheit der OECD entwickelten, Gebietstypisierungen zur Anwendung. Eine weitere Harmonisierung, vor allem bezüglich des regionalstatistischen Merkmalsangebots und dessen Integration in das europäische regionalstatistische Programm, wird angestrebt.

ad (3) Umweltpolitiken

Ein wichtiges Instrument für die europäische Umweltpolitik sind sogenannte Geodateninfrastrukturen. Ein Beispiel dafür ist INSPIRE, eine europäische Geodateninfrastruktur, die auf der INSPIRE-Richtlinie vom 15. Mai 2007 basiert.⁶⁸ Zweck dieser Richtlinie ist es, eine raumbezogene Dateninfrastruktur für eine gemeinschaftsweite und integrative Umweltpolitik aufzubauen⁶⁹: Alle Einrichtungen des öffentlichen Bereichs sollen über diese Geodateninfrastruktur ihre relevanten Daten nach einheitlichen technischen Standards bezüglich Download, View, Metadaten etc. für alle Bereiche von Politik, Gesellschaft, Wirtschaft und Wissenschaft interoperabel, d.h. möglichst überschneidungsfrei und leicht kombinierbar, bereitstellen.^{70,71}

Für Deutschland wurde die INSPIRE-Richtlinie durch die sogenannten Geodatenzugangsgesetze auf Bundes- und Länderebene umgesetzt.⁷² Mit dem Geoportal Deutschland wurde eine INSPIRE-konfor-

me Geodateninfrastruktur nach einheitlichen Prinzipien für alle föderalen Ebenen⁷³ geschaffen.

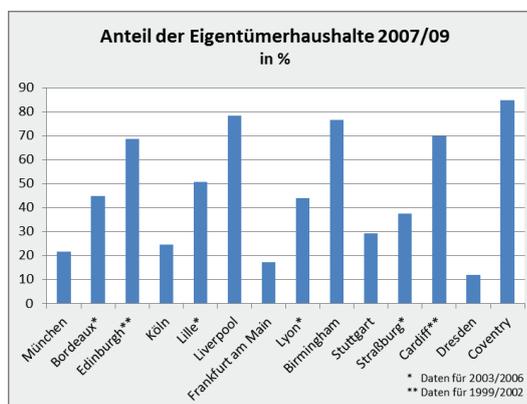
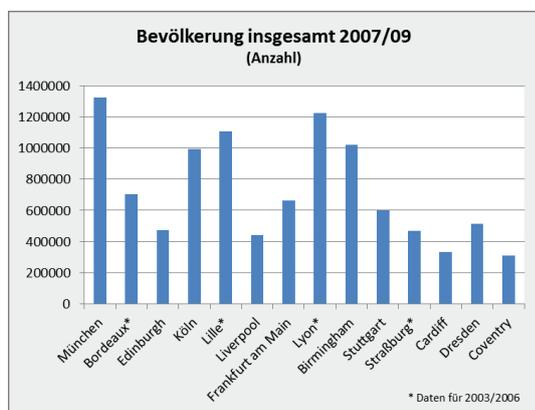
Nach Inkrafttreten der nationalen Rechtsakte zu INSPIRE haben die Statistischen Ämter des Bundes und der Länder unter Federführung des Statistischen Bundesamtes ein High-Level-Gremium eingerichtet, in dem wichtige Aspekte zur Relevanz von INSPIRE für die Bundesstatistik diskutiert und geklärt wurden. Auch hat sich das Statistische Bundesamt intensiv in den INSPIRE-Prozess auf europäischer Ebene eingebracht (u.a. durch Mitarbeit in einer thematischen Arbeitsgruppe wie auch durch die Beteiligung in offenen „Review and Testing“ im Sommer 2011) (Abbildung l).

Dabei wurden u.a. die fachlichen Themenbereiche vor allem von Anhang 3 der INSPIRE Richtlinie daraufhin untersucht, welche Daten aus dem bundesstatistischen Programm kleinräumig zur Verfügung gestellt werden können. Von den 21 im Anhang 3 genannten Themenbereichen sind etwa für die Hälfte prinzipiell Daten aus der Bundesstatistik verfügbar (siehe Abbildung m), wobei das Datenangebot in Tiefe und Breite bereichsspezifisch schwankt.

So ist die Bundesstatistik einer der Hauptdatenlieferanten für den Bereich „Bevölkerungsverteilung - Demografie/Population Distribution“, aber auch im Bereich „Gesundheit und soziale Sicherheit/Human Health and Safety“ oder „Landnutzung/Land Use“ gibt es eine Fülle von Daten.

Die Statistischen Ämter des Bundes und der Länder haben beschlossen, dass für INSPIRE bzw. die Geodaten-Infrastruktur Deutschland bzw. das Geoportal Deutschland kleinräumige Daten aus der Regionaldatenbank Deutschland⁷⁴ bereitgestellt werden. Diese Datenbank – ein Gemeinschaftsprodukt der Statistischen Ämter des Bundes und der Länder – enthält eine Fülle von deutschlandweit vergleichbaren statistischen Daten zu verschiedenen Statistikbereichen auf Kreisebene.⁷⁵ Sie beinhaltet mehr als 20 Mill. Werte für regionalstatistische Daten für die verschiedenen föderalen Ebenen: insgesamt etwa 110 Tabellen für 24 Bereiche aus 80 Statistiken auf Landes-, Regierungsbezirks- bzw.

k | Vergleich ausgewählter deutscher Städte mit ihren europäischen Partnerstädten



Quelle: Statistisches Bundesamt (Hrsg.): Lebens- und Arbeitsbedingungen in fünf deutschen Großstädten und ihren Partnerstädten in Frankreich und England – Ausgewählte Indikatoren, Wiesbaden 2013.

2 | Ausgewählte harmonisierte subnationale Datenangebote des Statistischen Bundesamtes sowie Gemeinschaftsveröffentlichungen der Statistischen Ämter des Bundes und der Länder kleinräumig, kartografisch sowie politikbegleitend

Datenquelle	Regionaleinheit	Karten	Aktualisierung	Herausgeber	Kurzbeschreibung	Beispiellink	
Datenbanken	GENESIS-Online	Deutschland, Bundesländer	ü	unterjährig	Destatis	> 500 Mill. Werte für > 200 Statistiken	https://www-genesis.destatis.de/genesis/online
	Regionaldatenbank	Deutschland, Bundesländer, Regierungsbezirke, Kreise/kreisfreie Städte, Gemeinden, -verbände	–	unterjährig	Gemeinschaftsveröffentlichung	24 Themenbereiche mit 113 Tabellen, davon 29 auf Gemeindeebene	https://www.regionalstatistik.de/genesis/online/logon
	Gemeindeverzeichnis-informations-system GV-ISys)	Deutschland, Bundesländer, Gemeinden, -verbände	–	unterjährig	Gemeinschaftsveröffentlichung	für > 11 100 Gemeinden, >1 200 Gemeindeverbände Angaben zur Bevölkerung, Fläche, Raumtypisierungen etc.	https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis_ol.html
Atlanten	Regionalatlas	Kreise/kreisfreie Städte	ü	unterjährig	Gemeinschaftsveröffentlichung	über 90 Indikatoren und rd. 600 Karten, Regionaldatenbank	https://www-genesis.destatis.de/gis/genView?GenMLURL=https://www-genesis.destatis.de/regatlas/ai002-1.xml&CONTEXT=REGATLAS01
	Außenhandelsatlas	Deutschland, Bundesländer	ü	jährlich	Destatis	Gesamtexport/-import und Handelsbilanzsaldo nach Warenklassifikationen	https://www-genesis.destatis.de/gis/genView?GenMLURL=https://www-genesis.destatis.de/ahAtlas/Atlas-T3.xml&CONTEXT=FOREIGN-TRADE01
	Luftverkehrs-atlas	Bund	ü	jährlich	Destatis	Passagier-, Luftfracht- und Luftpostaufkommen	https://www-genesis.destatis.de/gis/genView?GenMLURL=https://www-genesis.destatis.de/lvAtlas/LV-Atlas-03.xml&CONTEXT=AIR-TRANSPORT01&CLASSLIM=10000,100000,500000,1000000,5000000
	AgrarAtlas	Rasterzellen	ü	einmalig	Gemeinschaftsveröffentlichung	16 Karten zu Landwirtschaft, landwirtschaftl. Betrieben und Flächen, Tierbeständen etc.	http://www.atlas-agrarstatistik.nrw.de/
	Zensus 2011 – So wohnt Deutschland	Kreise/kreisfreie Städte	ü	einmalig	Destatis	Karten zu Neubauten, Leerstandsquoten, Eigentümerquoten, Wohnungsgrößen etc.	https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/Wohnen/zensusKarteTeaser.html
	Bundestagswahl	Wahlkreise	ü	4-jährlich	Bundeswahlleiter	Karten zu ausgewählten Strukturdaten und Wahlergebnissen	http://www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_09/onlineatlas/start_wahlatlas.html
	Europawahl	Deutschland, Bundesländer, Kreise/kreisfreie Städte	ü	5-jährlich	Bundeswahlleiter	Karten zu ausgewählten Strukturdaten und Wahlergebnissen	http://www.bundeswahlleiter.de/de/europawahlen/EU_BUND_09/onlineatlas/start_wahlatlas.html
	Genderatlas/ Atlas zur Gleichstellung von Männern und Frauen	Bundesländer, Kreise/kreisfreie Städte	tw	3-jährlich	BMFSFJ	Frauen-/Männeranteil zu Partizipation, Bildung, Ausbildung, Arbeit, Einkommen, Lebenswelt; teilweise im Regionalatlas	http://www.bmfsfj.de/BMFSFJ/Service/Publikationen/publikationen,-did=130048.html
Politikbegleitendes Monitoring, fach-spezifische Datenbanken	Nachhaltige Entwicklung in Deutschland	Deutschland	tw	fortlaufend/ regelmäßig	Destatis, Gemeinschaftsveröffentlichung	52 Indikatoren in GENESIS-Online; 16 Indikatoren im Regionalatlas	https://www.destatis.de/DE/ZahlenFakten/Indikatoren/Nachhaltigkeitsindikatoren/Nachhaltigkeitsindikatoren.html
	Sozialbericht-erstattung	Deutschland, Bundesländer, Regierungsbezirke, Kreise/kreisfreie Städte	tw	fortlaufend/ regelmäßig	Destatis, Gemeinschaftsveröffentlichung	Indikatoren zu Einkommens-/ -verteilung, Mindestsicherungsleistung, Qualifikationsniveau, Erwerbsbeteiligung; 16 Indikatoren im Regionalatlas	http://www.amtliche-sozialberichterstattung.de/armutsgefahrungsquoten_regional.html https://www.regionalstatistik.de/genesis/online;jsessionid=A54206AE99A68BEE961766235EE81CE7?sequenz=statistikTabellen&selectionname=22810
	Gesundheitsbericht-erstattung des Bundes/GBE	Deutschland, Bundesländer	tw	fortlaufend/ regelmäßig	Destatis, Gemeinschaftsveröffentlichung	> 2 Mrd. Zahlen und Kennziffern zu Gesundheitsdaten und -informationen, 5 Indikatoren im Regionalatlas	http://www.gbe-bund.de/
	Integration	Deutschland, Bundesländer	–	2-jährlich	Destatis	für Regionalatlas NUTS 2 geplant	http://www.berlin.de/imperia/md/content/lb-integration-migration/publikationen/berichte/bericht_2011_lag_indikatorenentwicklung_monitoring_teil_1_bericht.pdf?start&ts=1363878767&file=bericht_2011_lag_indikatorenentwicklung_monitoring_teil_1_bericht.pdf https://www.statistik-berlin-brandenburg.de/home/pdf/SP_Integrationsbericht-01-00_DE_2011_DT.pdf
	Bildungsbericht-erstattung	Deutschland, Bundesländer, Kreise/kreisfreie Städte	tw	2-jährlich	Gemeinschaftsveröffentlichung	datengestützte Informationen über Rahmenbedingungen, Input, Verläufe, Ergebnisse und Wirkungen von Bildungsprozessen	https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Bildungsstand/BildungDeutschland.html https://www.bildungsmonitoring.de/bildung/online/logon
Bildung und Wissenschaft	Bundesländer	ü	zuletzt 2009	Destatis	GENESIS-Online, 7 Indikatoren im Regionalatlas	https://www-genesis.destatis.de/genesis/online/data;jsessionid=F4115B54719FC6D5E0DC9F1BD95F5DF4.tomcat_GO_1_1?operation=statistikenVerzeichnisNextStep&levelindex=0&levelid=1401179487726&index=7&structurelevel=2	

ü= überwiegend, tw= teilweise Quelle: Eigene Darstellung

Regionsebenen, davon 29 Tabellen auf Gemeindeebene. Ein Teil dieser Daten und Indikatoren wird im Regionalatlas – ebenfalls ein Gemeinschaftsprodukt – interaktiv und kartografisch online allgemein zugänglich und kostenfrei angeboten. Dieser Atlas enthält zurzeit über 90 Indikatoren mit rd. 600 Karten; Datenangebote zu Indikatorensystemen, wie sie in der Datenbank des Statistischen Bundesamtes GENESIS-Online bereitgestellt werden (z.B. für die Bereiche nachhaltige Entwicklung in Deutschland, Wissenschafts-, Technologie- und Kulturindikatoren) sind in Vorbereitung.

Aus der Regionaldatenbank Deutschland wurden zunächst nur Daten zur Bevölkerung im Rahmen des sogenannten Monitoringprozesses von den statistischen Ämtern für INSPIRE gemeldet. Da für eine Geodateninfrastruktur wie INSPIRE vor allem statistische Angaben für vielfältige Zwecke benötigt werden, die z.B. eine Unterscheidung nach Tag- und Nachtbevölkerung erforderlich macht, wurden dazu bereits wichtige Komponenten aus der Bundesstatistik zu Wohnbevölkerung, zu Beschäftigten am Arbeitsort, zu Schülerinnen und Schülern etc. bereitgestellt.

In den nationalen (Geo-)Portalen (Geoportal Deutschland oder GovData) sind darüber hinaus weitere kleinräumige Datenangebote von den Statistischen Ämtern des Bundes und der Länder eingebunden. Dazu gehören z.B. der rasterbasierte Agraratlas oder Atlanten zu Bundestags- oder Europawahlen, ebenfalls Produkte des Statistischen Bundesamtes. Im Weiteren ist über diese Datenportale eine Vielzahl an ausgewählten bereichsspezifischen, auch kleinräumigen Datenangeboten aus der amtlichen Statistik zugänglich. Einen Überblick über deutschlandweit harmonisierte Daten vom Statistischen Bundesamt wie auch über Gemeinschaftsveröffentlichungen der Statistischen Ämter des Bundes und der Länder gibt Tabelle 2.

5. Ausblick

Auch wenn bislang auf Bundesebene ein unmittelbarer Bedarf an kleinräumigen Daten für politische Entscheidungen nur in ausgewählten Bereichen besteht, so ist doch ein zunehmendes allgemeines Interesse an flächendeckenden kleinräumigen Daten zu beobachten. Dieses Interesse resultiert zum einen aus der Erkenntnis, dass politische Vorgaben für die nationale, supra- oder internationale Ebene umso erfolgreicher sind, je mehr den regionalen Besonderheiten Beachtung geschenkt wird bzw. je mehr es auch in irgendeiner Weise benachteiligten Regionen ermöglicht wird, entsprechende Barrieren zu beseitigen und ihr Potenzial voll zu entfalten. Das bedeutet, auch von Regionen oder Städten wird zunehmend der Wunsch nach vergleichbaren flächendeckenden kleinräumigen Daten für Vergleichszwecke geäußert. Die Statistischen Ämter des Bundes und der Länder stellen bereits heute schon eine Fülle von weitgehend harmonisierten regionalstatistischen Daten für kleinräumige Analysen bereit. Der Ausbau dieses Angebots schreitet zügig fort.

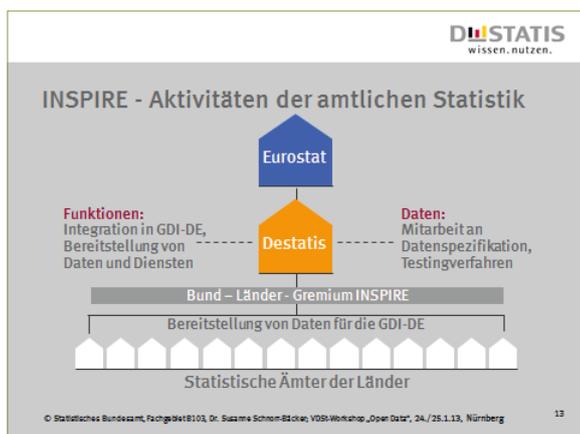
Dr. Susanne Schnorr-Bäcker

leitet im Statistischen Bundesamt (Destatis) das Referat *Koordinierung, Regionalstatistik, Indikatoren*. Sie ist Mitglied in verschiedenen Arbeitsgruppen auf nationaler und subnationaler Ebene und arbeitet in zahlreichen Gremien bei der Europäischen Kommission – besonders bei Eurostat –, als auch der OECD und den Vereinten Nationen mit.



Ihre aktuellen Arbeitsschwerpunkte sind „Neue Datenquellen/Big Data“, vor allem zusammen mit Eurostat, die „Better Life Initiative“ und „How is life in your region“ der OECD sowie die Bedeutung von Geodateninfrastrukturen (wie GDI-DE, INSPIRE und UN-GGIM) für das Programm der Bundesstatistik.

I | Mitarbeit der Statistischen Ämter des Bundes und der Länder bei INSPIRE



Quelle: Schnorr-Bäcker, S.: Weiterentwicklungen in der amtlichen Regionalstatistik, Vortrag und Präsentation auf dem VDSt-Workshop, 24./25. Januar 2013 in Nürnberg, Folie 13.

m | INSPIRE-Themenbereiche mit Bezug zur amtlichen Statistik

Themenbereiche	groß	mittel	gering
Agricultural and Aquaculture Facil. (AF)			X
Area Management (AM)			X
Buildings (BU)			X
Energy Resources (ER)			X
Human Health and Safety (HH)		X	
Land Cover (LC)			X
Land Use (LU)		X	
Population Distribution (PD)	X		
Production and Industrial Facilities (PF)			X
Statistical Units (SU)	X		
Utility and Governmental Services (US)			X

Quelle: Schnorr-Bäcker, S.: Weiterentwicklungen in der amtlichen Regionalstatistik, Vortrag und Präsentation auf dem VDSt-Workshop, 24./25. Januar 2013 in Nürnberg, Folie 15.

Anmerkungen

- 1 Siehe Regelungen im Umweltstatistikgesetz (UStatG) vom 16. August 2005 (BGBl. I S. 2446) oder im Straßenverkehrsunfallstatistikgesetz (STVUnfStatG) vom 15. Juni 1990 (BGBl. I S. 1078).
- 2 Siehe AgrarstatG, <http://www.atlas-agrarstatistik.nrw.de/> sowie Agrarstatistikgesetz (AgrStatG) in der Fassung der Bekanntmachung vom 17. Dezember 2009 (BGBl. I S. 3886), das zuletzt durch Artikel 13 Absatz 5 des Gesetzes vom 12. April 2012 (BGBl. I S. 579) geändert worden ist, <http://www.gesetze-im-internet.de/agrstatg/BJNR004690989.html>
- 3 Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz – E-GovG) vom 25. Juli 2013 (BGBl. I S. 2749), http://www.bmi.bund.de/DE/Themen/IT-Netzpolitik/E-Government/E-Government-Gesetz/e-government-gesetz_node.html
- 4 Zum genauen Wortlaut siehe die entsprechenden §§ im Bundesstatistikgesetz vom 22. Januar 1987 (BGBl. I S. 462, 565), das zuletzt durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749) geändert worden ist.
- 5 Der genaue Wortlaut von § 14 E-GovG ist: „(1) Wird ein elektronisches Register, welches Angaben mit Bezug zu inländischen Grundstücken enthält, neu aufgebaut oder überarbeitet, hat die Behörde in das Register eine bundesweit einheitlich festgelegte direkte Georeferenzierung (Koordinate) zu dem jeweiligen Flurstück, dem Gebäude oder zu einem in einer Rechtsvorschrift definierten Gebiet aufzunehmen, auf welches sich die Angaben beziehen. (2) Register im Sinne dieses Gesetzes sind solche, für die Daten auf Grund von Rechtsvorschriften des Bundes erhoben oder gespeichert werden; dies können öffentliche und nichtöffentliche Register sein.“
- 6 Siehe dazu auch die Ausführungen in European Commission: INSPIRE Infrastructure for Spatial Information in Europe, D2.8.III.1 v3.Orç3; 2013-02-04, S.1ff. http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SU_v3.Orç3.pdf
- 7 Siehe dazu die Ausführungen im Straßenverkehrsunfallstatistikgesetz, a. a. O. oder im Umweltstatistikgesetz a. a. O.
- 8 Siehe dazu auch die Ausführungen in Abschnitt 2.
- 9 Im Folgenden soll nicht weiter eingegangen werden auf geo- oder topografische Besonderheiten.
- 10 Siehe dazu z. B. die Publikation des Landesamtes für Statistik Niedersachsen: Gebietsstandsänderungen in Niedersachsen – ab 1978, vom 11. Februar 2014 unter http://www.statistik.niedersachsen.de/portal/live.php?navigation_id=33809&article_id=103365&psmand=40
- 11 Siehe dazu Statistische Ämter des Bundes und der Länder (Hrsg.): Gemeindeverzeichnis-Informationssystem (GV-ISys), Wiesbaden, Oktober 2013 bzw. <https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis/Gemeindeverzeichnis.html?sessionid=5327CA7B81C7788BB5F52F40A453B44E.cae1>
- 12 Siehe dazu auch die Regelungen im Raumordnungsgesetz (ROG) vom 22. Dezember 2008 (BGBl. I S. 2986), zuletzt geändert durch Artikel 9 des Gesetzes vom 31. Juli 2009 (BGBl. I S. 2585).
- 13 Auf einer derartigen Raumgliederung baut auch ein Simulationsansatz zur Versorgung der Bevölkerung auf; siehe auch Schnorr-Bäcker, S.: Ein Simulationsansatz zur Beurteilung der Güterversorgung der Bevölkerung unter räumlichem Aspekt, Diss., Gießen 1983.
- 14 Siehe dazu auch Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR), <http://www.bbsr.bund.de/BBSR/DE/Raumentwicklung/RaumentwicklungDeutschland/Da-seinsvorsorge/Projekte/ZentraleOrte.html>
- 15 Verordnung (EG) Nr. 1059/2003 des Europäischen Parlaments und des Rates vom 26. Mai 2003 über die Schaffung einer gemeinsamen Klassifikation der Gebietsseinheiten in der Statistik (NUTS) (ABl. L 154 vom 21. Juni 2003). Derzeit gilt die Verordnung (EU) Nr. 31/2011 der Kommission vom 17. Januar 2011, (ABl. L 13 vom 18. Januar 2011, S. 3).
- 16 Nach dem Gesetzestext besteht auch die Möglichkeit einer außerplanmäßigen Revision, wenn bestimmte nationale Interessen zu einer grundlegenden Änderung der regionalen Untergliederung eines Landes führen; als ein Beispiel sei die Aufnahme von Kroatien in die Europäische Union zum 1. Juli 2013 genannt.
- 17 Ab 1. Januar 2015 gilt im Hinblick auf die Übermittlung von Daten an die Europäische Kommission die Verordnung (EU) Nr. 1319/2013 (ABl. L342/1 vom 18. Dezember 2013) der Kommission vom 9. Dezember 2013 zur Änderung der Anhänge der Verordnung (EG) Nr. 1059/2003.
- 18 Zu Einzelheiten dieses Geostat-Projektes siehe European Forum for Geography and Statistics (Geostat/EFGS), <http://www.efgs.info/>
- 19 Ähnlich auch http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/documents/Defining_urban_areas.pdf
- 20 Siehe dazu besonders die Ausführungen http://epp.eurostat.ec.europa.eu/portal/page/portal/degree_urbanisation/introduction
- 21 Siehe dazu https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis_ol.html
- 22 Siehe dazu auch die Veröffentlichung von Dijkstra, L., Poelman, H.: Cities in Europe, the new OECD-EC Definition, herausgegeben von der Europäischen Kommission, Brussels 1/2012 bzw. http://ec.europa.eu/regional_policy/sources/docgener/focus/2012_01_city.pdf
- 23 Siehe dazu auch http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/city_urban
- 24 Siehe dazu http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/city_urban/spatial_units
- 25 Siehe <http://www.staettestatistik.de/urban-audit.html>
- 26 Siehe dazu auch Europäische Kommission: Cities in Europe, a.a.O., S. 6 ff.
- 27 Siehe dazu auch die Veröffentlichung der Europäischen Kommission: European Union, Regional Policy: Cities of tomorrow, Challenges, visions, ways forward, Brussels, October 2011 bzw. http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/citiesoftomorrow/citiesoftomorrow_final.pdf
- 28 Für Japan und die USA werden von der OECD andere Werte gewählt, wie z. B. eine Siedlungsdichte von 1 000 Einwohnern je km² für Kanada und die Vereinigten Staaten für Städte ab 50 000 Einwohnern; für Japan, Korea, Mexiko werden erst Städte mit einer Einwohnerzahl von mehr als 100 000 Personen betrachtet. Siehe dazu auch die Definitionen der OECD für funktionale Urban areas, S. 3, www.oecd.org/gov/regional-policy/definition-of-Functional-Urban-Areas-for-the-OECD-metropolitan-database.pdf
- 29 Siehe dazu http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/metropolitan_regions wie auch European Union: Metropolitan regions in the EU, in Regional Focus no. 1/2009 bzw. http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/metropolitan_regions
- 30 Siehe dazu auch die Ausführungen in RatSWD: Endbericht der AG „Georeferenzierung von Daten“ des RatSWD, Bericht der Arbeitsgruppe und Empfehlungen des Rates für Sozial- und Wirtschaftsdaten (RatSWD), Berlin 2012, insbesondere die Abschnitte 3, S. 10 ff. sowie 5.2.1., S. 37 ff.; http://ratswd.de/Geodaten/downloads/RatSWD_Endbericht_Geo-AG.pdf
- 31 Für Deutschland ist auf Bundesebene für diese Initiative das Bundesministerium des Innern zuständig; siehe dazu http://www.bmi.bund.de/DE/Themen/Moderne-Verwaltung/Open-Government/open-government_node.html
- 32 Siehe dazu auch <http://ec.europa.eu/digital-agenda/en/open-government>
- 33 Siehe dazu auch United Nations (Hrsg.): Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda, New York 2013 bzw. <http://www.post2015high.org/wp-content/uploads/2013/05/UN-Report.pdf>, besonders Executive Summary, Ziffer 4 und Kapitel 1, S. 4 ff.
- 34 Informationsfreiheitsgesetz vom 5. September 2005 (BGBl. I S. 2722), das durch Artikel 2 Absatz 6 des Gesetzes vom 7. August 2013 (BGBl. I S. 3154) geändert worden ist.
- 35 Gesetz über die Weiterverwendung von Informationen öffentlicher Stellen (Informationsweiterverwendungsgesetz – IWG) vom 13. Dezember 2006 (BGBl. I S. 2913); dieses Gesetz dient der Umsetzung der Richtlinie 2003/98/EG des Europäischen Parlaments und des Rates vom 17. November 2003 über die Weiterverwendung von Informationen des öffentlichen Sektors (ABl. EU Nr. L 345 S. 90).
- 36 Siehe Artikel 11 Absatz 4 des Vertrages über die Europäische Union, umgesetzt in der Verordnung EU 211/2011 bzw. <http://www.bundeswahlleiter.de/de/europawahlen/downloads/rechtsgrundlagen/EUV.pdf>, Artikel 11.
- 37 Zu Einzelheiten siehe auch http://www.bmi.bund.de/DE/Themen/Moderne-Verwaltung/Open-Government/Europaeische-Buergerinitiative_node.html
- 38 Siehe dazu auch <https://govdata.de/>; nähere Erläuterungen enthält z. B. <https://www.govdata.de/hilfe>
- 39 Siehe http://www.bmi.bund.de/DE/Themen/Moderne-Verwaltung/Open-Government/StatistischesBundesamt/statistischesBundesamt_node.html

- 40 Siehe dazu auch http://www.bundesregierung.de/Webs/Breg/DE/Themen/Nachhaltigkeitsstrategie/_node.html. Diese Strategie wird von einem statistischen Monitoring begleitet. Es wird vom Statistischen Bundesamt in zweijährlichem Turnus aktualisiert; siehe dazu auch https://www.destatis.de/DE/Publikationen/Thematisch/Umwelt-oekonomischeGesamtrechnungen/Umweltindikatoren/Indikatoren-PDF_0230001.pdf;jsessionid=CD2731A910418B8C596233D-F49274AC1cae3?__blob=publicationFile
- 41 Auch die Enquete-Kommission des Deutschen Bundestages „Wachstum, Wohlstand, Lebensqualität“ empfiehlt in ihrem Schlussbericht, angenommen vom Deutschen Bundestag am 6. Juni 2013, ein noch näher festzulegendes statistisches Monitoring „W3-Indikatoren“; siehe dazu Schlussbericht der Enquete-Kommission „Wachstum, Wohlstand, Lebensqualität – Wege zum nachhaltigen Wirtschaften und gesellschaftlichem Fortschritt in der Sozialen Marktwirtschaft“, hrsg. von der Bundeszentrale für politische Bildung, Bonn 2013, hier S. 28 f., <http://www.bpb.de/shop/buecher/schriftenreihe/175745/schlussbericht-der-enquete-kommission>
- 42 Siehe dazu http://www.bundesregierung.de/Webs/Breg/DE/Bundesregierung/BeauftragteFuerIntegrations/weitereschwerpunkte/monitoring/_node.html
- 43 Siehe dazu besonders <http://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/2.-Atlas-zur-Gleichstellung-in-Deutschland.property=pdf,bereich=bmfsfj,-sprache=de,rwb=true.pdf>
- 44 Siehe zur Gesundheitsberichterstattung des Bundes auch <http://www.gbe-bund.de/>
- 45 Siehe <https://www.bildungsmonitoring.de/bildung/online/logon>
- 46 Siehe dazu http://www.berlin.de/imperia/md/content/lb-integration-migration/publikationen/berichte/bericht_2011_lag_indikatorenentwicklung_monitoring_teil_1_bericht.pdf?start&ts=1363878767&file=bericht_2011_lag_indikatorenentwicklung_monitoring_teil_1_bericht.pdf
- 47 Siehe dazu <https://www.bildungsmonitoring.de/bildung/online/logon>
- 48 Einen Überblick dazu gibt Schnorr-Bäcker, S.; Heilemann, U.: Methoden und Grenzen der Marktbeobachtung im EU-Binnenmarkt und den Mitgliedsstaaten, Endbericht Forschungsvorhaben Nr. 61/09 im Auftrag des Bundesministeriums für Wirtschaft und Technologie, Wiesbaden im Mai 2011, besonders Kapitel 3, S. 60 ff.; über den aktuellen Stand informiert die offizielle Webseite http://ec.europa.eu/europe2020/index_en.htm
- 49 Siehe dazu die territoriale Agenda der Europäischen Union vom 24./25. Mai 2011 (kurz: Leipzig Charta 2007) und weitere Informationen unter http://www.bmvi.de/DE/StadtUndLand/EuropaeischeStadt/europaeische-stadt_node.html
- 50 Siehe dazu <http://www.stiglitz-sen-fitoussi.fr/en/index.htm>; zu Einzelheiten siehe den Bericht: Stiglitz, J.E.; Sen, A.; Fitoussi, J.-P.: Report by the Commission on the Measurement of Economic Performance and Social Progress, http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf
- 51 Siehe dazu: OECD (Hrsg.): How's life? 2013, Measuring Well-Being, Paris, 2013 bzw. <http://www.oecd.org/statistics/howlife.htm>
- 52 Siehe dazu <http://www.oecd.org/gov/regional-policy/How's-life-in-your-region-Flyer.pdf>
- 53 Siehe Schnorr-Bäcker, S.; Bömermann, H.: Regionalstatistisches Datenangebot zur Fortschrittsmessung, Zum OECD-Ansatz „How's life – Measuring Well-Being“, in: Zeitschrift für amtliche Statistik Berlin Brandenburg, 2/2013, S. 30 ff.
- 54 Siehe dazu auch <http://www.post2015hlp.org/> wie auch Anmerkung 36
- 55 Siehe dazu auch UN Statistical Commission (Hrsg.): Some national, regional and international efforts and practices in the measurement of sustainable development and human well-being, prepared by Friends of the Chair Group on broader measures of progress, New York, February 2014; <http://unstats.un.org/unsd/broader-progress/pdf/BG-FOC-Broader%20measures-Practices%20on%20broader%20measures%20of%20progress.pdf>
- 56 Siehe dazu UN Statistical Division in collaboration with the Friends of the Chair group on broader measures of progress, (Hrsg.): Compendium of statistical notes for the Open Working Group on Sustainable Development Goals (OWG), New York, März 2014 bzw. <http://unstats.un.org/unsd/broader-progress/work.html> sowie die Bewertung „Response to the SDSN indicator proposal“ vom April 2014 auf der o.g. Webseite.
- 57 Siehe dazu auch United Nations: Report of the High Level Panel, a.a.O., z. B. „Executive summary“ sowie [http://unstats.un.org/unsd/broader-progress/pdf/Summary%20of%20Comments%20on%20HLP%20Report%20\(2013-10-30\).pdf](http://unstats.un.org/unsd/broader-progress/pdf/Summary%20of%20Comments%20on%20HLP%20Report%20(2013-10-30).pdf)
- 58 Siehe dazu konsolidierte Fassung des Vertrags über die Arbeitsweise der Europäischen Union (ABl. EU C326/47 vom 26.10.2012), hier Artikel 2 sowie Artikel 174.
- 59 Zu Einzelheiten siehe besonders http://ec.europa.eu/regional_policy/how/index_de.cfm
- 60 Grundlage der aktuellen Kohäsionspolitik ist die Verordnung (EU) Nr. 1303/2013 des Europäischen Parlaments und des Rates vom 17. Dezember 2013 (ABl. EU L 347 vom 20.12.2013) mit gemeinsamen Bestimmungen über den Europäischen Fonds für regionale Entwicklung, den Europäischen Sozialfonds, den Kohäsionsfonds, den Europäischen Landwirtschaftsfonds für die Entwicklung des ländlichen Raums und den Europäischen Meeres- und Fischereifonds sowie mit allgemeinen Bestimmungen über den Europäischen Fonds für regionale Entwicklung, den Europäischen Sozialfonds, den Kohäsionsfonds und den Europäischen Meeres- und Fischereifonds und zur Aufhebung der Verordnung (EG) Nr. 1083/2006 des Rates, (ABl. EU L 210 vom 31.7.2006); zu weiteren Rechtsakten siehe auch http://ec.europa.eu/regional_policy/information/legislation/index_en.cfm
- 61 Siehe dazu auch http://ec.europa.eu/regional_policy/what/future/index_en.cfm
- 62 Eine Übersicht über die Publikationen enthält http://ec.europa.eu/regional_policy/information/reports/index_en.cfm
- 63 Siehe dazu auch Europäische Kommission: In Europas Zukunft investieren, Fünfter Bericht über den wirtschaftlichen, sozialen und territorialen Zusammenhalt, Brüssel November 2010.
- 64 Siehe Europäische Kommission: Die regionale und urbane Dimension der Krise, Bericht der Kommission, Brüssel Juni 2013.
- 65 Siehe dazu auch http://ec.europa.eu/regional_policy/activity/urban/audit/index_en.cfm
- 66 Für Deutschland siehe Urban Audit <http://www.staettestatistik.de/urban-audit.html>
- 67 Vgl. Statistisches Bundesamt (Hrsg.): Lebens- und Arbeitsbedingungen in fünf deutschen Großstädten und ihren Partnerstädten in Frankreich und England – Ausgewählte Indikatoren, Wiesbaden 2013.
- 68 Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), (ABl. L 108 S.1) bzw. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ.L.2007:108:0001:0014:EN:PDF>
- 69 Siehe INSPIRE-Richtlinie a.a.O., hier Erwägungsgrund (3) ff.
- 70 Siehe dazu auch Commission Regulation (EU) No 1253/2013 of 21 October 2013 amending Regulation (EU) No. 1089/2010 implementing Directive 2007/2/EC as regards interoperability of spatial data sets and services; zu diesen und weiteren Rechtsakten siehe <http://inspire.ec.europa.eu/index.cfm/pageid/3>
- 71 Zum Aufbau von INSPIRE vor allem in technischer Hinsicht siehe z. B. http://www.geodaten.niedersachsen.de/portal/live.php?navigation_id=8655&article_id=25473&psmand=28
- 72 Für den Bund siehe Geodatenzugangsgesetz (GeoZG) vom 10. Februar 2009 (BGBl. I S. 278), das durch Artikel 1 des Gesetzes vom 7. November 2012 geändert worden ist.
- 73 Siehe dazu auch <http://www.geoportal.de/GE/DE/GDI-/gdi-de.html?lang=de>
- 74 Siehe <https://www.regionalstatistik.de/genesis/online/logon>
- 75 Welche Tabellen aus dieser Datenbank abrufbar sind, darüber gibt der RegioStat-Katalog Auskunft; siehe <https://www.destatis.de/DE/Publikationen/Thematisch/Regionales/RegioStatkatalog.html>

Zensus 2011

▣ Kleinräumige Gliederung, Georeferenzierung und Rasterdarstellung im Zensus

von Tobias Kirchner, Franziska Pflanz, Andreas Techen, Lars Wagenknecht

Die räumliche Darstellung von statistischen Ergebnissen und Indikatoren kann grundsätzlich zwei kartografischen Mustern folgen: Einerseits können die Ausprägungen des zu betrachtenden Merkmals vorgegebenen Gebietseinheitskategorien, etwa Landkreisen, Gemeinden oder untergemeindlichen kleinräumigen Gliederungen, zugeordnet werden. Darüber hinaus besteht die Möglichkeit der Ergebnisdarstellung in Form einer georeferenzierten Rasterung. Hierbei stützt man sich nicht auf administrative Gebietseinheiten, sondern erzeugt ein Netz aus lückenlosen, identischen Flächensegmenten. Bei der Datenaufbereitung und Ergebnisveröffentlichung des Zensus 2011 sind beide Wege möglich. Doch während der Aufbau und die Aufbereitung kleinräumiger Gliederungssysteme in der amtlichen Statistik keineswegs ein neues Thema ist, betritt der Zensus mit der Georeferenzierung und Rasterung seiner Daten teilweise Neuland. Dabei sind etwa mit der Zuordnung der Zensusanschriften zu den Geokoordinaten vielfältige methodische Probleme zu lösen.

1. Kleinräumige Auswertung von Zensusdaten

Mit der Veröffentlichung der endgültigen Ergebnisse des Zensus 2011 ist ein Datensatz verfügbar, der vielfältige neue Möglichkeiten für die kleinräumige Auswertung von Personen-, Haushalts-, Gebäude- und Wohnungsdaten bietet. Zur Beantwortung kommunaler Fragestellungen ist entscheidend, dass die Daten auch unterhalb der Gemeindeebene – beispielsweise differenziert nach Gemeindeteilen, Planungsräumen oder anderen Gebietseinteilungen – entsprechend strukturiert verfügbar sind und den relevanten Informationsbedarf abdecken. Diese Ergebnisse sind für die kommunale Ebene wichtig, da die lokalen Akteure dadurch bei ihren Planungen und Entscheidungen unterstützt werden können und damit die Basis für eine effiziente Gestaltung kommunaler Politik und Verwaltung verbessert wird.

Die kleinräumige Gliederung versteht sich als Lokalisierungs- und Zuordnungssystem. Sie gründet sich auf Straße und Hausnummer, also auf die Adresse als Ortsangabe und eine bis zum (Bau-)Block und zur Blockseite differenzierte räumliche Gliederung des Gemeindegebietes [1]. Unterhalb der Gemeindeebene darf nach § 10 Absatz 2 BStatG¹ die Blockseite als Merkmal der regionalen Zuordnung genutzt werden. Daraus ergibt sich, dass auf Ebene der Blockseite Auswertungen erstellt werden können, soweit die Grundsätze zur statistischen Geheimhaltung von Einzelangaben gewahrt bleiben. Die Auswertungstabellen unterliegen dabei den gesetzlichen Vorgaben der statistischen Geheimhaltung (§ 16 BStatG). Damit wird vorab sichergestellt, dass durch veröffentlichte Tabellen keine Rückschlüsse auf Einzelangaben von Auskunftspflichtigen möglich sind.

Durch Aggregationen der Blockseiten sind Auswertungen auf allen höheren Gliederungsebenen realisierbar. Da die Erfordernisse der statistischen Geheimhaltung berücksichtigt werden müssen, ist die fachliche Tiefe der Auswertungsmöglichkeiten abhängig von der Größe der gewählten Gliederungsebene. Generell können deshalb für einen Ortsteil detailliertere Ergebnisse ausgewertet werden als für (Bau-)Blöcke und Blockseiten.

Für Berlin stehen dem Amt für Statistik Berlin-Brandenburg (AfS) für die Auswertung der Zensusergebnisse bereits vielfältige und regional tiefe Gliederungsebenen zur Verfügung, von denen die Lebensweltlich Orientierten Räume (LOR) sicherlich zu den wichtigsten gehören. Und auch für Brandenburg sind die Zensusdaten mit einer flächendeckenden kleinräumigen Gliederung auf Ortsteilebene verknüpft.

Zusätzlich haben die Kommunen und Ämter in Brandenburg die Möglichkeit, individuelle, kleinräumig gegliederte Auswertungen unterhalb der Gemeindeebene zu erhalten. Aktuell haben 54 Städte/Gemeinden und elf Ämter dem AfS insgesamt 110 kleinräumige Gliederungssysteme übermittelt und eine entsprechende Datenaufbereitung veranlasst. Rund 72% der übermittelten Gliederungssysteme kommen dabei aus Städten und Gemeinden mit mindestens 10 000 Einwohnern².

¹ Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749).

² Einwohner zum Zensus-Stichtag 9. Mai 2011

Die Struktur dieser Gliederungssysteme ist sehr vielfältig. Die Städte und Gemeinden mit einer Einwohnerzahl unter 10 000 haben häufig nur eine Gliederungsebene – die des Ortsteils – in ihre Gliederung aufgenommen. Nach Blockseiten aufbereitete Zensusergebnisse wurden lediglich von 16 Städten und Gemeinden angefordert. Einige größere Städte haben zu den klassischen drei Ebenen – Stadtteile/Ortsteile, (Bau-)Blöcke, Blockseiten – auch weitere Gliederungsebenen, wie statistische Bezirke oder Quartiere, generiert.

Bei aller kommunalpolitischen Relevanz von diesen überwiegend administrativen und stadtplanerischen Gliederungsebenen fehlt es ihnen in Hinblick auf flächendeckende Auswertungen an Flexibilität und Vergleichbarkeit. So forderte unter anderem die Arbeitsgruppe „Georeferenzierung von Daten“ des Rates für Sozial- und Wirtschaftsdaten (RatSWD) die Schaffung von kleinräumigen, vergleichbaren, regionalen, statistischen Einheiten für Deutschland, entsprechend der europäischen Rahmenrichtlinie INSPIRE (INfrastructure for SPatial InfoRmation in Europe) [2].

Seit dem letzten Jahr darf in Deutschland gemäß der getroffenen Neuregelung im Bundesstatistikgesetz³ eine regionale Zuordnung von Erhebungsmerkmalen an eine geographische Gitterzelle erfolgen [3]. Damit ist erstmals in Deutschland die gesetzliche Grundlage dafür geschaffen worden, amtliche Fachstatistiken wie den Zensus rasterbasiert auszuwerten. Grundlage dieser Rasterauswertungen sind die statistischen Merkmale, die im Fall des Zensus mit einer konkreten Anschrift verknüpft sind. In einem ersten Schritt werden diese Anschriften georeferenziert, das heißt einem Koordinatenpaar, den sogenannten Geokoordinaten, zugeordnet.

2. Zuordnung der Anschriften zu Geokoordinaten

Mit dem Zensus 2011 liegt für Deutschland erstmalig eine georeferenzierte Datei aller Anschriften von Gebäuden mit Wohnraum vor. Dieses Anschriften- und Gebäuderegister diente als zentrales Steuerungs- und Koordinierungsinstrument der verschiedenen Zensuserhebungen und ist bei der Ergebnisauswertung die Datengrundlage zur Herstellung des Raumbezuges. Ein solcher Raumbezug besteht am unmittelbarsten durch die Verwendung von Geokoordinaten. Diese liegen im Zensus auf Ebene der Anschriften vor. Eine Georeferenzierung der erhobenen Gebäude-, Wohnungs- und Personenmerkmale besteht insofern indirekt über die Anschrift. Befinden sich mehrere Wohngebäude an einer Anschrift, werden alle Gebäude dem Koordinatenpunkt der Anschrift zugeordnet.

Die Koordinaten im Anschriften- und Gebäuderegister stammen aus den Lieferungen der „Georeferenzierten Adressdaten – Bund“ (GAB) des Bundesamtes für Kartographie und Geodäsie. Zusammen mit den Datenlieferungen der Melderegister und dem erwerbsstatistischen Register der Bundesagentur für Arbeit wurden sie zum Aufbau und zur Pflege des Anschriften- und Gebäuderegisters genutzt. Die Zusammenführung dieser drei Datenquellen erfolgte auf der Anschriftenebene, das heißt über die Kombination der Merkmale Amtlicher Gemeindeschlüssel, Postleitzahl, Straße, Hausnummer, Hausnummernzusatz und wenn nötig auch über den Ortsteil. Insgesamt wurden zwischen 2008 und 2011 vier GAB-Lieferungen in das Anschriften- und Gebäuderegister integriert. Im Gegensatz zu den personenbezogenen Daten der Meldebehörden und der Bundesagentur für Arbeit, werden die GAB-Register auf Anschriftenebene geführt, ohne dass dabei nach Wohn- und Nicht-Wohnanschriften unterschieden wird. Deshalb wurde eine Anschrift im Anschriften- und Gebäuderegister erst dann zur Anschrift mit Wohnraum erklärt, wenn sie aus mindestens zwei Datenquellen stammte, etwa aus den GAB-Daten und aus dem Melderegister. Anschriften, die etwa lediglich in der GAB-Lieferung vorkamen, mussten von den statistischen Landesämtern manuell hinsichtlich vorhandenen Wohnraums geprüft werden. [4]

Nicht alle Anschriften im Anschriften- und Gebäuderegister fanden sich auch in den Georeferenzierten Adressdaten des Bundes. Ein Teil der Anschriften besaß deshalb noch keine Koordinatenwerte. Am Ende der Erhebungsphase des Zensus waren in Berlin 99,2% und im Land Brandenburg 96,1% aller Anschriften des Anschriften- und Gebäuderegisters geokodiert. Für die Ergebnisdarstellung von Bedeutung sind allerdings nur die sogenannten zensusrelevanten Anschriften⁴. Bei der Betrachtung dieser Anschriftengruppe fällt das Ergebnis der Zuordnung mit Geokoordinaten, zumindest für Brandenburg, besser aus. Demnach lagen am Ende der Erhebungsphase in Berlin für unveränderte 99,2% und in Brandenburg für 97,6% der zensusrelevanten Anschriften Koordinatenwerte vor. Bundesweit betrug die Quote 98,3%. Die eigentliche Herausforderung bestand nun darin, die fehlenden Geokoordinaten für bundesweit 332 516 zensusrelevante Anschriften zu ergänzen. Davon entfielen 2 602 Anschriften auf Berlin und 16 632 Anschriften auf Brandenburg.

Der Großteil dieser noch offenen Anschriften konnte durch das Statistische Bundesamt (Destatis) geokodiert werden. Durch die Nutzung weiterer, ausschließlich zu diesem Zweck verwendeter aktueller GAB-Lieferungen aus den Jahren 2012 und 2013 erfolgte die maschinelle Koordinatenzuordnung von bundesweit 71,8% der noch offenen Fälle. Weitere 6,6% der fehlenden Koordinaten wurden durch Destatis auf Basis manueller Recherchen in öffentlich zugänglichen Quellen ergänzt. Um die Menge der restlichen zensusrelevanten Anschriften ohne Koordinaten erheblich zu reduzieren, erfolgte der

³ Siehe Artikel 13 (Änderung des Bundesstatistikgesetzes) des Gesetzes zur Förderung der elektronischen Verwaltung sowie zur Änderung weiterer Vorschriften vom 25. Juli 2013 (GVBl. I S. 2749).

⁴ Eine Anschrift gilt als zensusrelevant, wenn sie eines der folgenden fünf Kriterien erfüllt: Relevanz für die Ziehung der Haushaltsstichprobe, Vorliegen eines Melderegistereintrages zum Zensusstichtag, es handelt sich um eine Sonderanschrift, das Ergebnis der manuellen Wohnraumprüfung oder der Gebäude- und Wohnungszählung (GWZ) lautet „mit Wohnraum“.

Ausschluss derjenigen Anschriften, an denen trotz Zensusrelevanz am Ende kein Zensusergebnis vorlag. Dies ist der Fall, wenn an der Anschrift im Ergebnis weder eine existente Person noch Wohnraum ermittelt wurde, was auf weitere 8,1% der zensusrelevanten Anschriften ohne Koordinaten zutraf. Die Kenntnis darüber lag erst nach Abschluss der Datenaufbereitung in der Auswertungsdatenbank vor. Im nächsten Schritt erfolgte die Geokodierung von sogenannten reinen Meldeanschriften, die im Anschriften- und Gebäuderegister für jede Gemeinde angelegt wurden, um z.B. Personendatensätze von wohnungslos Gemeldeten zuordnen zu können. Für diese besonderen Anschriften wurden die Koordinatenwerte der jeweiligen Rathausanschrift vergeben, wenn diese im Anschriften- und Gebäuderegister bereits geokodiert vorlag. Für andere noch offene Anschriften konnte unter Zuhilfenahme der Koordinaten von Nachbaranschriften eine maschinelle Ergänzung erfolgen. Voraussetzung dafür war die Lage der Anschrift zwischen zwei Nachbaranschriften mit vorhandenen Geokoordinaten, die maximal 100 m voneinander entfernt liegen. In diesem Fall wurde der Koordinatenmittelwert gebildet und der noch nicht georeferenzierten Anschrift zugewiesen.

Am Ende der Arbeiten zur Anreicherung mit Geokoordinaten beim Statistischen Bundesamt blieben bundesweit noch 34 574 Anschriften, also rd. 10% der ursprünglich im Anschriften- und Gebäuderegister nicht geokodierten zensusrelevanten Anschriften, zur abschließenden Prüfung und Ergänzung durch die statistischen Landesämter übrig. Auf Berlin entfielen 258 und auf Brandenburg 1265 Anschriften.

Bei der Ergänzung der fehlenden Koordinaten wurde für Berlin und Brandenburg unterschiedlich vorgegangen. Für Berlin konnte auf die Geokoordinaten der Anschriften des Regionalen Bezugssystems (RBS) mit Stand Mai 2011 zurückgegriffen werden. Jeder Anschrift im RBS ist stets ein Koordinatenwert zugeordnet. Diese entsprechen größtenteils den Hauskoordinaten der Vermessungsämter, wurden aber zum Teil auch manuell positioniert. Allen 258

noch offenen zensusrelevanten Anschriften konnte per Abgleich eine RBS-Koordinate zugewiesen werden. Da die RBS-Koordinaten im Koordinatensystem „DHDN Soldner Berlin“ vorliegen, wurden die Werte noch in das im Zensus verwendete System „WGS84“ umgerechnet.

Für die restlichen Anschriften ohne Koordinaten im Land Brandenburg mussten die Geokoordinaten manuell recherchiert werden. Um die genaue Lage der Anschrift und die dazugehörigen Koordinatenwerte zu ermitteln, wurden verschiedene Onlineportale, wie Lika-Online, Brandenburg-VIEWER, Geoportal, Straßenkatalog und google maps genutzt. Dabei konnte allerdings nicht in jedem Fall die genaue Koordinatenposition der Anschrift bestimmt werden. Hilfsweise wurden die Koordinaten der unmittelbar benachbarten Anschrift verwendet oder die Straßenmitte bzw. der Mittelpunkt oder Eingangsbereich einer Wohn-/Kleingartenanlage als Koordinatenpunkt gewählt.

Nach der Zuordnung der Geokoordinaten können nun einzelne Anschriften zu beliebig großen Flächen (Rastern) aggregiert und die anschriftenbezogenen, statistischen Merkmale raumbezogen ausgewertet und kartographisch dargestellt werden.

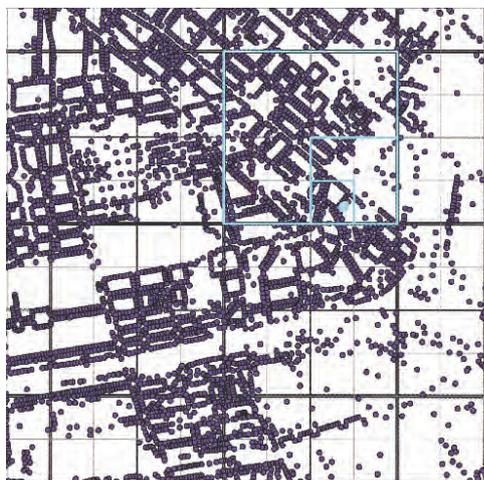
3. Rasterbasierte kartographische Darstellung

In der rasterbasierten Darstellung statistischer Ergebnisse wird ein lückenloses Netz aus identischen Flächeneinheiten⁵ erzeugt. In der Regel kommen hierbei quadratische Rasterzellen – auch Grid genannt – zur Anwendung, wobei die Kantenlänge jeder Zelle, gegebenenfalls unter Berücksichtigung des Datenschutzes, je nach Anwendungs- oder Auswertungszweck festgelegt werden kann [5]. Für jede Rasterzelle, egal welcher Größe, wird in der Regel das Koordinatenpaar des linken unteren bzw. südwestlichen Eckpunktes angegeben, um die Rasterzelle zu definieren beziehungsweise zu identifizieren.

Je nach Ausdehnung des zu rasternden Gebietes ist die Wahl einer Projektion vorzunehmen, um die Erdoberfläche mit einer möglichst geringen Abweichung auf die Ebene zu projizieren [6]. Zur Umsetzung der INSPIRE-Richtlinie wurde das Referenzsystem ETRS89 im Rahmen der Geodateninfrastruktur (GDI-DE) in Deutschland als Standard definiert [7]. Als Projektionsmethode kommt die Lambertsche flächentreue Azimutalprojektion zur Anwendung.

Oftmals werden für einen Indikator Grids mit verschiedenen Rasterweiten und dem gleichen Nullpunkt erstellt, wobei mit einer Vergrößerung des Rasters auch ein Informationsverlust einhergeht [5]. Die Kantenlängen der unterschiedlichen Grids sollten dabei jeweils ein Vielfaches der kleinsten Rasterzelle sein, sodass eine hierarchische Struktur entsteht. Somit besteht die Möglichkeit, einzelne oder alle Rasterzellen auf die nächst höhere (Aggregati-

1 | Zuordnung von Anschriften zu hierarchischen Rastern (250 m, 500 m, 1000 m)



⁵ Die Netzstruktur einer Rasterung kann generell auch das Relief berücksichtigen, womit sich eine dreidimensionale Einteilung der Rasterzellen ergibt.

on) oder die nächst niedere (Disaggregation) Rasterzellenebene zu transformieren. Die Darstellung der Zensusergebnisse wird europaweit auf einer Rastergröße von 1km vorgenommen [8]. In Deutschland erfolgt eine Auswertung auf Rastergrößen von 100 m, 1 km und 10 km Kantenlänge.

Abbildung 1 verdeutlicht das Prinzip der hierarchischen Rasterung mit Hilfe von 250 m-, 500 m- und 1000 m-Rastern. Der markierten Anschrift werden die drei ebenfalls markierten Rasterzellen zugewiesen.

Die Zuordnung der Merkmale zu einer Rasterzelle erfolgt über ein Koordinatenpaar, welches in dem gleichen Bezugssystem vorliegt wie das zu nutzende Raster. Hierbei wird mittels einer räumlichen Zuordnung in einem GIS an den Ausgangsdatensatz ein Feld angehängt, in dem die zugehörige Rasterzelle gespeichert wird. Für Punktdaten, etwa Adressen, ist somit eine eindeutige Zuweisung zu einer Rasterzelle möglich. Die Zuordnung von Linien- oder Polygonlayern kann im GIS mit speziellen Abfragen – etwa zu Flächenanteilen bei der Ermittlung der vorrangigen Landnutzung – erfolgen.

Im Rahmen des Zensus 2011 wird ausschließlich eine Zuordnung von Punktkoordinaten, die für alle Anschriften mit Zensusergebnis vorliegen, zu den jeweiligen Grids vorgenommen. Somit ist jedem Gebäude mit Wohnraum eine 100 m-, eine 1 km- und eine 10 km-Rasterzelle zugeordnet, was dann rasterbasierte Auswertungen auf den Ebenen Wohnungen, Haushalte und Personen zulässt. Sowohl die Koordinaten selbst als auch die Anschriften stehen für Auswertungen der Zensusdaten nicht zur Verfügung, da sie aus datenschutzrechtlichen Gründen spätestens vier Jahre nach Erhebungsbeginn gelöscht werden müssen.

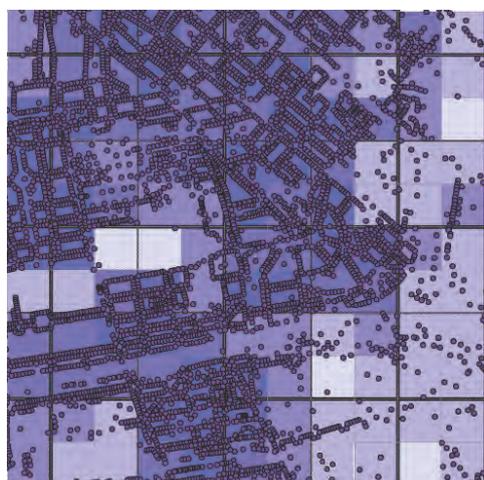
Für jede Zelle können nun statistische Verteilungswerte, wie Summen oder Durchschnittswerte, berechnet und dargestellt werden [6]. In Abbildung 2 ist dem 250 m-Raster exemplarisch die klassifizierte Anzahl der darin befindlichen Gebäude auf 250 m-Basis zugewiesen.

Da in der amtlichen Statistik eine Ausweisung von Werten kleiner 3 aus Datenschutzgründen oftmals nicht erfolgen darf, empfiehlt sich in diesen Fällen die Aggregation auf die nächsthöhere Ebene. Durch Kombination mehrerer hierarchischer Ebenen können dann Mischrasterkarten erstellt werden, sodass weniger Zellen unter einem vorab definierten Grenzwert liegen und somit mit Sachinformation gefüllt werden können. In Abbildung 3 sind diejenigen Flächeneinheiten markiert, die einer Aggregation unterzogen wurden, sodass aus Sicht der statistischen Geheimhaltung eine flächendeckende Ausweisung von Werten erfolgen kann. Die Kategorie „unter 3 Gebäude“, die in Abbildung 2 dargestellt ist, kann somit eliminiert werden.

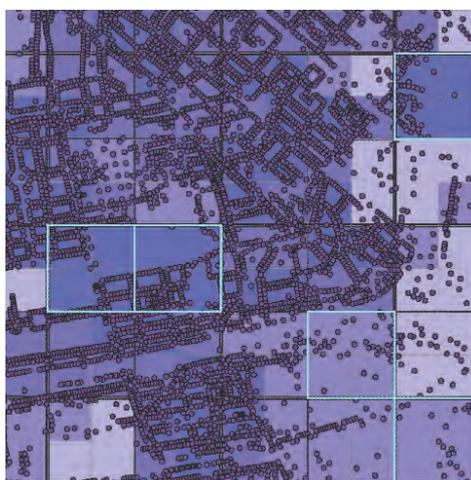
Als Vorteil des Rasters gegenüber einer Darstellung administrativer Einheiten ist hervorzuheben, dass die Raumkategorien durch die einheitliche Ausdehnung der Rasterzellen vergleichbar sind. Auch Zeitreihen sind somit stabil darstellbar, da sie unabhängig von Gebietsstandsänderungen, wie etwa Eingemeindungen, sind [5]. Laut BBSR [9] können durch Rasterdarstellungen sowohl innerstädtische Strukturen als auch Übergänge zwischen Stadt und Umland besser aufgezeigt werden.

Die Datenschutzerfordernisse können eingehalten werden, da bei hierarchischer Rasterung die Möglichkeit besteht, zu schwach besetzte Zellen mit benachbarten Rasterzellen zu aggregieren, um einen Grenzwert zu erreichen, der ausgewiesen werden kann bzw. darf. Die so entstehenden Mischrasterkarten folgen dem Grundsatz „so detailliert wie möglich, so stark aggregiert wie (datenschutzrechtlich) nötig“ [10]. Darüber hinaus können unterbesetzte Zellen nach wie vor gekennzeichnet und nicht dargestellt beziehungsweise in der gleichen Art und Weise wie leere Rasterzellen dargestellt werden.

2 | Gebäudeanzahl als Rasterdarstellung (Kantenlänge 250 m)



3 | Gebäudeanzahl als Mischrasterkarte (Kantenlänge 250 m und 500 m)



4. Neue Möglichkeiten für die Nutzung von Zensusdaten

Mit der Georeferenzierung und Rasterung der Zensusdaten besteht erstmals eine einheitliche Möglichkeit, amtliche Bevölkerungs- und Wohnungsdaten unabhängig von administrativen Gebietseinheiten und deren Änderungen kleinräumig auszuwerten. So können, je nach Rastergröße, in unterschiedlicher Auflösung und Detailliertheit raumbezogene Sozialstrukturanalysen durchgeführt werden. Diese Ergänzung der klassischen kleinräumigen Gliederungssysteme wird zukünftig nicht nur die Möglichkeiten bei der Veröffentlichung statistischer Ergebnisse in Form von kartographischen Informationssystemen erweitern, sondern vor allem die Dateninfrastruktur für kommunale Entscheidungsträger und auch für die Wissenschaft verbessern. Auch wenn durch den nunmehr drei Jahre zurückliegenden Zensus 2011 der tatsächliche Informationsgewinn dieser neuen Auswertungs- und Darstellungsmöglichkeiten ein wenig getrübt ist, werden auch hier wichtige Maßstäbe für den Raumbezug anderer amtlicher Statistiken und insbesondere für die Auswertung zukünftiger Volkszählungen gesetzt.

Tobias Kirchner, Diplom-Geograph, ist seit 2011 im Referat *Zensus* des Amtes für Statistik Berlin-Brandenburg tätig, aktuell im Bereich Gebäude- und Wohnungszählung mit den Schwerpunkten Aufbereitung und Auswertung raumbezogener Daten sowie thematische Kartographie. Vorher war er Mitarbeiter bei der Gesellschaft für Markt- und Absatzforschung, Ludwigsburg.

Franziska Pflanz, B. A. Business Administration, ist seit 2011 im Referat *Zensus* des Amtes für Statistik Berlin-Brandenburg tätig, aktuell im Bereich Datenaufbereitung mit Schwerpunkt kleinräumige Auswertung.

Dr. Andreas Techen studierte Sozialökonomie in Kiel und promovierte in Kiel und Nürnberg über soziale Netzwerke. Im Amt für Statistik Berlin-Brandenburg war er bis August 2014 tätig – zunächst als Referatsleiter für Energie- und Industriestatistik, später als Koordinator für die Auswertungsphase und zuständig für die Datenaufbereitung des Zensus 2011.

Lars Wagenknecht, Diplom-Geograph, ist seit 2010 im Referat *Zensus* des Amtes für Statistik Berlin-Brandenburg tätig, aktuell im Bereich Datenaufbereitung mit Schwerpunkt Kartographie und Raumbezug.

Literatur

- [1] Deutscher Städtetag 1991: „DST-Beiträge zur Statistik und Stadtforschung“, Reihe H, Heft 39, S. 5.
- [2] RatSWD – Rat für Sozial- und Wirtschaftsdaten 2012: Endbericht der AG „Georeferenzierung von Daten“.
- [3] Bundesgesetzblatt Jahrgang 2013 Teil I Nr. 43, S. 2757.
- [4] Heidrich-Riske, Holger; Scholz, Bettina; Stepien, Halina 2013: GIS-gestützte Ermittlung der „EU-Orte“ im Rahmen des Zensus 2011 für die Datenlieferungen an Eurostat, „Wirtschaft und Statistik“, Heft 7/2013, S. 467-475.
- [5] Wonka, Erich; Kaminger, Ingrid; Katzlberger, Gernot 2009: Regionalstatistische Auswertungen mit geographischen Rastern in der Raumplanung, „Informationen zur Raumentwicklung“, Heft 10/11 2009, S. 661-675.
- [6] Kaup, Stefan; Rieffel, Philippe 2013: „Rasterbasierte Regionalstatistik“, Hrsg.: ILS – Institut für Landes- und Stadtentwicklungsforschung, Dortmund.
- [7] RatSWD – Rat für Sozial- und Wirtschaftsdaten 2012: „Georeferenzierung von Daten – Situation und Zukunft der Geodatenlandschaft in Deutschland“, Berlin.
- [8] eurostat: „Statistics Explained – Population grids“. http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_grids [Zugriff: 16.04.2014]
- [9] BBSR – Bundesinstitut für Bau-, Stadt- und Raumforschung 2013: Stadtbeobachtung im BBSR auf neuen Wegen, „Informationen aus der Forschung des BBSR“, Nr. 2/2013, S. 5.
- [10] Strobl, Josef 2005: Hierarchische Aggregation – Detailinformation versus Datenschutz am Beispiel adressbezogen georeferenzierter Datensätze, „Salzburger Geographische Arbeiten“, Bd. 38, S. 163-171.

Fachgespräch mit Dr. Andreas Techen

Die rasterbasierten Auswertungen stehen erst am Anfang



Dr. Andreas Techen war bis zu seinem Ausscheiden im August 2014 im Referat Zensus des Amtes für Statistik Berlin-Brandenburg für die Datenaufbereitung zuständig.

In welcher Form wird die amtliche Statistik der breiten Öffentlichkeit rasterbasierte Ergebnisse des Zensus 2011 zur Verfügung stellen?

Die kleinräumigen, rasterbasierten Auswertungen der Zensusdaten mit ihrem großen Analysepotenzial stehen erst am Anfang. Natürlich werden auch diese Ergebnisse veröffentlicht, sobald sie vorliegen. So ist unter anderem ein Online-Tool in Planung, über das individuelle Rasterauswertungen und ihre Visualisierung über Karten möglich sein werden. In der Landwirtschaftsstatistik gibt es das mit dem Agrar-Atlas schon seit einigen Monaten und der Zensus wird mit einem solchen Angebot folgen.

Herr Dr. Techen, Sie sprechen von der enormen Wichtigkeit kleinräumiger Zensusdaten für die kommunale Politik. Wie können beispielsweise Brandenburger Gemeindeverwaltungen auf die Daten zugreifen?

Nutzen die Kommunen mit abgeschotteten Statistikstellen die Einzeldaten des Zensus im Rahmen des § 22 ZensG für kommunalstatistische Zwecke, sind natürlich auch Auswertungen auf Basis der Rasterzellen möglich, die Bestandteil des Datensatzes sind. Allen anderen Gemeinden bietet das Amt für Statistik Berlin-Brandenburg auf Anfrage maßgeschneiderte kleinräumige Sonderauswertungen an.

Welche Rastergrößen sind Ihrer Meinung nach für Berlin und welche für Brandenburg am sinnvollsten?

Das hängt entscheidend von der Fragestellung ab. Allgemein gilt, dass die Aussagekraft der Ergebnisse mit der Größe der Rasterzellen zunimmt, ganz einfach, weil so auch die Fallzahl steigt und die statistische Geheimhaltung weniger stark in die Daten eingreifen muss. Natürlich machen manche Fragestellungen, z. B. der Raumplanung, keinen Sinn, wenn sie in einem 10x10 km-Raster ausgewertet werden. Hier muss man Erkenntnisziel, Datenqualität und Datenschutz sorgfältig abwägen. In Berlin wird diese Abwägung aufgrund der dichteren Besiedelung häufig leichter fallen als in einem Flächenland wie Brandenburg.

Inwieweit können Einzeldaten des Zensus 2011 für Auswertungen genutzt werden?

Einzeldaten werden natürlich nicht original veröffentlicht. Jedoch ermöglicht das Bundesstatistikgesetz den Zugang zu faktisch anonymisierten Einzeldaten für die Wissenschaft. So wie in anderen Statistiken auch werden diese Daten über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder im Laufe des Jahres 2015 bereitgestellt.

Fachstatistische Anwendungen

Die kleinräumige Gliederung

Grundlage eines Raumbewachungssystems für ein strategisches Controlling in der Landeshauptstadt Potsdam

von Lutz Rittershaus

In Bundes-, Landes- und Kommunalstatistiken werden eine Vielzahl von Daten für die Städte erhoben und den Bedarfsträgern zur Verfügung gestellt. Diese Daten sind wichtige Informationsquellen und werden für Veröffentlichungen und Vergleiche mit anderen Städten genutzt. Um spezielle Planungs- und Verwaltungsaufgaben in der Kommune lösen zu können, werden jedoch räumlich differenzierte Daten benötigt. Daten für die Gesamtstadt genügen den Ansprüchen nicht. Um aufgaben- und kundenorientiert Informationen zu erstellen, wurde deshalb in vielen Städten das System der kleinräumigen Gliederung eingeführt.

Kleinräumige Gliederung

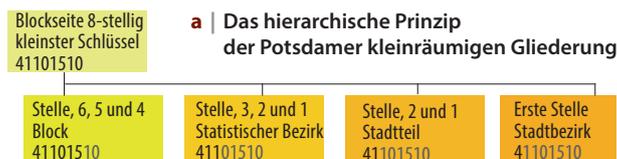
Die kleinräumige Gliederung ist die Einteilung der Kommune in kleinräumige, sich auf die Adresse beziehende Bereiche. Die unterste Ebene dieser Bereiche ist die Blockseite.

Daten, die in den Fachämtern vorliegen und sich auf Adressen, Blockseiten bzw. auf andere aggregierte Einheiten der kleinräumigen Gliederung beziehen, haben einen Raumbezug. Hierbei wird unterschieden nach direktem Raumbezug, das sind Adressen mit Koordinaten, und indirektem Raumbezug: Dabei handelt es sich um die Elemente der hierarchischen Gliederung, wie Blockseite, Block oder Stadtteil.

Der Raumbezug oder auch die Georeferenz verbindet die Daten aus den unterschiedlichen Fachämtern mit dem Ort dieser Daten. Das heißt, die Daten sind einer bestimmten Adresse oder anderen aggregierten Einheiten der kleinräumigen Gliederung zugeordnet. Durch den Raumbezug werden diese Daten zu Geodaten, also örtlich referenzierten Daten.

Zusätzlich kann die Zugehörigkeit einer Adresse im Stadtgebiet zu jeder weiteren Gebietsgliederung bestimmt werden, da sich die Elemente der kleinräumigen Gliederung zu weiteren Gebietseinteilungen, wie Schulbezirke, Planungsräume oder Wahlbezirke, zusammenfassen lassen. Durch die Nutzung eines Raumbezugssystems auf der Grundlage der kleinräumigen Gliederung ist es möglich, sehr effektiv Raumbezugs- und Sachdaten zu verknüpfen und eine kundenorientierte Informationswelt aufzubauen.

Gepflegt und fortgeschrieben wird die kleinräumige Gliederung mit dem Programm „AGK“ (Adresszentraldatei, Gebäudedatei und Kleinräumige Gliederung) aus dem KOSIS-Verbund (Kommunales Statistisches Informationssystem).



Zusammen mit dem Geographischen Informationssystem (GIS) wird die kleinräumige Gliederung im Bereich für Statistik und Wahlen der brandenburgischen Landeshauptstadt Potsdam seit 1991 auf- und ausgebaut bzw. fortgeschrieben und ist ein Bestandteil des gesamten Statistischen Informationssystems der Landeshauptstadt.

Die Potsdamer kleinräumige Gliederung besteht aus 8 Stadtbezirken, 34 Stadtteilen, 84 Statistischen Bezirken und ca. 2 000 Blöcken. Ein Block enthält als kleinste Einheit die Blockseite. Jede Blockseite hat eine 8-stellige Schlüsselnummer, die die Kennung der übergeordneten Gebietseinheiten enthält.

Als Orientierung für die Grenzen der kleinräumigen Gliederung dient der Aufbau der Stadt bzw. deren topographische Struktur.

Straßen, Wege, Ufer oder auch Gräben bilden die Blockgrenzen. In Ausnahmefällen werden Blöcke aber auch durch administrative Linien wie Flur- oder Grundstücksgrenzen beschrieben.

Die Blockgrenzen verlaufen in der Landeshauptstadt Potsdam straßenmittig. Zur Digitalisierung wurden Stadtkarten im Maßstab 1:5 000 und 1:10 000 verwendet, auf denen, nach der Erstbegehung 1991, die Blockgrenzen eingezeichnet und digitalisiert wurden.

Die nunmehr digitale Stadtgrundkarte unterliegt einer ständigen Aktualisierung, die Blockgrenzen werden analog zu dieser Aktualisierung der moderneren und digitalen Kartengrundlage angepasst.

Visualisierung der kleinräumigen Gliederung

Das Geographische Informationssystem ist ein Programm für computergestützte Kartierung.

Die meisten der im Amt vorliegenden statistischen Daten haben einen Raumbezug (Adresse) und somit eine geografische Komponente.

Mit der GIS-Software werden die (aggregierten) statistischen Daten über die ID mit den geographischen Daten verbunden und die Ergebnisse in einer digitalen Karte visualisiert. Das heißt, der Raumbezug dieser Daten kann graphisch dargestellt werden. Dieses erfolgt beispielsweise mit Hilfe unterschiedlichster thematischer Karten oder auch raumbezogener Berechnungen für die verschiedenen Fachämter.

Dadurch werden Strukturen und Zusammenhänge erkennbar. Die so erhaltenen Informationen können ausgewertet und für kommunale Planung und Verwaltung eingesetzt werden, aber auch private Unternehmen nutzen dieses Wissen zunehmend. Als Beispiel sei hier die Standortplanung eines Supermarktes genannt.

Möglich sind diese Auswertungen bzw. Darstellungen für sämtliche andere Gebiete, die sich aus der kleinräumigen Gliederung aggregieren.

Jede digitalisierte Fläche, jede Linie und jedes Punktobjekt enthält außerdem Daten, die implizit vorhanden sind: Dies sind Daten zur Fläche, zur Länge und die Koordinaten des digitalen Objektes. Diese Daten können für weitere Analysen und Berechnungen genutzt werden.

Aufbau der kleinräumigen Gliederung in der Landeshauptstadt Potsdam

Die Blockseite ist das kleinste digitalisierte Element der kleinräumigen Gliederung.

„Blockseite ist innerhalb eines Gemeindegebiets die Seite mit gleicher Straßenbezeichnung von der durch Straßeneinmündungen oder vergleichbare Begrenzungen umschlossenen Fläche.“ (§ 10 Abs. 3 BstatG) (Abbildung b).

Eine Blockseite kann aber auch ein Abschnitt eines Ufers, ein Feld- oder ein Waldweg sein. Besonders im ländlichen Raum sind Blockseiten ohne Straßenzuordnung häufig zu finden. Eine Identifizierung ist zum Beispiel über die Beschreibung möglich (Abbildung c).

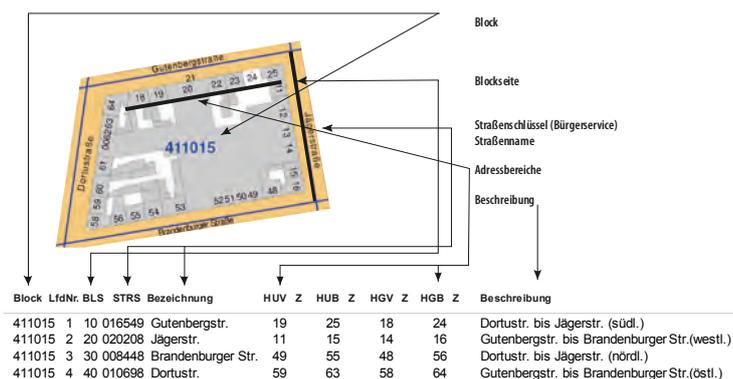
Mehrere zusammenhängende Blockseiten bilden einen Block. Statistische Blöcke werden in der Regel von Straßen, natürlichen oder baulichen Grenzen umschlossen und weisen möglichst eine homogene Struktur auf.

Der Block 411015 wird durch die Blockseiten 41101510, 41101520, 41101530 und 41101540 begrenzt.

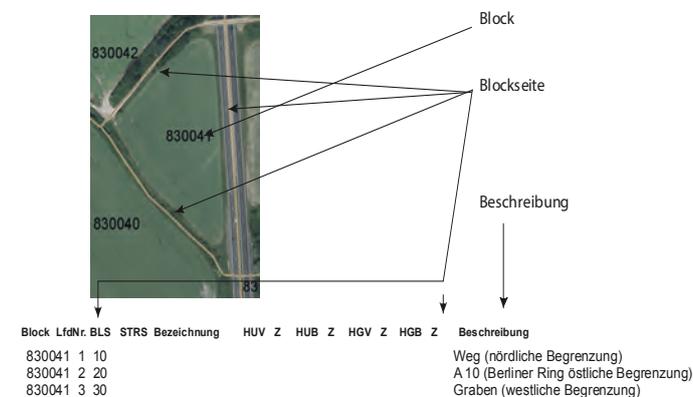
Fasst man alle Blöcke zusammen, die an den ersten drei Stellen ihrer 6-stelligen Schlüsselnummer die gleiche Kennung haben, ergibt dies einen Statistischen Bezirk.

Die Blöcke, deren Schlüsselnummer mit 411 beginnen, ergeben den Statistischen Bezirk 411 – Barocke Stadterweiterung.

b | 5. und 4. Ebene Blockseite und Block

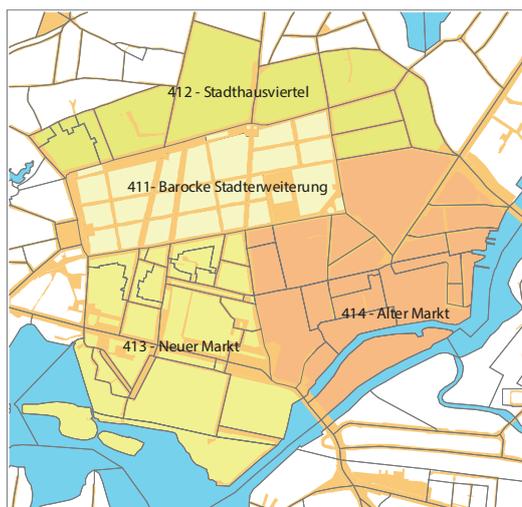


c | Block ohne Adressen bzw. Gebäude



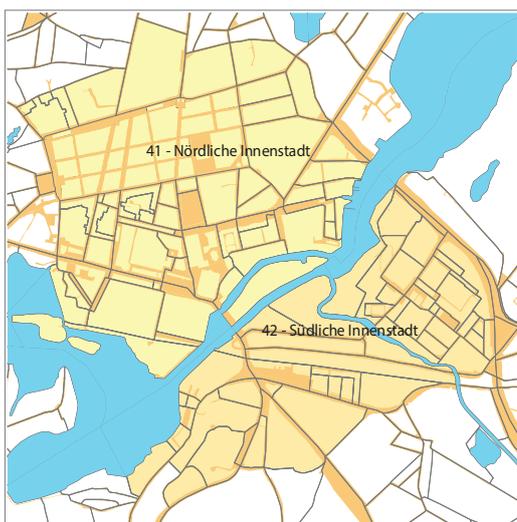
d | 3. Ebene – Statistischer Bezirk



e | 2. Ebene – Stadtteil

Der Stadtteil aggregiert sich aus den Blöcken bzw. Statistischen Bezirken, die an den ersten zwei Stellen ihrer 6-stelligen bzw. 3-stelligen Schlüsselnummer die gleiche Schlüsselnummer haben.

Die Statistischen Bezirke mit der Kennung
411 – Barocke Stadterweiterung,
412 – Stadthausviertel,
413 – Neuer Markt und
414 – Alter Markt
ergeben zusammengefasst den Stadtteil
41 – Nördliche Innenstadt.

f | 1. Ebene – Stadtbezirk

Nach dem gleichen Prinzip aggregiert sich der Stadtbezirk. Sämtliche Objekte der kleinräumigen Gliederung, die an der ersten Stelle ihrer Schlüsselnummer die gleiche Ziffer haben, bilden einen Stadtbezirk: hier den Stadtbezirk 4 – Innenstadt.

Beispiele raumbezogener Auswertungen

Wie bereits erwähnt, enthält jede digitalisierte Fläche, jede Linie und jedes Punktobjekt Daten, die implizit vorhanden sind, das heißt diese Daten werden beim Digitalisieren automatisch ermittelt, die Blocknummer (Block-Id) wird manuell vergeben. Die Daten, die aus unterschiedlichen Registern oder Erhebungen vorliegen, werden im Bereich Statistik auf

Blockseite bzw. Block aggregiert und können so über die Blocknummer mit der Geometrie verknüpft und visualisiert werden.

Das Gleiche gilt für die anderen Ebenen der kleinräumigen Gliederung oder andere digital vorhandene Geometrien, wie zum Beispiel Planungsräume, Wahlbezirke, Grundschulbezirke und Schulstandorte.

In der Landeshauptstadt Potsdam sind die Gebiete, die im Bereich Statistik ausgewertet werden, „blockgenau“ abgegrenzt. So hat jeder Block weitere Attribute, wie Wahlbezirk, Sozialraum oder Planungsraum, zugeordnet (Abbildung g).

Dadurch ist es möglich, schnell standardisierte statistische Auswertungen zu erzeugen und auch Gebiete, die nicht zur kleinräumigen Gliederung gehören, über die Blöcke abzubilden.

Verknüpfung kleinräumiger Gebiete mit Sachdaten – kleinräumige Auswertungen

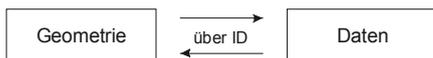
Je nach Datengrundlage lassen sich die Sachdaten als Tabelle, Diagramme, Punktdichtekarte oder als Farbabstufung mit Hilfe des GIS visualisieren.

Die Blockseiten sind aus Gründen des Datenschutzes nur in wenigen Fällen für kleinräumige Auswertungen geeignet. Hierfür eignen sich am besten die Statistischen Bezirke. Sie sind nicht zu klein, sodass der Datenschutz gewährleistet wird, und nicht zu groß, sodass die unterschiedlichen Strukturen der Stadt abgebildet werden können.

Für einige Fachbereiche werden die statistischen Daten auf Grundlage fachspezifischer Räume ausgewertet (Neubaugebiete, Sanierungsgebiete oder Entwicklungsbereiche).

Für diese Beobachtungsgebiete wurden die Blöcke so abgegrenzt, dass sie deren Struktur abbilden.

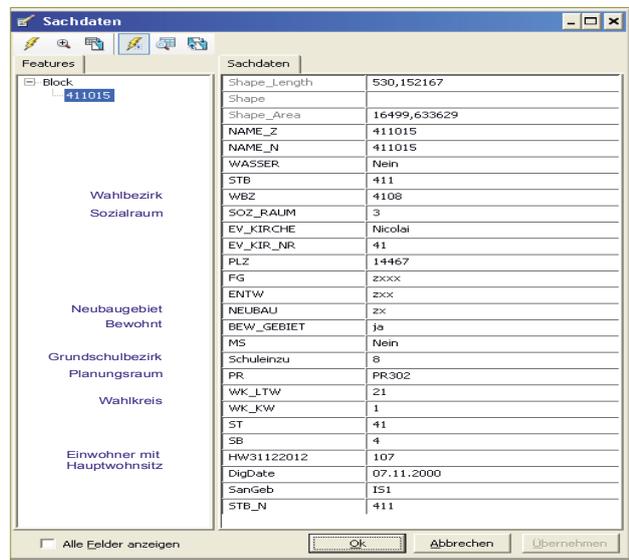
Das Geographische Informationssystem ist so aufgebaut, dass Nutzerinnen und Nutzer sowohl über die Geometrie zu den Daten kommen als auch über die Daten zur Geometrie. Dieses erfolgt über die Such- bzw. Abfragefunktion.



g | Beispiel 1

statistischer Block			
	Shape Length	Shape Area	BlockId
	488,913665	14899,441975	411001
	671,421213	25605,313073	411002
	525,693576	14331,163411	411003
	559,645892	18413,016609	411004
	670,131423	27594,548795	411005
	561,532185	15578,777348	411006
	502,363115	15166,322639	411007
	492,925766	14712,411853	411008
	603,441693	20057,16436	411009

h | Beispiel 2



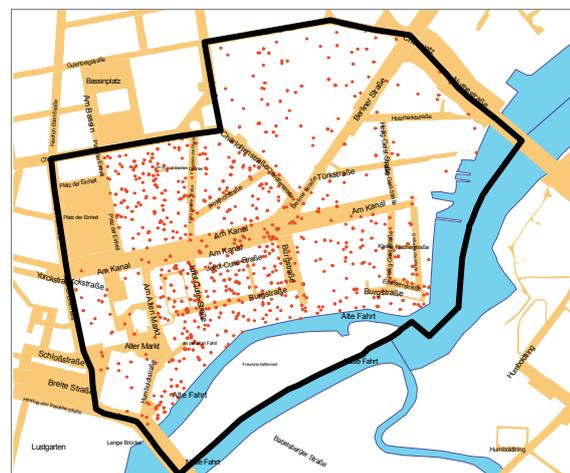
Beispiele thematischer Karten auf Blockebene

Punktdichtekarte – Einwohner mit Hauptwohnsitz im Statistischen Bezirk 414.

Ein Punkt = 5 Einwohner.

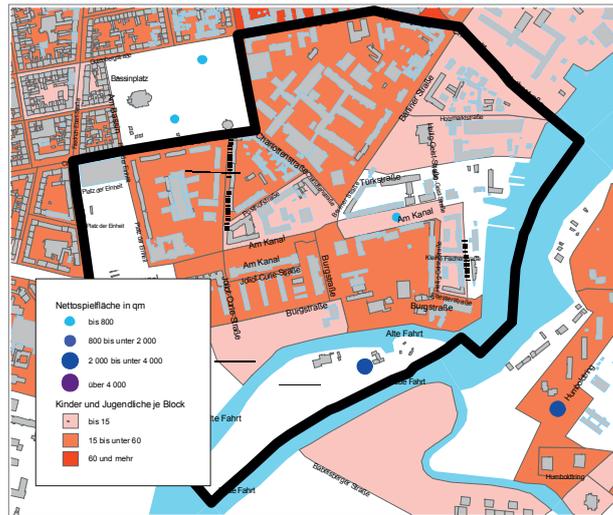
Im Beispiel der Abbildung i wird die Blockgeometrie mit den Daten aus dem Einwohnermelderegister verknüpft, wobei die Blockgrenzen im GIS auf „nicht sichtbar“ gesetzt werden.

i | Kartenausschnitt



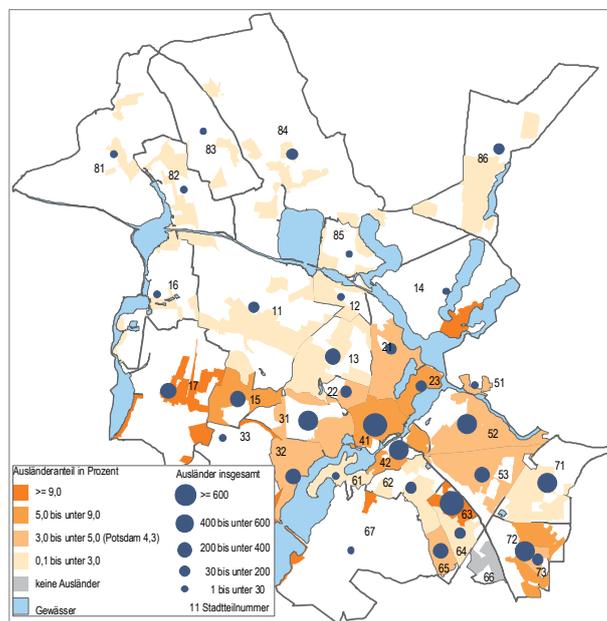
n | Kartenausschnitt

Anzahl der Kinder und Jugendlichen unter 18 Jahre nach Blöcken sowie Spielplatzstandorte und deren Nettospielfläche.



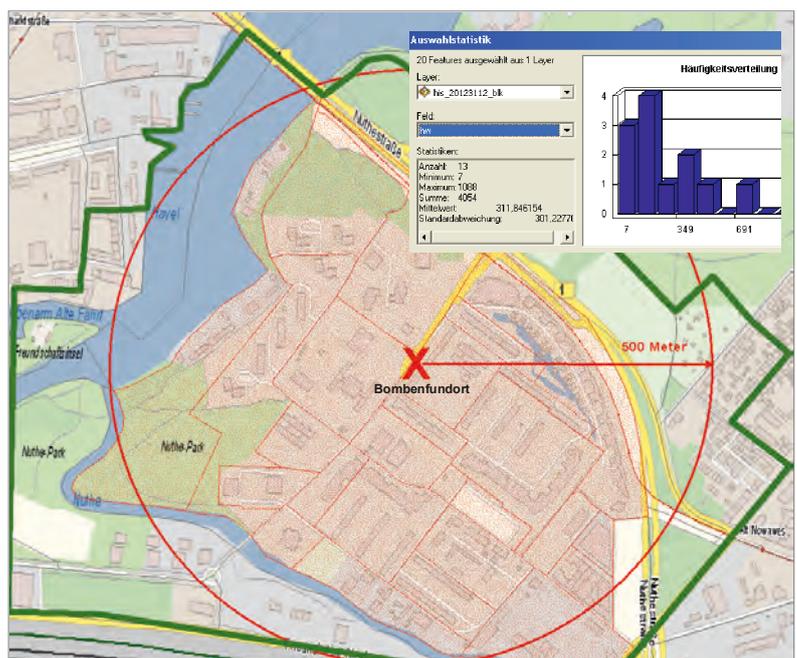
o | Kartenausschnitt

Ausländer und Ausländeranteil nach Stadtteilen



p | Kartenausschnitt

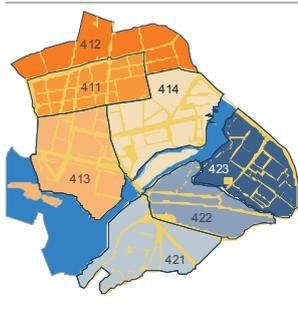
Ermittlung der Einwohneranzahl und Festlegung eines Sperr- und Räumkreises zur Evakuierung der Bevölkerung z. B. bei Bombenentschärfung nach einem Bombenfund. Der Evakuierungsradius beträgt 500 Meter.



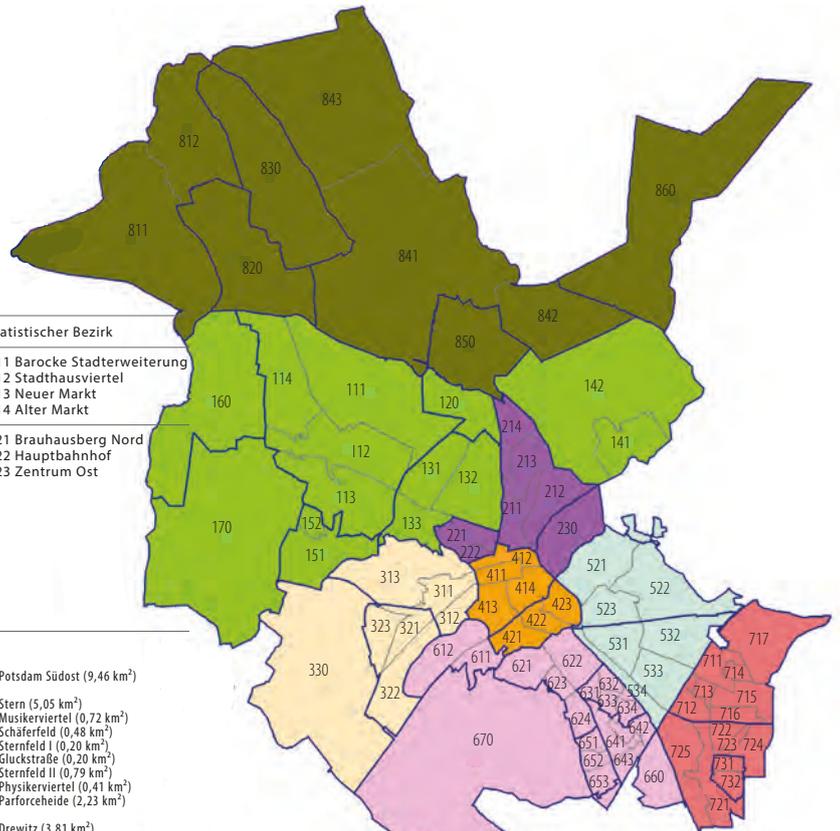
q | Karte

Kleinräumige Gliederung der Landeshauptstadt Potsdam

Hierarchische Gebietsgliederung der Landeshauptstadt Potsdam am Beispiel des Stadtbezirkes 4 - Innenstadt



Stadtbezirk	Stadtteil	Statistischer Bezirk
4 Innenstadt	41 Nördliche-Innenstadt	411 Barocke Stadterweiterung
		412 Stadthausviertel
		413 Neuer Markt 414 Alter Markt
42 Südliche-Innenstadt	421 Brauhausberg Nord	421 Brauhausberg Nord
		422 Hauptbahnhof
		423 Zentrum Ost



1 Potsdam Nord (51,17 km ²)	4 Innenstadt (4,17 km ²)	7 Potsdam Südost (9,46 km ²)
11 Bornim (13,94 km ²)	41 Nördliche Innenstadt (2,48 km ²)	71 Stern (5,05 km ²)
111 Bornim Nord (5,61 km ²)	411 Barocke Stadterweiterung (0,42 km ²)	711 Musikerviertel (0,72 km ²)
112 Bornim Ortslage (2,82 km ²)	412 Stadthausviertel (0,47 km ²)	712 Schäferfeld (0,48 km ²)
113 Katharinenholz (2,86 km ²)	413 Neuer Markt (0,84 km ²)	713 Sternfeld I (0,20 km ²)
114 Schlängenbruch (2,65 km ²)	414 Alter Markt (0,74 km ²)	714 Glückstraße (0,20 km ²)
12 Nedlitz (1,63 km ²)	42 Südliche Innenstadt (1,68 km ²)	715 Sternfeld II (0,79 km ²)
120 Nedlitz (1,63 km ²)	421 Brauhausberg Nord (0,65 km ²)	716 Physikerviertel (0,41 km ²)
13 Bornstedt (4,73 km ²)	422 Hauptbahnhof (0,44 km ²)	717 Parforceheide (2,23 km ²)
131 Das Kurze Feld (1,42 km ²)	423 Zentrum Ost (0,58 km ²)	72 Drewitz (3,81 km ²)
132 Bornstedter Feld (2,22 km ²)	5 Babelsberg (9,72 km ²)	721 Gewerbegebiet Drewitz (0,35 km ²)
133 Bornstedter Ortslage (1,08 km ²)	51 Klein Glienicke (0,28 km ²)	722 Drewitz I (0,15 km ²)
14 Sacrow (10,50 km ²)	510 Klein Glienicke (0,28 km ²)	724 Drewitz II (0,72 km ²)
141 Sacrow Ortslage (1,53 km ²)	52 Babelsberg Nord (5,09 km ²)	725 Drewitzer Nuthewiesen (1,51 km ²)
142 Königswald (8,97 km ²)	521 Park Babelsberg (1,73 km ²)	73 Kirchsteigfeld (0,59 km ²)
15 Eiche (2,35 km ²)	522 Neu Babelsberg (2,39 km ²)	731 Kirchsteigfeld Nord (0,24 km ²)
151 Eiche Ortslage (2,01 km ²)	523 Weberplatz (0,96 km ²)	732 Kirchsteigfeld Süd (0,35 km ²)
152 Altes Rad (0,33 km ²)	53 Babelsberg Süd (4,34 km ²)	8 Nördliche Ortsteile 66,31 km ²)
16 Grube (6,66 km ²)	531 Lutherplatz (0,99 km ²)	81 Uetz-Paaren (13,38 km ²)
160 Grube (6,66 km ²)	532 Medienstadt (1,55 km ²)	811 Uetz (9,23 km ²)
17 Golm (11,44 km ²)	533 Gewerbegebiet Babelsberg (1,13 km ²)	812 Paaren (4,18 km ²)
170 Golm (11,44 km ²)	534 Nuthewiesen Babelsberg (0,65 km ²)	82 Marquardt (5,22 km ²)
2 Nördliche Vorstädte (6,63 km ²)	6 Potsdam Süd (23,43 km ²)	820 Marquardt (5,22 km ²)
21 Nauener Vorstadt (3,94 km ²)	61 Templiner Vorstadt (2,01 km ²)	83 Satzkorn (6,64 km ²)
211 Kappellenberg (0,82 km ²)	611 An der Vorderkappe (0,60 km ²)	830 Satzkorn (6,64 km ²)
212 Neuer Garten (1,26 km ²)	612 Hermannswerder (1,41 km ²)	84 Fahrland (27,44 km ²)
213 Pfingstberg (1,91 km ²)	62 Teltower Vorstadt (2,81 km ²)	841 Fahrland Ortslage (13,86 km ²)
214 An der Roten Kaserne (0,84 km ²)	621 Brauhausberg Süd (0,99 km ²)	842 Krampnitz (3,32 km ²)
22 Jägervorstadt (0,96 km ²)	622 Teltower Vorstadt Nord (1,03 km ²)	843 Kartzow (10,25 km ²)
221 Am Ruinenberg (0,74 km ²)	623 Kunersdorfer Straße (0,29 km ²)	85 Neu Fahrland (3,39 km ²)
222 Am Weinberg (0,21 km ²)	624 Siedlung Eigenheim (0,48 km ²)	850 Neu Fahrland (3,39 km ²)
23 Berliner Vorstadt (1,73 km ²)	63 Schlaatz (1,00 km ²)	86 Groß Glienicke (10,21 km ²)
230 Berliner Vorstadt (1,73 km ²)	631 Gartenanlage am Schlaatz (0,22 km ²)	860 Groß Glienicke (10,21 km ²)
3 Westliche Vorstadt (16,59 km ²)	632 Schlaatz Nord (0,23 km ²)	
31 Brandenburger Vorstadt (4,06 km ²)	633 Schlaatz Zentrum (0,25 km ²)	
311 Zimmerplatz (0,75 km ²)	634 Schlaatz Süd (0,28 km ²)	
312 Kiewitt (0,53 km ²)	64 Waldstadt I (1,27 km ²)	
313 Park Sanssouci (2,78 km ²)	641 Waldstadt I Nord (0,55 km ²)	
32 Potsdam West (3,10 km ²)	642 Stadtrand siedlung (0,41 km ²)	
321 Potsdam West Zentrum (0,66 km ²)	643 Waldstadt I Süd (0,30 km ²)	
322 Luftschiffhafen (1,57 km ²)	65 Waldstadt II (0,87 km ²)	
323 Forststraße (0,86 km ²)	651 Waldstadt II Nord (0,24 km ²)	
33 Wildpark (9,43 km ²)	652 Waldstadt II Zentrum (0,25 km ²)	
330 Wildpark (9,43 km ²)	653 Waldstadt II Süd (0,37 km ²)	
	66 Industriegelände (1,17 km ²)	
	660 Industriegelände (1,17 km ²)	
	67 Forst Potsdam Süd (12,65 km ²)	
	670 Forst Potsdam Süd (12,65 km ²)	

Stadtbezirke

- 1-Potsdam Nord
- 2-Nördliche Vorstädte
- 3-Westliche Vorstadt
- 4-Innenstadt
- 5-Babelsberg
- 6-Potsdam Süd
- 7-Potsdam Südost
- 8-Nördliche Ortsteile

Stadtteil

- 11-Bornim
- 111-Bornim Nord

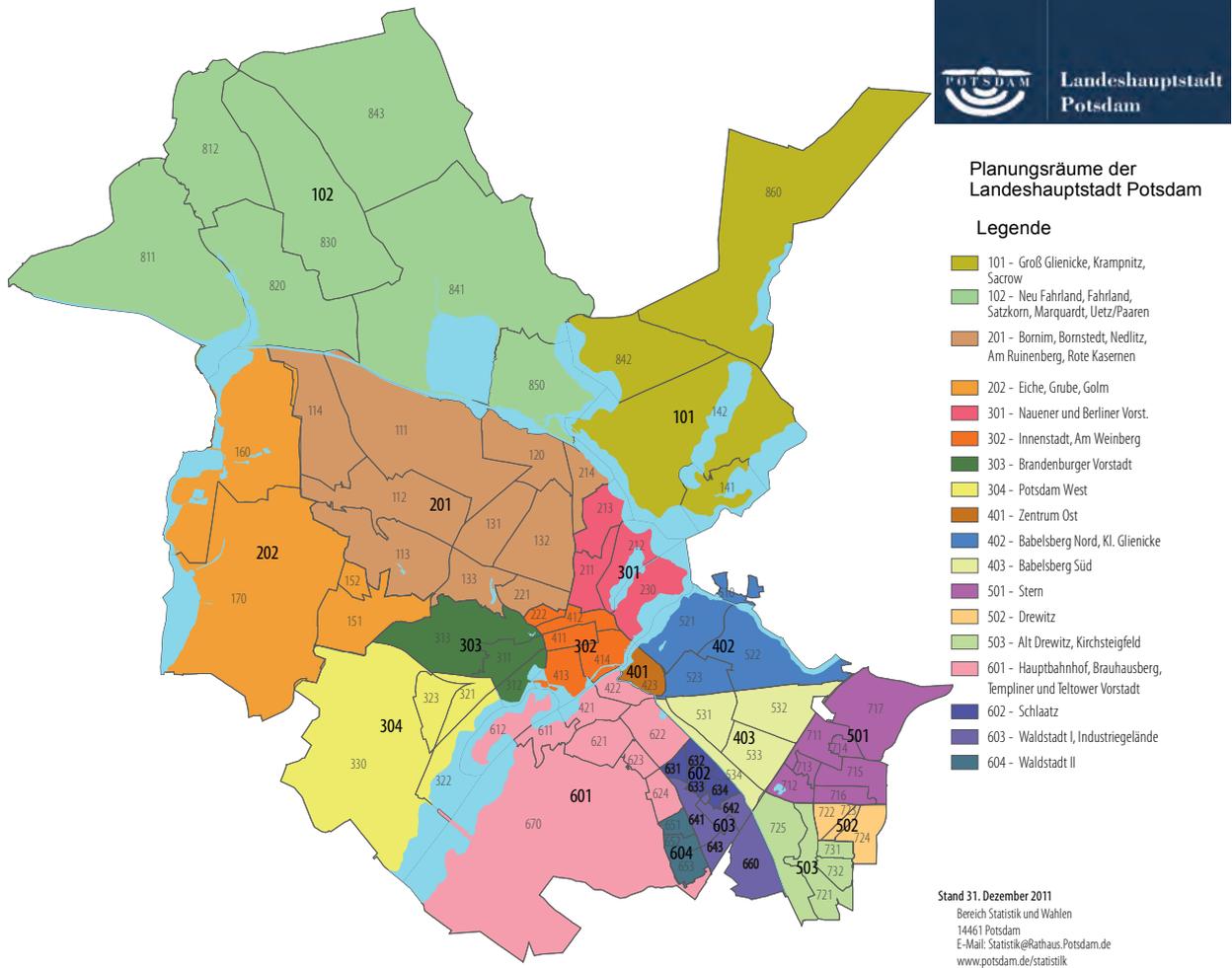
Statistischer Bezirk

- 111-Bornim Nord

Stand: 31.12.2013

r | Karte

Einteilung von Gebieten auf Grundlage der kleinräumigen Gliederung –
Bildung von Planungsräumen durch Aggregation von Statistischen Bezirken



Lutz Rittershaus ist seit 1993 Mitarbeiter im Bereich *Statistik und Wahlen* der Landeshauptstadt Potsdam. Er hat als zuständiger Sachbearbeiter die kleinräumige

Gliederung der Landeshauptstadt mit aufgebaut und ist verantwortlich für die Pflege des Raumbezugssystems und für alle Arbeiten im Geographischen Informationssystem.

Fachstatistische Anwendungen

▣ Datenschutzkonzept zur Nutzung von SOEPgeo im Forschungsdatenzentrum SOEP am DIW Berlin

von Jan Goebel und Bernd Pauer

In den empirischen Sozial- und Wirtschaftswissenschaften werden Forschungsdaten immer wichtiger, aber auch komplexer. In die Erstellung von Daten fließt nicht zuletzt ein großer zeitlicher und/oder personeller Aufwand. Daher sollten Daten für die Forschung möglichst breit nutzbar sein, um ihren maximalen Nutzen entfalten zu können. Empirische Daten tragen dazu bei, dass wissenschaftliche Aussagen auf einer transparenten Grundlage stehen. Sie anderen Wissenschaftlerinnen und Wissenschaftlern zur Verfügung zu stellen, gehört zur guten wissenschaftlichen Praxis. Aus diesen Gründen sollten empirische Daten, welche die Grundlage von wissenschaftlichen Analysen bilden, immer auch als Forschungsdaten angemessen archiviert und allen Forschenden für eine potenzielle Nachnutzung oder Reanalyse zur Verfügung stehen.

1. Einleitung

Bereits zu Beginn des vergangenen Jahrhunderts wurden die Beziehungen zwischen dem Menschen und dem Raum festgestellt (Werlen 2008). In diesem Zusammenhang wurden in der sozialökologischen Forschung der Chicagoer Schule die Interaktionen zwischen der räumlichen Stadtumwelt und dem Handeln des Individuums untersucht (Burgess 1925). Die auf diesen Ansätzen aufbauenden bzw. weiterführenden Kategorisierungen von Städten und Stadtregionen bestehen für Deutschland zunächst als Modell der Stadtregionen (Boustedt 1953). Weitere spezifisch für Deutschland entwickelte Typisierungen, wie die BIK-Regionen oder die laufende Raumb Beobachtung des Bundesinstituts für Bau-, Stadt- und Raumforschung (BBSR), beschreiben den Raum mangels detaillierter räumlicher Daten auf administrativen Raumeinheiten.

Dabei kommt der Bedeutung des Raumes für die Sozialforschung eine immer größer werdende Rolle zu, da durch raumbezogene Informationen die Beziehungen zwischen dem Menschen und dem Raum, also die Interaktion zwischen dem Individuum und dem räumlichen Kontext, besser untersucht werden können. Einerseits ist der Bedeutungszuwachs des Raumes auf die gestiegene Verfügbarkeit von räumlichen Daten zurückzuführen, was letztlich auch auf technische Entwicklungen in den Informationstechnologien wie den geographischen Informationssystemen (GIS), globalen Positionierungssystemen (GPS) und vor allem satellitengestützter Erdbeobachtung zurückzuführen ist. Die Relevanz raumbezogener Daten für die Sozialforschung wird als spatial turn für die Sozialwissenschaften bezeichnet (Goodchild 2007) und die Georeferenzierung von sozialwissenschaftlichen Daten als eine der großen positiven Herausforderungen der letzten Jahre interpretiert (RatSWD 2010).

Ein gänzlich neuer Zugang zur raumbezogenen Auswertung der Daten des SOEP ist über die vom Forschungsdatenzentrum des SOEP (FDZ SOEP) eingerichtete Infrastruktur „SOEPgeo“ möglich. An Gastarbeitsplätzen wird im FDZ SOEP die Möglichkeit bereitgestellt, Geo-Koordinaten für die einzelnen Haushalte zu nutzen. Das Konzept ermöglicht es, dass unter streng kontrollierten Datenschutzbedingungen Mikrodaten des SOEP zusammen mit geokodierten externen Daten für wissenschaftliche Zwecke ausgewertet werden können. Die folgenden Ausführungen beschreiben im Detail, wie hoch sensitive geokodierte Adressinformationen von befragten Haushalten Wissenschaftlerinnen und Wissenschaftlern zugänglich gemacht werden können, ohne dass es zu datenschutzrechtlichen Abstrichen kommt.

1.1 Das SOEP

Das Sozio-oekonomische Panel (SOEP) ist eine repräsentative Wiederholungsbefragung, die bereits seit nahezu 30 Jahren läuft (Wagner et al. 2007). Im Auftrag des DIW Berlin werden jedes Jahr in Deutschland über 20 000 Personen aus rund 11 000 Haushalten von TNS Infratest Sozialforschung befragt. Die Daten geben Auskunft zu Fragen über Einkommen, Erwerbstätigkeit, Bildung oder Gesundheit. Weil jedes Jahr die gleichen Personen befragt werden, können langfristige soziale und gesellschaftliche Trends besonders gut verfolgt werden. Eine detaillierte Beschreibung und weiterführende Dokumentation des Surveys findet sich auf der Webseite des FDZ SOEP (<http://www.diw.de/soepfdz>) oder auf dem Metadatenportal des SOEP (<http://data.soep.de/studies/2>).

1.2 Regionaldaten im SOEP

Das SOEP bietet vielfältige Möglichkeiten für regionalisierte Analysen. Die enthaltenen regionalen Informationen lassen sich in zwei Kategorien unterteilen:

1. Beschreibung des regionalen oder räumlichen Kontexts
 - a) Informationen gewonnen aus der Befragungssituation zur Nachbarschaft bzw. Wohnumgebung der Befragten,
 - b) Informationen zur Beschreibung der näheren Umgebung an Hand von verknüpften Indikatoren der microm GmbH,
 - c) Informationen zu Regionstypisierungen, in denen der befragte Haushalt lebt.
2. Offizieller Schlüssel der regionalen Einheiten zum jeweiligen Erhebungsjahr (administrative Einheiten, Postleitzahlen oder Koordinaten), um eine Verknüpfbarkeit mit externen Daten (z.B. aus der amtlichen Statistik) zu gewährleisten.

Bis auf die Informationen unter Punkt 1 a) sind alle diese Informationen datenschutzrechtlich sensibel und daher nicht in der Datenlieferung des Standard Scientific Use Files enthalten. Alle Informationen können aber für die unabhängige wissenschaftliche Forschung genutzt werden, allerdings unter zum Teil stark erhöhten Datenschutzerfordernissen, insbesondere im Zugang zu diesen Daten. Eine Übersicht über die im SOEP vorhandenen Regionalinformationen findet sich in Goebel 2013.

1.3 Koordinaten

Im Rahmen einer Kooperation mit der microm Micromarketing-Systeme und Consult GmbH war es seit 2004 möglich, den SOEP-Haushalten Indikatoren zu ihrer konkreten kleinräumigen Umgebung zuzuspielen. microm entwickelt seit 1992 mikrogeographische Daten, ursprünglich mit dem Ziel, Unternehmen bei der räumlichen Verortung ihrer Kunden- oder Zielgruppen zu unterstützen. Teile dieser Informationen sind auch für sozialwissenschaftliche Fragestellungen von Bedeutung, wie beispielsweise die Variable „microm Status“ oder der Anteil von Personen mit Migrationshintergrund. Die microm Datenbasis ermöglicht es flächendeckend, die ca. 41 Mill. bundesdeutschen Haushalte hinsichtlich verschiedenster Kriterien räumlich abzubilden. Die räumliche Verortung der SOEP-Haushalte wird direkt beim Erhebungsinstitut (TNS Infratest Sozialforschung), das als Einzige Namen und Klartext-Adresse vorhält, vorgenommen. Die verknüpften Daten werden von Infratest (ohne Klartext-Adressen und Namen) an das FDZ SOEP geliefert.

Die dabei zugespielten Daten enthalten auch die Koordinaten, die zur Verknüpfung der Datenquellen notwendig sind. Am FDZ SOEP werden diese Daten wiederum aufgespalten, sodass die inhaltlichen Indikatoren (z.B. microm status oder Anteil der Migrantinnen und Migranten) an einem normalen Gastarbeitsplatz Wissenschaftlerinnen und Wissenschaftlern zugänglich sind. Diese Zugangsmöglichkeiten beinhalten jedoch keinen Zugriff auf die Koordinaten. Damit Wissenschaftlerinnen und Wissenschaftlern auch das große Potenzial von geo-

grafisch referenzierten Koordinaten nutzen können, musste erst eine zusätzliche Infrastruktur aufgebaut werden. Einer der großen Vorteile bei der Integration der räumlichen Komponente über die Verortung der Befragungshaushalte ist, dass herkömmliche SOEP-Auswertungen immer auf administrative Raumeinheiten beschränkt sind, während über die Gebietsauswahl durch Geo-Koordinaten jeder beliebige Raum definiert werden und eine Unabhängigkeit von Gebietsstandsveränderungen in den Verwaltungseinheiten erzielt werden kann.

Zentraler Teil des entwickelten Datenschutzkonzeptes ist dabei, dass die Geo-Koordinaten der SOEP-Haushalte von den Erhebungsinformationen grundsätzlich getrennt gehalten werden. Keine Wissenschaftlerin und kein Wissenschaftler hat zu irgendeinem Zeitpunkt Zugriff auf Koordinate und Befragungsinformation. Die Erzeugung von inhaltlichen Indikatoren ist nur innerhalb eines speziell geschützten und abgeschotteten Systems möglich. Der Datennutzer hat daher keinen gleichzeitigen Zugriff auf die SOEP-Erhebungsdaten und die Geo-Koordinaten der SOEP-Haushalte. Die Ergebnisse werden nur streng anonymisiert präsentiert.

2. Allgemeine Beschreibung von SOEPgeo

SOEPgeo ermöglicht es den Nutzerinnen und Nutzern des SOEP, geokodierte Daten für wissenschaftliche Zwecke innerhalb des DIW Berlin auszuwerten. Bisher konnten Wissenschaftlerinnen und Wissenschaftler diese Informationen nicht nutzen. Mit SOEPgeo ist es nunmehr möglich, dies in einer speziellen Arbeitsumgebung zu realisieren. Wissenschaftlerinnen und Wissenschaftler müssen eine Datenschutzverpflichtung unterzeichnen und der Zugriff auf die Daten wird vollständig protokolliert.

Grundlage des Konzepts ist es, dass während allen von Nutzerinnen und Nutzern ausgeführten Analysen die Geo-Koordinate der SOEP-Haushalte von den eigentlichen Erhebungsinformationen getrennt gehalten wird. Zur Erzeugung von inhaltlichen Indikatoren innerhalb eines Geo-Informationen Systems (GIS) oder innerhalb des Statistik Pakets R sind nur die Koordinaten ohne weitere Informationen über den Haushalt oder Personen in diesem Haushalt notwendig. Ein Zuspätspielen der in einem GIS erzeugten Indikatoren erfolgt durch Beschäftigte des FDZ ohne die Möglichkeit eines Zugriffs durch die Nutzerinnen und Nutzer. Im Folgenden wird das Konzept im Detail dargestellt.

3. Grundlagen des Datenschutzkonzepts

3.1 Juristische Grundlagen

Das Datenschutzkonzept beruht auf dem Konzept der „faktischen Anonymisierung“ von Daten, das auch von der amtlichen Statistik angewandt wird. Zugänglich gemacht werden „Scientific Use Files“ (SUF). Dabei sind Daten theoretisch deanonymisierbar (im Gegensatz zu absolut anonymisierten Mikrodaten, die als Public Use Files für jedermann verfügbar sind). Die Deanonymisierung würde einen unverhältnismäßigen Aufwand bedeuten und

ist mit juristischen Sanktionen belegt. Das heißt, dass die Auswertung nur auf Basis eines speziellen Datennutzungsvertrages, der ausschließlich mit vertrauenswürdigen Forschungseinrichtungen abgeschlossen wird, möglich ist und die erhobenen Daten faktisch anonymisiert sind (insbesondere keine Namen oder Klartexte enthalten). Das Deanonymisieren durch indirekte Verfahren wird deutlich erschwert und ist mit juristischen Sanktionen belegt.

Personenbezogene Daten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlichen Person (§3 Abs.1 BDSG¹). Keine personenbezogenen Daten sind zusammenfassende Angaben oder aggregierte Daten, da sie keiner Einzelperson zugeordnet werden können. Bestimmt sind die Daten, wenn sie sich direkt auf eine bestimmte Person beziehen, d.h. einen unmittelbaren Rückschluss auf die Identität einer Person, in erster Linie anhand ihres Namens und Geburtsdatums, zulassen. Bestimmbar sind sie, wenn sie eine Person nicht als solche identifiziert, jedoch mithilfe anderer Informationen und Zusatzwissen identifizierbar machen, d.h. die Möglichkeit besteht, die Identität einer Person festzustellen.

Wie der Interministerielle Ausschusses für Geoinformationswesen treffend feststellt, besteht das Problem bei georeferenzierten Daten darin, „dass sich die Abgrenzung, wann es sich um ein – datenschutzrechtlich nicht geschütztes – bloßes Sachdatum handelt und in welchen Fällen ein Personenbezug vorliegt, für den Einzelfall als sehr schwierig erweisen kann. Damit ein Geodatum überhaupt Personenbezug aufweisen kann, muss es sich auf eine bestimmte oder bestimmbar Person beziehen.“²

Da es sich bei den Geo-Koordinaten der Haushalte um sensitive Daten handelt, die in Verbindung mit den Befragungsdaten eine Deanonymisierung von teilnehmenden Personen enorm erleichtern können, werden die üblichen Maßnahmen für Scientific Use Files durch SOEPgeo deutlich verstärkt. Durch SOEPgeo wird ausgeschlossen, dass ein Deanonymisierungsversuch der Befragungsdaten auf Basis der Geo-Koordinaten durchgeführt werden kann.

Im Zentrum des technischen Schutzes steht, dass im Rahmen von SOEPgeo keinerlei Mikrodaten in Verbindung mit Geo-Koordinaten genutzt werden können. Eine Verknüpfung und damit missbräuchliche Nutzung der Mikrodaten, um die einzelnen Koordinaten einer bestimmbarer Person zuzuordnen, wird durch software- und hardware-technische Vorkehrungen für Benutzer ausgeschlossen. Das heißt, die Deanonymisierung ist nur mit unverhältnismäßigem Aufwand möglich. Die Deanonymisierung von ausgewählten Einzelfällen bringt auch keinen erkennbaren Nutzen und Forscherinnen und Forscher würden ihre Karriere ruinieren, wenn durch die vollständige Protokollierung des Datenzugriffs Deanonymisierung bzw. Deanonymisierungsversuche aufgedeckt werden.

Alle Zugriffe werden vollständig protokolliert. Das heißt – im Gegensatz zur Distribution von faktisch anonymisierten Scientific Use Files – ist eine voll-

ständige Kontrolle der Analysen jederzeit möglich, und es werden keine Mikrodaten weitergegeben. Ein weiterer wesentlicher Bestandteil des Datenschutzes von SOEPgeo ist, dass der Zugriff nur während eines Aufenthalts am DIW Berlin möglich ist.

3.2 Technische Grundlagen

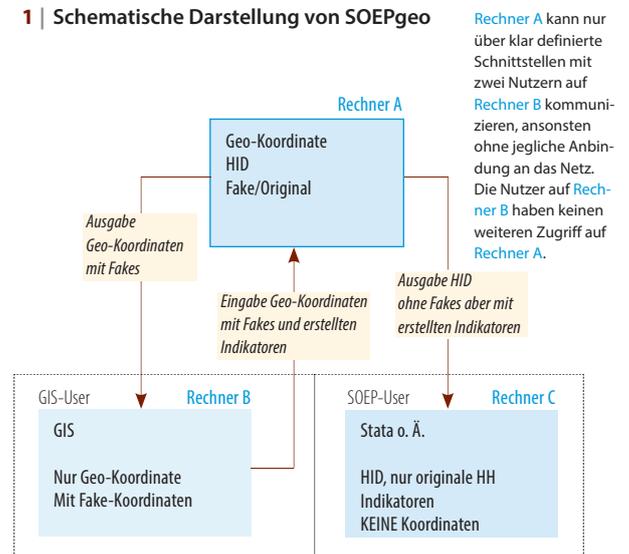
Zur Umsetzung des Datenschutzes von SOEPgeo wurden im DIW Berlin für das FDZ SOEP drei virtuelle Server³ speziell konfiguriert (ein Schema des Setups findet sich in Abbildung 1). Auf dem ersten Rechner (Rechner A) ist als Einziges die Zuordnung der SOEP-Haushalts-ID zu den jeweiligen Geo-Koordinaten gespeichert. Dieser Rechner ist dafür verantwortlich, entweder die Geo-Koordinaten (ohne Haushalts-ID Zuordnung) oder die SOEP-Haushalts-IDs mit den neu erstellten Indikatoren (ohne die entsprechenden Geo-Koordinaten) über vordefinierte Schnittstellen auszugeben. Der zweite Rechner (Rechner B) wird zur Analyse der Geo-Koordinaten genutzt und der dritte Rechner (Rechner C) zur Analyse der SOEP-Befragungsdaten inklusive der aus dem Raumbezug gewonnenen zusätzlichen Indikatoren (ohne Geo-Koordinaten). Die Analyse der Geo-Koordinaten und die Analyse der eigentlichen SOEP-Daten erfolgt dabei auf den getrennten Systemen im jeweiligen Datenkontext in den Rollen als GIS-Nutzer oder SOEP-Nutzer. Beide Rollen können sich nicht überschneiden und es ist auch nicht möglich, Daten direkt zu übermitteln.

1 Bundesdatenschutzgesetz (BDSG) in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), zuletzt geändert durch Artikel 1 des Gesetzes vom 14. August 2009 (BGBl. I S. 2814).

2 Siehe Behördenleitfaden zum Datenschutz bei Geodaten und -diensten, http://www.imagi.de/SharedDocs/Downloads/IMAGI/DE/Imagi/behoerdenleitfaden.pdf?__blob=publicationFile, Zugriff 30.07.2014.

3 Auf dem Hostsystem laufen nur diese drei virtuellen Maschinen.

1 | Schematische Darstellung von SOEPgeo



3.2.1 Hostsystem der SOEPgeo Maschinen

Die Hardware und der darauf installierte Hypervisor für die drei virtuellen Maschinen der SOEPgeo Installation steht hinter der allgemeinen Perimeter Firewall des DIW Berlin und hat nur die zum Betrieb notwendigen Dienste aktiviert, in diesem Fall die Management Software zur Verwaltung der drei Maschinen.

Der Zutritt ist durch die Installation im Rechenzentrum des DIW Berlin in einem abgeschlossenen Datenschutzzschränk gesichert.

Die drei SOEPgeo Systeme haben namentlich bekannte Systemadministratoren, die nach einer Anmeldung explizit in die Administratorenrolle wechseln müssen, was, wie der Zugang, ebenfalls protokolliert wird.

3.2.2 Rechner A: Kombination von Haushalts-ID und Geo-Koordinate

Der eingesetzte Rechner für die Kombination der Geo-Koordinaten und der SOEP-Haushalts-ID ist nur von Rechner B oder Rechner C über eine interne Netzwerkschnittstelle, die auch nicht aus dem DIW-internen Netz angesprochen werden kann, erreichbar. Er kennt keine sogenannte „Default Route“ und ist somit nicht in der Lage, mit anderen Rechnern zu kommunizieren.

Zugriff auf den Rechner A ist nur für die Systemadministratoren mit dem Secure Shell Public-Key Verfahren möglich. Da auf diesem Rechner die Verbindung der Koordinate zu den Haushalts-IDs liegt, gelten für diesen Rechner besondere Schutzmaßnahmen.

Die dortige Datenbank beinhaltet den Umsteigeschlüssel von der Geo-Koordinate zur SOEP-Haushalts-ID und die Information, ob für diese Geo-Koordinate überhaupt ein SOEP-Haushalt im SOEP befragt wurde, es sich also um eine „wahre“ Koordinate oder um eine gefälschte Koordinate handelt. Um die eindeutige Zuordnung der Koordinate zu einem Haus mit einem SOEP-Haushalt aufzulösen, werden zusätzliche Geo-Koordinaten in die Datenbank mit aufgenommen (im Verhältnis 1:1). Da das Ziehungsdesign des SOEP zu einer Klumpung von Geo-Koordinaten führt, wurden die gefälschten Geo-Koordinaten nur im Umkreis der im SOEP bereits vertretenen Haushalte hinzugefügt, um auch auf diesem Weg eine Unterscheidung von gefälschter und richtiger Koordinate nicht zu ermöglichen.

Werden von einem GIS-Nutzer Indikatoren generiert, so werden anschließend durch Beschäftigte des FDZ SOEP auf Rechner A die Geo-Koordinaten entfernt und die SOEP-Haushalts-ID angespielt. Die Indikatoren mit Haushaltsnummern werden dann auf den jeweiligen Nutzerbereich des Rechners C geschrieben. Auf diesen Bereich können die jeweils definierten SOEP-Nutzer des Rechners C zugreifen.

3.2.3 Rechner B + C: Analyserechner

Zugriff auf die Rechner B + C ist nur mit dem Secure Shell Public-Key Verfahren innerhalb des DIW Berlin möglich. Auf den Analyserechnern werden (neben dem Systemadministrator) zwei getrennte Nutzerkreise eingerichtet: die GIS-Nutzer auf Rechner B und die SOEP-Nutzer auf Rechner C.

Der GIS-Nutzer darf nur auf die Datenbestände der Geo-Koordinaten zugreifen und kann nicht unterscheiden, hinter welcher der Koordinaten sich eine reale SOEP-Adresse befindet und hinter welcher nicht. Die Geo-Koordinaten werden ohne Verknüpfung zur SOEP-Haushalts-ID auf diesen Rechner durch die Systemadministration abgespeichert und einmal jährlich aktualisiert. Als einziges Attribut enthalten die Geo-Koordinaten ein Erhebungsjahr, um eine zeitliche Zuordnung zu ermöglichen. Die gefälschten Koordinatenpunkte bilden dabei ebenso reale Mobilitätsmuster ab wie die tatsächlich beobachteten Haushalte, das heißt, auch durch den Einbezug der temporalen Information können Wissenschaftlerinnen und Wissenschaftler daher eine echte und eine gefälschte Koordinate nicht unterscheiden.

Der GIS-Nutzer kann nicht auf Rechner A zugreifen und daher nicht die Verknüpfung von Koordinate und Haushalts-ID lesen. Der Datenfluss geht hier ausschließlich über das Konto der Systemadministration in Richtung Rechner A und beinhaltet die innerhalb des GIS-Systems erstellten Indikatoren auf Geo-Koordinaten-Ebene. Als GIS-Software kommen die folgenden Open Source Software Produkte zum Einsatz:

- R (<http://www.r-project.org/>),
- QGIS (<http://www.qgis.org/>),
- GRASS (<http://grass.osgeo.org/>).

In der SOEP-Nutzer-Rolle darf nur auf die Datenbestände des SOEP, ohne Geo-Koordinaten, aber mit den im GIS gebildeten Indikatoren, zugegriffen werden. Dabei kann ebenfalls nicht auf die Datenbank des Rechners A zugegriffen werden. Die Rechte beschränken sich auf das Lesen der Indikatoren in Verbindung mit den SOEP-Haushalts-IDs, wie sie von Rechner A rausgeschrieben wurden. Als Statistik-Software kommt zum einen Stata (<http://www.stata.com>) und zum anderen die Open Source Software R (<http://www.r-project.org/>) zum Einsatz.

3.3 Zugangskontrolle

Jedes Einloggen auf den Rechner A und B und jeder Zugriff auf die Postgres Datenbank des Rechners A wird protokolliert und mit Datum und Nutzerkennung von der SOEP-Gruppe elektronisch gespeichert. Dadurch kann jegliche Fehlverwendung auch im Nachhinein erkannt und individuell zugeordnet werden.

Möchten Wissenschaftlerinnen und Wissenschaftler für ihre SOEP-basierte Forschung auf Koordinateninformationen zugreifen und dafür SOEPgeo nutzen, so muss der Vertragsnehmer neben einem gültigen SOEP-Standardweitergabevertrag auch einen Sondernutzungsvertrag unterzeichnen. Erst aufgrund dieser geprüften Informationen können Wissenschaftlerinnen und Wissenschaftler als SOEPgeo-Datennutzer am DIW Berlin arbeiten.

3.4 Zugriffskontrolle

Die Zugriffskontrolle auf Rechner B+C erlaubt den jeweiligen Nutzerinnen und Nutzern den Lesezugriff auf die bereitgestellten Daten. Gleichzeitig sind die Ergebnisse und Verfahren anderer Nutzerinnen und Nutzer nicht von vornherein zugreifbar.

Arbeits- oder Projektgruppen mit mehreren Nutzerinnen und Nutzern können unter einer gemeinsamen Gruppe zusammengefasst werden. Feiner granulierte Zugriffsrechte ließen sich noch darüber hinaus abbilden, wurden aber bislang nicht benötigt.

4. Nutzungsbeispiele

Im Folgenden soll anhand ausgewählter Beispiele dargestellt werden, wie Forscherinnen und Forscher bisher den neuen Infrastruktur Service SOEPgeo nutzen und welche Forschungsfragen dadurch analysiert werden konnten.

Bauerschuster, Falck und Woessmann nutzten den zeitlich und räumlich ungleichmäßigen Ausbau und damit den Zugang zu Breitband-Internet in Deutschland, um den Zusammenhang von Internet und sozialem Kapital zu untersuchen. Dazu nutzten sie die Daten der Telekom, um über die Distanz zum nächsten OPAL-fähigen Zugangspunkt festzustellen, wann für welche Haushalte Breitband-Internet zur Verfügung stand. Ihre Ergebnisse zeigen, dass kein negativer, sondern eher ein positiver Effekt vom Internetzugang auf soziales Kapital ausgeht (Bauerschuster, Falck und Woessmann 2014).

Verschiedene Forscherinnen und Forscher nutzen die Geo-Koordinaten, um den Einfluss von Naturkatastrophen auf das Verhalten oder Befinden der Menschen zu untersuchen. Goebel, Kerkel u. a. (2013) untersuchten den Einfluss der Katastrophe von Fukushima auf die Umweltsorgen in Deutschland und kontrollierten dabei die Distanz zum nächsten Atomkraftwerk in Deutschland, Frankreich oder Tschechien. Berlemann, Steinhardt und Tutt 2014 stellten auf einer Präsentation anlässlich des „Spring Meeting of Young Economists“ (2014) ihre Arbeiten zum Einfluss des Elbe-Hochwassers 2002 auf das Spar- und Umzugsverhalten vor. Ebenfalls die Daten aus Satellitenaufnahmen zu einer Hochwasser-Katastrophe (allerdings die des Jahres 2013) nutzten Avdeenko, Kregel und Martinoty 2014, um den Zusammenhang mit Umweltsorgen und Einstellungen zum Klimawandel zu untersuchen.

Die allgemeineren Auswirkungen der unmittelbaren Umgebung zeigen unter anderem Kolbe, Kregel und Wüstemann (2014). Sie nutzten flächendeckende Daten aus dem European Urban Atlas für deutsche Großstädte, um relevante Grünflächen zu identifizieren. Ihre Analyse beleuchtet damit die quantitative Bedeutung von Grünflächen für das Wohlbefinden der Stadtbewohner.

Kregel und Zerrahn (2014) untersuchten den Einfluss von Windenergieanlagen auf die Wohnzufriedenheit, indem sie die Distanz zum nächstgelegenen Windrad im Zeitraum 2000 bis 2012 errechneten. Die Kombination von Paneldaten und flächendeckender Geoinformationen machte es hier möglich, auch kausale Erklärungsmodelle zu schätzen. Ebenfalls flächendeckende Daten, allerdings zur Luftverschmutzung, nutzten Voigtländer u. a. 2011. Mit Hilfe von Daten zur Luftbelastung aus EURAD-IM untersuchten sie den Einfluss auf die im SOEP erfragten subjektiven Gesundheitsmaße.

Eine wichtige Anwendung von Geo-Koordinaten in der sozialwissenschaftlichen Forschung ist die Möglichkeit, Varianz unterhalb der Gemeindeebene zu nutzen, also den Effekt von Stadt je nach Lage innerhalb der Stadt differenziert schätzen zu können. Goebel und Wurm (2010) nutzten den Abstand zu einer verallgemeinerten Stadtgrenze, die sich nicht nach der offiziellen Gemeindegrenze richtet, sondern nach der faktischen Bebauungsstruktur. Dabei verschnitten sie Daten von CORINE Land Cover mit der Lage der SOEP-Haushalte und berechneten den Abstand zum Zentrum und zur Stadtgrenze. Sie zeigten dabei, dass das Armutrisiko am Stadtrand am niedrigsten ist und in beide Richtungen (zum Stadtzentrum und in Richtung ländlicher Raum) zunimmt.

5. Fazit

Am FDZ SOEP ist seit Beginn der Pilotphase im Jahr 2010 die Infrastruktur zur Nutzung von GEO-Koordinaten in Verbindung mit sozialwissenschaftlichen Befragungsdaten für Wissenschaftlerinnen und Wissenschaftler zugänglich. Die bisherigen Erfahrungen sind durchweg positiv.

Insgesamt nimmt die wissenschaftliche Community den Service sehr gut an, obwohl die Arbeitsbedingungen schwieriger sind als am eigenen Arbeitsplatzrechner und teilweise das Arbeiten durch die technische Trennung mehr Zeit beansprucht. Die Einschätzung, dass der Zugang zu den Originaldaten für wissenschaftliche Analysen wichtiger ist als der Komfort bei den Analysen, hat sich als richtig herausgestellt. Trotz der erhöhten Anforderungen an die Wissenschaftlerinnen und Wissenschaftler wird der Service von SOEP-Nutzern honoriert.

Insgesamt gesehen hält sich auch der Arbeitsaufwand für die Prüfung der Ergebnisse im Rahmen der normalen Outputprüfungen an einem Gastarbeitsplatz in einem Forschungsdatenzentrum in Grenzen.

Ein Problem, das sich in Zukunft stellt, ist damit verbunden, dass immer mehr geokodierte flächendeckende Daten und auch neue Schätzmethode verfügbar werden. Diese an sich sehr erfreuliche Entwicklung hat jedoch den Nachteil, dass die Datenmengen, die die Server verarbeiten müssen, steigen. Die Einbindung von Abstandsmatrizen, also berechnete Distanzen von jeder beobachteten Einheit zu jeder anderen, stellt im Moment noch sehr hohe Anforderungen an die Rechner und ist auch im gegenwärtigen Setting nur schwer umzusetzen.

Eine weitere Herausforderung wird es sein, den Zugang zu SOEPgeo auch von Gastarbeitsplätzen in anderen Forschungsdatenzentren per Remote-Zugriff zu ermöglichen, wie es bereits für Regionalinformationen bis zu den Kreisen möglich ist (das sogenannte FDZ-im-FDZ-Modell, siehe auch Heining und Bender 2012).



Dr. Jan Goebel ist stellvertretender Leiter des Sozio-ökonomischen Panel (SOEP). Sein Arbeitsbereich: Data-Operation und Forschungsdatenzentrum (SOEP FDZ), die Arbeitsschwerpunkte sind: Datenweitergabe SOEP/SOEPinfo, Regionaldaten, Armuts- und Ungleichheitsforschung, Einkommensverteilung und -dynamik, Item-Non-Response und Imputation.

Bernd Pauer ist Mitarbeiter der Abteilung *Forschungsinfrastruktur* im DIW Berlin, IT-Sicherheit, er ist stellvertretender Datenschutzbeauftragter des DIW Berlin.

Literatur

- Avdeenko, Alexandra, Christian Krekel und Laurine Martinoty (2014). "Natural Disasters and Environmental Concerns: The Case of the 2013 Flood in Germany". Unveröffentlichter Vortrag SOEP User Conference 2014. Berlin. url: http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.465838.de/107_soep2014abstract_avdeenko_et.al.pdf (besucht am 06. 08. 2014).
- Bauernschuster, Stefan, Oliver Falck und Ludger Woessmann (2014). "Surfing alone? The internet and social capital: Evidence from an unforeseeable technological mistake". In: *Journal of Public Economics* 117, S. 73–89. issn: 0047-2727. doi: <http://dx.doi.org/10.1016/j.jpubeco.2014.05.007>. url: <http://www.sciencedirect.com/science/article/pii/S0047272714001145>
- Berlemann, Michael, Max Steinhardt und Jascha Tutt (2014). Natural disasters and the savings rate. Some micro-evidence from a natural experiment. Unveröffentlichter Vortrag "Spring Meeting of Young Economists".
- Boustedt, Olaf (1953). „Die Stadtregion. Ein Beitrag zur Abgrenzung städtischer Agglomerationen“. In: *Allgemeines Statistisches Archiv* 37, S. 13–26.
- Burgess, Ernest W. (1925). "The growth of the city: an introduction to a research project". In: *The City*. Hrsg. von Robert E. Park, Ernest W. Burgess und Roderick D. McKenzie. Univ. of Chicago Press.
- Goebel, Jan (2013). „Regionalisierungsmöglichkeiten des Sozio-ökonomischen Panels (SOEP)“. In: *Regionale Standards*, Hrsg. von der Arbeitsgruppe Regionale Standards. 2. vollständig überarbeitete und erweiterte Auflage. GESIS.
- Goebel, Jan, Christian Krekel u. a. (2013). Natural Disaster, Policy Action, and Mental WellBeing: The Case of Fukushima. SOEPpaper 599. Berlin: DIW/SOEP. url: http://www.diw.de/documents/publikationen/73/diw_01.c.430617.de/diw_sp0599.pdf
- Goebel, Jan und Michael Wurm (2010). „Räumliche Unterschiede im Armutsrisiko in Ost- und Westdeutschland“. In: *Leben in Ost- und Westdeutschland: Eine sozialwissenschaftliche Bilanz der deutschen Einheit 1990-2010*. Hrsg. von Peter Krause und Ilona Ostner. Frankfurt/Main: Campus. url: http://www.diw.de/documents/publikationen/73/diw_01.c.362278.de/diw_sp0321.pdf
- Goodchild, Michael F. (2007). "The Morris Hansen Lecture 2006 Statistical Perspectives on Spatial Social Science". In: *Journal of Official Statistics* 23.3, S. 269–283. url: <http://www.geog.ucsb.edu/~good/papers/436.pdf>
- Heining, Jörg und Stefan Bender (2012): „Technische und organisatorische Maßnahmen für den Fernzugriff auf die Mikrodaten des Forschungsdatenzentrums der Bundesagentur für Arbeit“. FDZ-Methodenreport, 08/2012, Nürnberg.
- Kolbe, Jens, Christian Krekel und Henry Wüstemann (5.–9. November 2014). "The Greener, the Happier? The Effects of Urban Green and Abandoned Areas on Residential Well-Being". Unveröffentlichter Vortrag, 67th Annual Meeting of the Gerontological Society of America. Washington, USA.
- Krekel, Christian und Alexander Zerrahn (5.–6. Juni 2014). "Sowing the Wind and Reaping the Whirlwind? The Effect of Wind Turbines on Residential Well-Being". Unveröffentlichter Vortrag, Energy, Environment and Well-Being : Workshop. Universität Oldenburg. Delmenhorst.
- (RatSWD), German Data Forum, Hrsg. (2010). *Building on Progress. Expanding the Research Infrastructure for the Social, Economic and Behavioral Sciences*. Budrich UniPress. url: <http://www.ratswd.de/publikationen/building-on-progress>
- Voigtländer, Sven u. a. (2011). Using geographically referenced data on environmental exposures for public health research: a feasibility study based on the German Socio-Economic Panel Study (SOEP). SOEPpaper 386. Berlin: DIW Berlin. url: http://www.diw.de/documents/publikationen/73/diw_01.c.375907.de/diw_sp0386.pdf
- Wagner, Gert G., Joachim R. Frick und Jürgen Schupp (2007). "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements". In: *Schmollers Jahrbuch (Journal of Applied Social Science Studies)* 127.1, S. 139–169.
- Werlen, Benno (2008). *Sozialgeographie*. 3. Aufl. Bern: Hauptverlag.

Geheimhaltung

Informationenverlust durch Anonymisierung am Beispiel der Berliner Einwohnerregisterdaten

Anpassungsgüte nichtparametrischer Dichteschätzer bei diskretisierten Daten

von Sebastian Schmon

„Geodaten haben wachsende Bedeutung“, schreibt der Rat für Sozial- und Wirtschaftsdaten (2012) im Abschlussbericht der Arbeitsgruppe „Georeferenzierung von Daten“. Das Ergebnis der Arbeitsgruppe gibt einen wichtigen Trend in der statistischen Analyse von Daten an, wo georeferenzierte Daten immer mehr Beobachtung finden. Als Schlussfolgerung wird von der Arbeitsgruppe unter anderem festgehalten, dass „eine datenschutzrechtliche Prüfung erforderlich [sei], ab welchen Auflösungsschwellen eine ‚Genauigkeit‘ bei personenbezogenen Geodaten keine Schutzbedürftigkeit mehr hervorruft.“¹

Einführung

Während von der datenschutzrechtlichen Seite eine Vergrößerung der Auflösungsgenauigkeit wünschenswert ist, können daraus für Anwenderinnen und Anwender von georeferenzierten Daten Probleme resultieren. Eine Anonymisierung von Daten geht mit einem Informationsverlust einher, der Auswirkungen auf die spätere Schätzung von statistischen Zusammenhängen haben kann. Beispielsweise müssen Erwartungstreue und Konsistenz von Schätzfunktionen nicht mehr unbedingt erfüllt sein.

Eine relativ einfache und intuitive Methode, um die räumliche Verteilung eines statistischen Merkmals darzustellen, ist die Visualisierung mittels einer (bivariaten) Dichtefunktion, über Höhenlinien oder durch unterschiedliche Farbintensität (Abbildung a). Häufig genutzte Methoden, um die Dichtefunktion aus den Daten zu schätzen, sind Histogramm- oder Kerndichteschätzer.

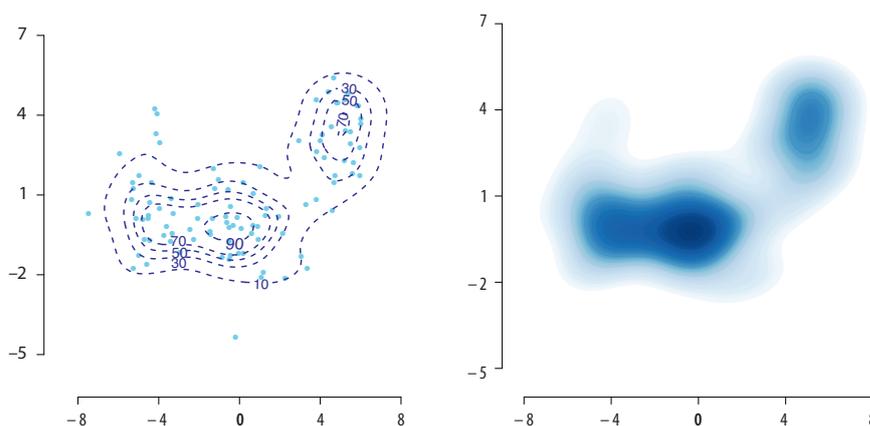
Dieser Beitrag soll der Frage nachgehen, wie stark sich die Anonymisierung von Geodaten auf die Interpretierbarkeit statistischer Dichteschätzer auswirkt. Anonymisierung wird in diesem Zusammenhang als eine Diskretisierung der Daten verstanden, also als eine Minderung der Auflösungsgenauigkeit der Koordinaten.

Die Ausführungen unterteilen sich in zwei Abschnitte. Eine Simulation mit Zufallsstichproben soll zeigen, wie sich die Diskretisierung der Daten auf die Struktur der Schätzung auswirkt. Gleichzeitig bietet sie die Möglichkeit, die geschätzte mit der bekannten, vor Experimentbeginn festgelegten, wahren Dichte zu vergleichen. Im zweiten Abschnitt wird anhand des Berliner Einwohnerregisters (Stand 31.12.2012) die praktische Relevanz der vorgestellten Problematik beleuchtet.

Dichteschätzung

Die wohl bekannteste Methode zur Schätzung einer Dichtefunktion ist das Histogramm. Es unterteilt die Stichprobe in Intervalle (univariater Fall) oder Rechtecke (bivariater Fall). Anschließend wird die Dichte aus der relativen Häufigkeit der untersuchten Merkmalsausprägungen geschätzt.

a | Kerndichteschätzung aus einer gemischten Normalverteilung, visualisiert durch ein Streudiagramm mit Höhenlinien (links) und durch „gefärbte Höhenlinien“ (rechts; ein dunklerer Farbton stellt eine höhere Dichte dar)



¹ Rat für Sozial- und Wirtschaftsdaten (2012): Georeferenzierung von Daten. Situation und Zukunft der Geodatenlandschaft in Deutschland. SCIVERO Verlag, Berlin.

Nachteile des Histogramm-Schätzers sind unter anderem seine Abhängigkeit vom gewählten Ursprungspunkt x_0 ; ebenso lässt die konstante Schätzung über das Intervall oder Rechteck keine „glatten“ Funktionen als Ergebnis zu.

Der Kerndichteschätzer lässt sich als Verallgemeinerung des Histogramms auffassen. Er „umgeht“ die genannten Problematiken durch seine Konstruktion, sodass sich auch „glatte“ Funktionen schätzen lassen. Die Methode des Kerndichteschätzers folgt der Idee, dass sich Beobachtungen vermehrt dort häufen, wo Realisationen wahrscheinlicher sind. Ein natürlicher Schätzer für eine Dichtefunktion ergibt sich nun aus der umgekehrten Überlegung, dass die theoretische Dichte in der unmittelbaren Umgebung der Beobachtungspunkte tendenziell höher sein sollte als in Bereichen ohne Realisierung. Zu diesem Zweck wird zu jeder Beobachtung einer Stichprobe eine symmetrisch abfallende Kernfunktion mit Modalwert an der Beobachtungsstelle konstruiert. Die geschätzte Dichtefunktion $\hat{f}_h(x)$ entsteht dann durch das Aufsummieren der Kernfunktionen.

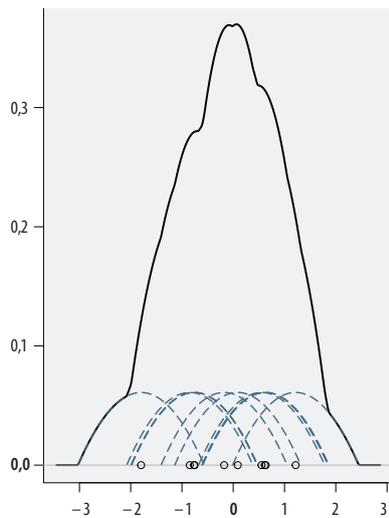
Diese Vorgehensweise ist für den univariaten Fall in Abbildung b illustriert. Die dargestellte Kernfunktion – der Epanechnikov-Kern – ist nur eine mögliche Kernfunktion unter vielen. Die Wahl des Kerns ist allerdings weniger kritisch als die Bandbreite (h , univariater Fall) bzw. die Bandbreitenmatrix (H , bivariater Fall), welche die Ausdehnung der Kernfunktion bestimmt.

Das Hauptaugenmerk der statistischen Inferenz liegt bei der (Kern-)Dichteschätzung auf der Entscheidungsregel zur Bestimmung des Bandbreitenparameters. Die auftretende Problematik ist in Abbildung c dargestellt: Eine zu kleine Bandbreite resultiert in sehr „gezackten“ Linien mit Gipfeln um die jeweiligen Beobachtungen, $\hat{f}_h(x)$ ist also sehr variabel. Allerdings ist die Verzerrung (Bias) sehr klein, die Schätzung bewegt sich in der Nähe der Dichtefunktion $f(x)$ (links). Wählt man hingegen eine sehr große Bandbreite, führt das zu einer sehr glatten Schätzfunktion. Die Varianz sinkt zwar, aber die Verzerrung nimmt zu, da die Charakteristika der unterliegenden Daten „überglättet“ werden. Eine sorgfältige Wahl des Glättungsparameters ist also wesentlich für die Anpassungsgüte. Dieses Dilemma wird auch als *Bias-Varianz Tradeoff* bezeichnet.

Gebräuchliche Schätzmethoden der Bandbreite basieren auf der Minimierung des asymptotischen mittleren integrierten quadrierten Abstandes (AMISE) der geschätzten Dichte von der wahren Dichte oder auf Kreuzvalidierungsverfahren. Aber auch einfache „Daumenregeln“, welche die Bandbreite als Faktor der geschätzten Stichprobenvarianz und des Stichprobenumfangs messen, können oftmals schon zu brauchbaren Ergebnissen führen.

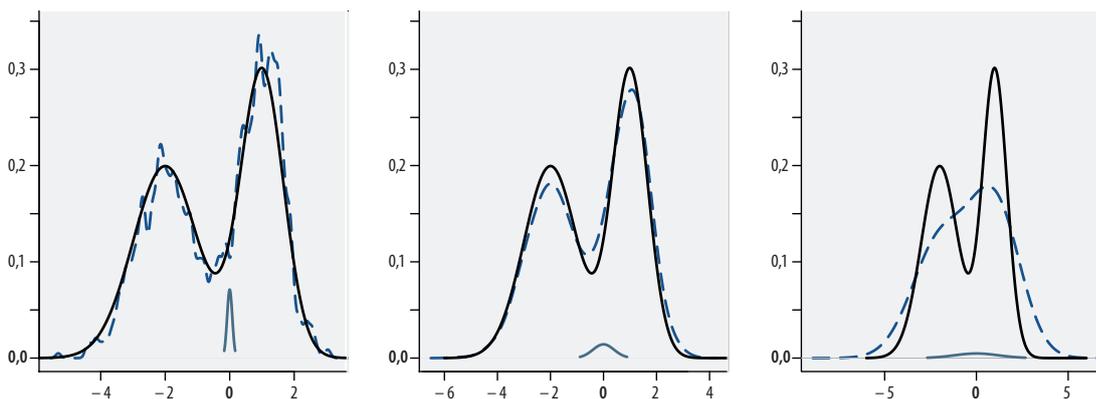
Eine ausführliche Darstellung verschiedener Methoden zur Bandbreitenwahl findet man beispielsweise bei Wand und Jones (1995)² oder Silverman (1986)³.

b | Kerndichteschätzung und Kernfunktionen auf Basis einer Stichprobe aus einer Standardnormalverteilung (n=10, Bandbreite=0.55) Schätzung mit Epanechnikov-Kern



2 Wand, Matt P. und M. C. Jones (1995): Kernel smoothing. Chapman & Hall Ltd.
 3 Silverman, Bernard W. (1986): Density Estimation for Statistics and Data Analysis. Chapman & Hall Ltd.

c | Drei Kerndichteschätzungen einer Mischverteilung mit unterschiedlichen Bandbreiten $h=0.08$ (links), $h=0.4$ (mitte) und $h=1.2$ (rechts). Zusätzlich wurde die Kernfunktion zur jeweiligen Bandbreite skizziert (Veranschaulichung nach Wand und M. Jones 1995)



Simulation

Der erste Schritt der Analyse besteht aus einer Simulation. Der Vorteil der Simulation liegt darin, dass die „Wahrheit“ bekannt ist. Das heißt, die geschätzte Dichtefunktion $\hat{f}_h(x)$ kann mit der tatsächlichen Dichtefunktion $f(x)$, aus welcher die simulierten Daten entstammen, verglichen werden.

Die Grundlage der Simulation bildet eine bivariate Mischverteilung (Z) aus drei normalverteilten Zufallsvariablen

$$Z = \frac{1}{3}N(\mu_1, \Sigma_1) + \frac{1}{3}N(\mu_2, \Sigma_2) + \frac{1}{3}N(\mu_3, \Sigma_3),$$

mit Erwartungswertvektoren μ_i und Kovarianzmatrizen Σ_i . Wie der Ausdruck bereits andeutet, wird jeweils mit Wahrscheinlichkeit $1/3$ aus der ersten, zweiten oder dritten Normalverteilung gezogen.

Abbildung d zeigt exemplarisch eine Stichprobe aus einer solchen Mischverteilung. Für die weitere Untersuchung soll die geschätzte Dichte der generierten Daten mit der geschätzten Dichte einer diskretisierten Form des Datensatzes verglichen werden. Die Gitterweiten entsprechen den Vergrößerungen, die aus Datenschutzgründen eventuell notwendig sein könnten. Im Rahmen der Simulation werden für die Gitterweiten (d) die Werte 0,25, 0,5, 0,75, 1 und 1,5 verwendet. Dabei bedeutet $d=0$, dass die Daten nicht vergrößert wurden und die Originaldaten (die gezogene Stichprobe) verwendet wurde.

Um die Anpassungsgüte an die tatsächliche Verteilung zu messen, wird der empirische *mean squared error* verwendet. Dabei wird für jeden Punkt der Abstand der Schätzung zur tatsächlichen Dichtefunktion berechnet und quadriert. Um die punktweise Anpassungsgüte auf die gesamte Stichprobe zu erweitern, werden die quadrierten Abweichungen der

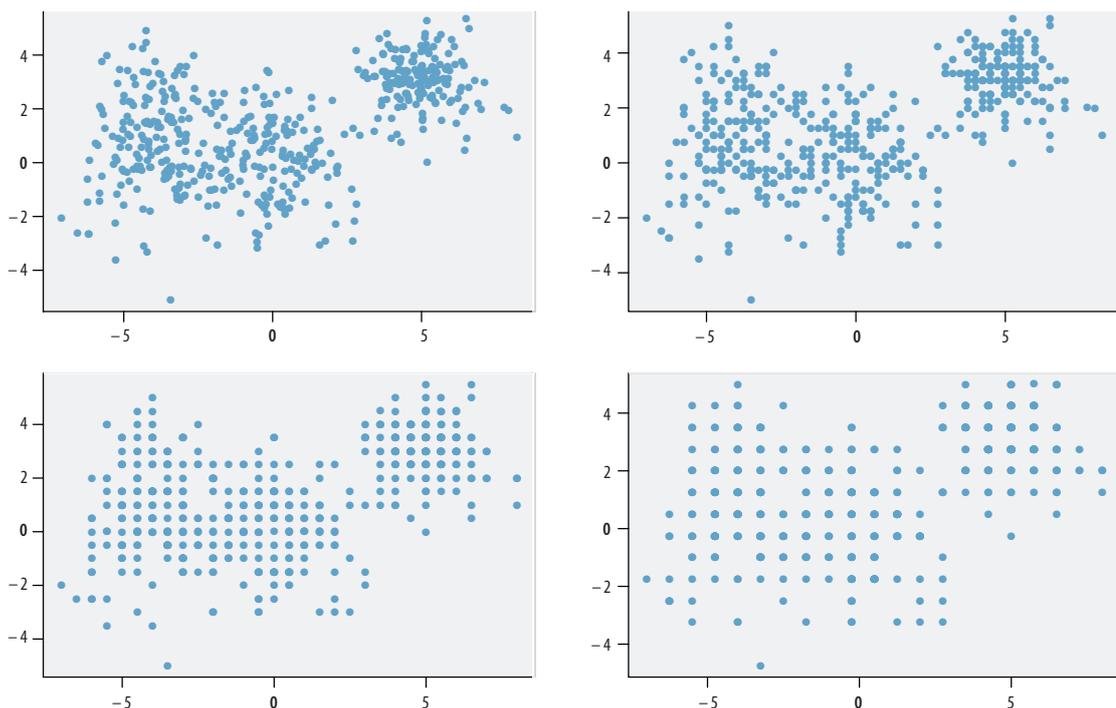
Schätzwerte von der theoretischen Dichtefunktion aufsummiert. Das Verhältnis der summierten Abstandsquadrate gibt somit ein Vergleichskriterium zur Bestimmung des Diskretisierungseffektes. Da es sich um Realisationen von Zufallsvariablen handelt, ist es allerdings möglich, dass eine Realisation des Kerndichteschätzers nur zufällig sehr nahe an der theoretischen Dichtefunktion liegt. Damit eine realistische Betrachtung der Ergebnisse möglich ist, wurde jede Simulation 500 Mal wiederholt.

Aufgrund der wichtigen Rolle, die der Bandbreitenwahl zukommt, wurden neben dem Einfluss unterschiedlich starker Diskretisierungsstufen auch verschiedene Methoden zur Wahl des Bandbreitenparameters in die Simulation aufgenommen: Der *two-stage plug-in selector* ist ein auf dem AMISE-Kriterium basierender moderner Schätzer; die *normal-reference rule* – auch „Silverman’s rule of thumb“ – ist eine einfache Daumenregel, welche auf einer Normalverteilungsannahme beruht. Silvermans Daumenregel überglättet die wahre Dichte allerdings leicht (siehe Silverman 1986).

Simulationsergebnisse

Die Konstruktionsweise des Kerndichteschätzers legt nahe, dass eine Anonymisierung durch Vergrößerung der Auflösungsgenauigkeit – also durch größere Gitterweiten – dazu führt, dass sich bei zu groben Gittern jeweils spitze Hügel um die Erhebungszentren bilden. Des Weiteren ist zu erwarten, dass der modernere two-stage plug-in selector zumindest für wenig diskretisierte Daten bessere Ergebnisse liefert. Die Simulationsergebnisse, die in Abbildung e visualisiert sind, bestätigen diese Vermutungen weitgehend.

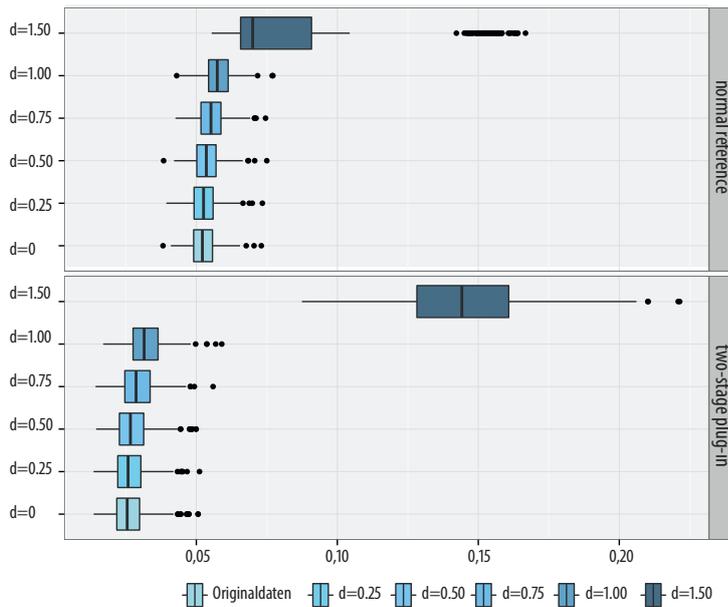
d | Visualisierung von vier Diskretisierungsstufen. Originaldaten (links oben), $d 1=0.25$ (rechts oben), $d 2=0.5$ (links unten) und $d 3=0.75$ (rechts unten)



In fünf von sechs Fällen liefert der two-stage plug-in selector tatsächlich eine bessere Schätzung als die einfache Daumenregel. Werden die Daten allerdings stark vergrößert (in diesem Fall ein Gitter mit Gitterlänge $d=1.5$), liefert die two-stage plug-in Methode sehr schlechte Ergebnisse. Eine mögliche Erklärung, wieso Silvermans Daumenregel bei sehr großen Gitterabständen systematisch bessere Ergebnisse lie-

fert, ist die einleitend erwähnte Tendenz zur „Überglättung“. Dies kann gerade bei vergrößerten Daten von Vorteil sein. Wie in Abbildung f zu sehen ist, führt eine kleine Bandbreite zu gezackten Schätzfunktionen mit jeweiligen Hochpunkten um die Ausprägungen. Genau das gleiche Phänomen führt hier zu sehr großen kumulierten Fehlern. Die geschätzte Dichtefunktion schwankt im Fall der two-stage plug-in Schätzer sehr stark um die Gitterpunkte, an welchen sich die Punkte aufgrund der Diskretisierung ansammeln. Dadurch entsteht eine Dichtefunktion mit jeweiligen Hochpunkten um die erstellten Gitterpunkte.

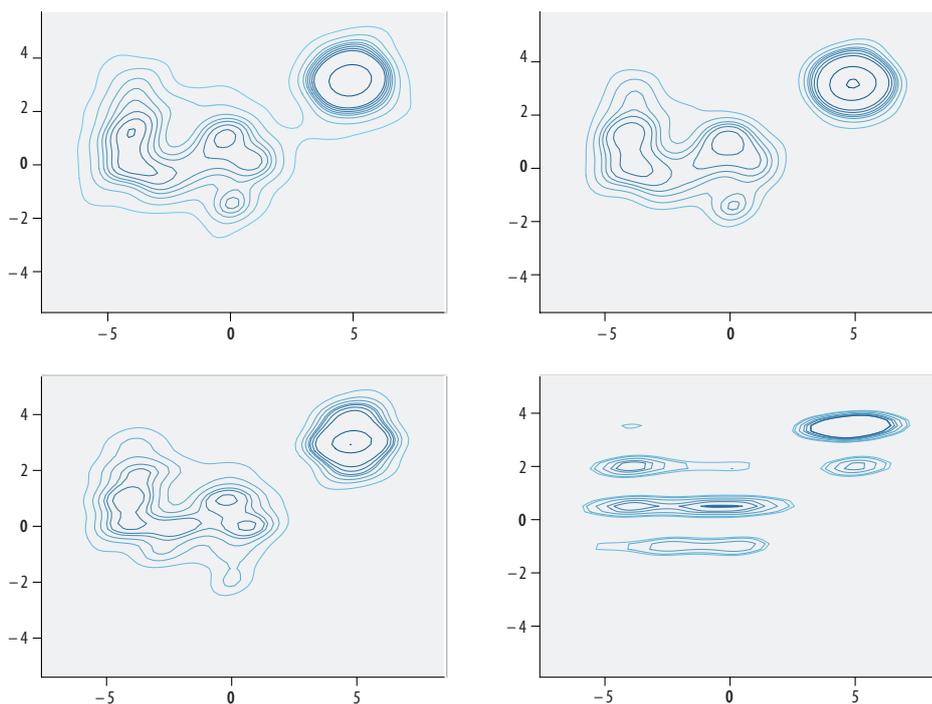
e | Vergleich der kumulierten absoluten Fehler der Simulationen mit normal reference rule (links). Vergleich der kumulierten absoluten Fehler der Simulationen mit two-stage plug-in selector (rechts)



Empirische Analyse

Im Folgenden soll die praktische Relevanz der untersuchten Fragestellung anhand eines Datensatzes illustriert werden. Im Speziellen stellt sich die Frage, bis zu welcher Gitterlänge der Kerndichteschätzer den wahren Zusammenhang in der Praxis noch korrekt darstellen kann. Im Gegensatz zur Simulation ist hier die „wahre“ Dichtefunktion nicht mehr bekannt. Die Simulationsergebnisse lassen jedoch den Schluss zu, dass die Schätzung auf Basis der Originaldaten die unterliegende Dichtefunktion erwartungsgemäß am besten

f | Visualisierung der Kerndichteschätzung durch Höhenlinien. Schätzung auf Basis der Originaldaten (links oben), mit Gitterweite $d=0.5$ (rechts oben), Gitterweite $d=1$ (links unten) und Gitterweite $d=1.5$ (rechts unten)



beschreibt, sodass diese hier als Referenz gewählt wird. Der verwendete Datensatz wurde vom Amt für Statistik Berlin-Brandenburg (AfS) über die kontrollierte Datenfernverarbeitung zur Verfügung gestellt und enthält das Einwohnerregister Berlin, d. h. melderechtlich registrierte Einwohnerinnen und Einwohner am Ort der Hauptwohnung zum 31.12.2012. Die Adressen sind in Form einer Projektion in die Ebene mit x - und y -Koordinaten, gemessen in Metern, abgespeichert, sodass die vorgestellten Kerndichteschätzer angewandt werden können. Zu den jeweiligen Koordinaten ist sowohl die Anzahl der Einwohner hinterlegt als auch die Anzahl von Einwohnern mit den Charakteristika

- Migrationshintergrund,
 - Migrationshintergrund, Herkunftsgebiet Türkei,
 - Migrationshintergrund, Herkunftsgebiet Vietnam,
 - jüngere Einwohner, d. h. Alter unter 18 Jahre,
 - ältere Einwohner, d. h. 60 Jahre oder älter,
 - weibliche ältere Einwohner, d. h. 60 Jahre oder älter.
- Exemplarisch werden hier die Einwohner mit türkischem Migrationshintergrund betrachtet.

Die Analyse der Daten folgt dem Muster der Simulation. Zunächst werden Gitter unterschiedlicher Größe erzeugt, die einer unterschiedlich starken Diskretisierung entsprechen. Daraufhin werden alle Koordinaten ihrem nächsten Gitterpunkt zugeordnet. Die verwendeten Gitterlängen sind $d=500$ m, $d=1000$ m, $d=1500$ m, $d=2000$ m und $d=2500$ m. Die Wahl der Bandbreite erfolgt mit dem two-stage plug-in selector. In Abbildung g ist zunächst der Kerndichteschätzer der Originaldaten visualisiert. Zusätzlich zur geschätzten Einwohnerdichte mit türkischem Migrationshintergrund sind Land- und Bezirksgrenzen der Stadt Berlin sowie Gewässer zur Orientierung hinzugefügt. Ein dunklerer Blauton bedeutet hierbei einen höheren Wert der Dichtefunktion an der betrachteten Stelle. Für jeden unterschiedlich stark diskretisierten Datensatz wurde die Dichtefunktion geschätzt.

Qualitative Auswertung

Bivariate Kerndichteschätzer haben den Vorteil, dass sie ein sehr intuitives Verständnis der Verteilung eines Merkmals in der Ebene geben können. Daher werden zunächst die grafischen Darstellungen der Schätzungen auf Basis der Diskretisierungsstufen verglichen. Abbildung h zeigt die Ergebnisse der Kerndichteschätzung auf Basis der vergrößerten Datensätze. Die Schätzungen mit den Gitterlängen $d=500$ m und $d=1000$ m bilden den Zusammenhang, wie er in Abbildung g dargestellt ist, sehr gut nach. Bei größeren Gitterlängen, namentlich $d=2000$ m und $d=2500$ m, ist wiederum derselbe Effekt wie bei der Simulationsstudie zu beobachten: Die Wahrscheinlichkeitsmasse der Dichtefunktion sammelt sich an den Punkten, an denen sich die Beobachtungen konzentrieren. Der Kerndichteschätzer legt also spitze Hügel um die gewählten Gitterpunkte, womit die komplexe Struktur der Bevölkerungsverteilung in Berlin kaum mehr sichtbar ist.

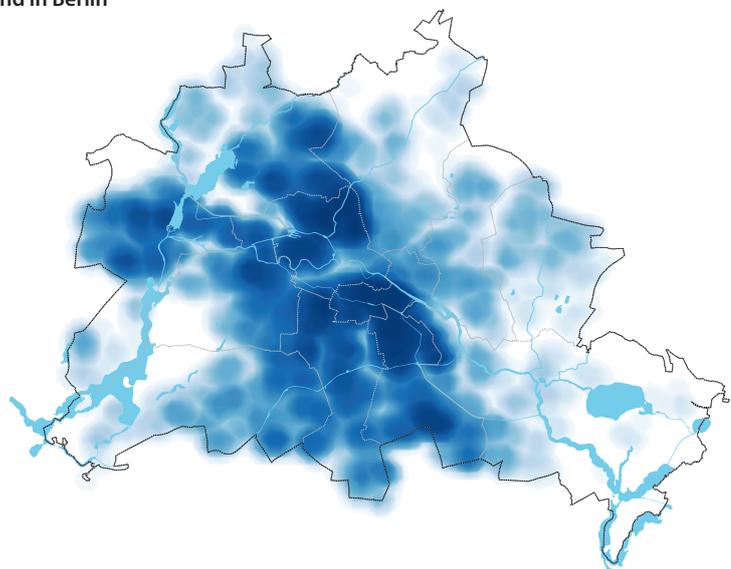
Quantitative Auswertung

Um die Anpassungsgüte der geschätzten Dichtefunktion an die den Daten unterliegende Dichte zu quantifizieren, bietet sich, wie schon in der Simulation, die Bildung der Fehlerquadratsumme an. Weil hier die „wahre“ Dichtefunktion nicht bekannt ist, muss ein anderer Referenzwert gefunden werden. Die Simulationsergebnisse legen nahe, dass die Schätzung auf Basis der Originaldaten in der Regel die beste Anpassung liefert. Anstatt die Abweichung der Schätzung der diskretisierten Daten in Bezug auf eine theoretische Dichtefunktion f zu messen, wird nun die Differenz zu deren Schätzung \hat{f}_H berechnet. Die Ergebnisse sind in der unten stehenden Tabelle dargestellt. Der Fehler, der durch eine Auflösung von 500 m brei-

Quotient der Abstandsquadratsumme zur geschätzten Dichte ohne Diskretisierungsfehler Gitterweite $d=500$ m als Referenzkategorie

Gitterweite.....	500	1000	1500	2000	2500
Fehlerquotient...	1	10.6	73.2	212.7	384.0

g | Schätzung der Dichte der Einwohner mit türkischem Migrationshintergrund in Berlin



Quelle: Statistische Mikrodaten aus dem Berliner Einwohnerregister, 2012, eigene Berechnungen.

ten Gittern entsteht, wurde auf 1 normiert. Während bei einer Vergrößerung der Gitterlänge von 500 m auf 1000 m der Anstieg der Fehlerquadratsumme noch vergleichsweise gering ausfällt, führen größere Gitterbreiten zu stark erhöhten Fehlertermen.

Fasst man die qualitativen und quantitativen Resultate zusammen, lässt sich folgendes Fazit ziehen: Im Falle des vorliegenden Merkmals führt eine Anonymisierung der Adressdaten durch eine Ansammlung der Punkte auf ein Gitter der Länge 500 oder 1000 Meter nur zu geringen Verzerrungen. Wählt man eine größere Gitterlänge, resultieren stark ansteigende Abweichungen von der direkten Schätzung, die sich wiederum durch die charakteristischen „Dichtehügel“ um die Gitterpunkte bemerkbar machen.

Fazit

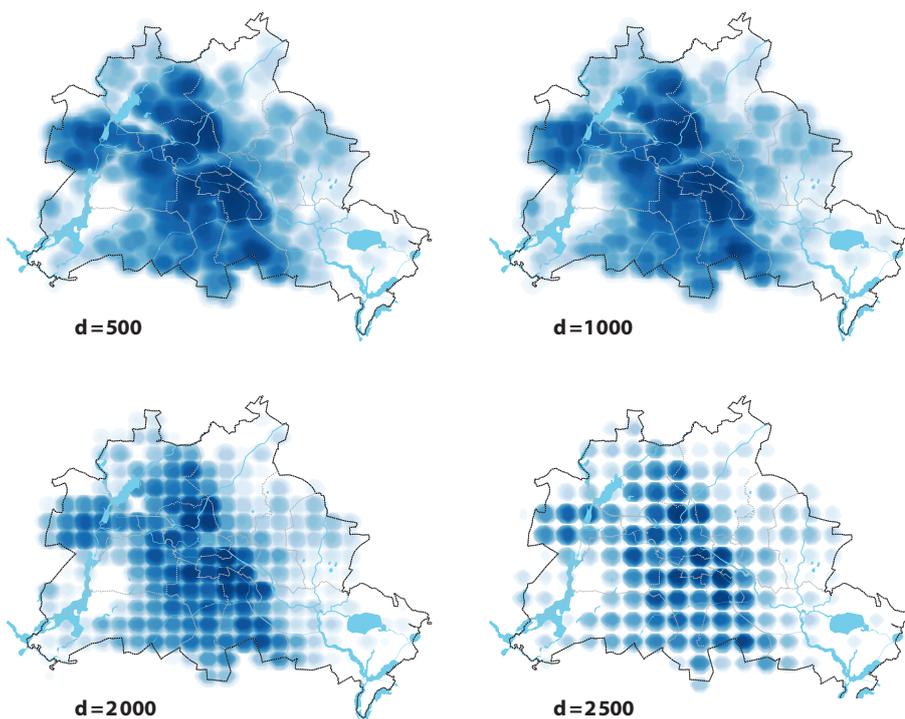
Räumliche Daten, wie Adressdaten, können aufschlussreiche Informationen über ein Merkmal liefern. Um georeferenzierte Daten zur statistischen Analyse zur Verfügung stellen zu können, ist es allerdings notwendig, Maßnahmen zu ergreifen, die dem Datenschutz genügen. Dafür muss unter anderem geklärt werden, welche Erhebungs- oder Speichergenauigkeit für eine ausreichende Anonymisierung sorgen.⁴ Auf der Seite der Anwenderinnen und Anwender stellt sich die Frage, welche Auswirkungen verschiedene Anonymisierungsmethoden auf die Ergebnisse einer statistischen Datenauswertung ha-

ben oder welche Methoden sich eignen, Diskretisierungsfehler aus den Ergebnissen herauszurechnen. Die Entwicklung solcher Verfahren ist Gegenstand weiterer Forschung.

Die Ergebnisse der Simulation deuten überdies an, dass Methoden zur Bandbreitenwahl, die bei nicht-diskretisierten Daten gute Ergebnisse liefern, bei diskretisierten Datensätzen möglicherweise schlechtere Resultate liefern können. Mit welcher Regel zur automatisierten Bandbreitenwahl die Schätzungen am wenigsten unter dem Informationsverlust leiden, könnte Inhalt weiterer Simulationen sein. Die Analyse anhand des Datensatzes ergibt für das beobachtete Merkmal (Einwohner in Berlin mit Migrationshintergrund, Herkunftsland Türkei), dass eine Gitterlänge von bis zu ca. 1000 m die Struktur der geschätzten Dichtefunktion relativ wenig verändert. Für weitere Untersuchungen ist die Generalisierbarkeit dieser Ergebnisse von besonderem Interesse, d.h. die Suche nach allgemeinen Toleranzwerten für georeferenzierte Daten.

Sebastian Schmon erhielt seinen Bachelor in Volkswirtschaftslehre an der Freien Universität Berlin und studiert nun Mathematik und Statistik. Die vorgestellten Ergebnisse entstanden im Rahmen der Abschlussarbeit im Fach Volkswirtschaftslehre.

⁴ Rat für Sozial- und Wirtschaftsdaten (2012): Georeferenzierung von Daten. Situation und Zukunft der Geodatenlandschaft in Deutschland. SCIVERO Verlag, Berlin.



h | Vergleich der Schätzungen auf Basis unterschiedlich diskretisierter Daten Gitterlänge

Quelle: Statistische Mikrodaten aus dem Berliner Einwohnerregister, 2012, eigene Berechnungen.

Geheimhaltung

Statistische Geheimhaltung bei der Auswertung georeferenzierter Daten

von Jörg Höhne und Julia Höninger

Auch bei der Auswertung von georeferenzierten Statistiken, die dem Bundesstatistikgesetz oder Landesstatistikgesetzen unterliegen, muss die amtliche Statistik sicherstellen, dass das Statistikgeheimnis gewahrt wird. Bei der Anwendung der gängigen Geheimhaltungsregeln in der statistischen Geheimhaltungsprüfung ist die Mindestfallzahl bei georeferenzierten Daten ein besonderer Diskussionspunkt. Als Methoden können auch bei diesem Datentyp traditionelle oder datenverändernde Geheimhaltungsverfahren verwendet werden. Des Weiteren werden die Enthüllungsrisiken der geografischen Differenzbildung und das Prüfen bei der Darstellung von Verhältniszahlen in Karten erläutert.

1. Warum statistische Geheimhaltung?

Die Wahrung des Statistikgeheimnisses ist eine zentrale Aufgabe der amtlichen Statistik. Aus den Veröffentlichungen der statistischen Ämter darf kein Rückschluss auf die Einzelangaben der Befragten gezogen werden. Schon in der Begründung zum Bundesstatistikgesetz (BStatG)¹ 1987 wird betont, dass die statistische Geheimhaltung nicht nur eine gesetzliche Aufgabe, sondern auch essentiell für das Vertrauensverhältnis zwischen Befragten und den statistischen Ämtern ist. Dieses ist wiederum wichtig, um qualitativ hochwertige Statistiken erheben zu können (Statistische Ämter des Bundes und der Länder 2006, S. 9).

Um das Statistikgeheimnis zu wahren, werden alle Auswertungen vor der Veröffentlichung einer Prüfung unterzogen, und zwar der sogenannten statistischen Geheimhaltung. Zur Identifikation von Tabellenzellen, die möglicherweise nicht veröffentlicht werden dürfen, werden in der amtlichen Statistik üblicherweise drei Geheimhaltungsregeln angewandt. Diese Regeln können grundsätzlich auch auf Auswertungen von georeferenzierten Daten angewandt werden. Jede geografische Einheit, für die eine Information veröffentlicht werden soll, ob nach Koordinaten, nach Rasterzellen oder anderen Regionalschlüsseln, wird dabei wie ein Tabellenfeld behandelt.

In Verbindung mit dem E-Government-Gesetz (EGovG)² ist die Speicherung von Daten in Rastern mit einer Größe von 100 x 100 Metern erlaubt, so ist es bei der Auswertung und Veröffentlichung von Ergebnissen jedoch stets notwendig, die Geheimhaltung zu prüfen und falls notwendig durch geeignete Geheimhaltungsverfahren das Statistikgeheimnis zu wahren.

Bei Veröffentlichungen aus Bundesstatistiken gilt grundsätzlich das BStatG. Bei anderen Datenproduzenten unterliegen Auswertungen dem Bundesdatenschutzgesetz (BDSG)³. Hier gilt ein Geheimhaltungserfordernis nur dann, wenn die Daten personenbezogen sind. Bei Auswertungen georeferenzierter Daten ist bei diesen der Personenbezug zuerst zu prüfen. Einige Geofachdaten beziehen sich auf Sachen, wie beispielsweise Grundstücke. Wenn der jeweilige Gegenstand, der Ort oder das Grundstück einer natürlichen Person zugeordnet werden kann, sind die Veröffentlichungen nach dem BDSG zu prüfen. Unter welchen Bedingungen eine Bestimmbarkeit angenommen werden muss, ist laut Karg (2008, S. 8) letztlich nicht endgültig geklärt und in der Praxis der Aufsichtsbehörden, der Literatur und Rechtsprechung umstritten. Da bei Auswertungen von Geofachdaten aus Bundesstatistiken nach §16 BStatG alle Einzelangaben geschützt werden müssen, ist es nicht von Bedeutung, ob die Einzelan-

1 Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 13 des Gesetzes vom 25. Juli 2013 (BGBl. I S. 2749).

2 Durch das Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz – EGovG) vom 25. Juli 2013

(BGBl. I S. 2749) wird in §10 BStatG (Erhebungs- und Hilfsmarkale) die „geografische Gitterzelle“ eingefügt. Diese umfasst eine Fläche von 100 x 100 Metern (= 1 Hektar oder 1 ha). Damit wurde eine Rechtsgrundlage für die Speicherung und Verbreitung georeferenzierter statistischer Angaben geschaffen. Ziel ist die flexiblere

räumliche Auswertung von Statistiken. Die statistische Geheimhaltung ist aber weiterhin zu wahren. In §13 BStatG werden auch in Adressdateien die „Geokoordinaten“ ergänzt. Hierdurch erfolgte die Klarstellung, dass die dauerhafte Speicherung der Geokoordinate in Adressdateien zulässig ist.

3 Bundesdatenschutzgesetz (BDSG) in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), zuletzt geändert durch Artikel 1 des Gesetzes vom 14. August 2009 (BGBl. I S. 2814).

gaben von natürlichen oder juristischen Personen gemacht wurden oder aus Verwaltungsdaten entnommen sind.

2. Geheimhaltungsprüfung

a) Anwendung der Geheimhaltungsregeln

Die gängigen Geheimhaltungsregeln in der amtlichen Statistik sind die Mindestfallzahl-, die Randwert- und die Dominanzregel. Die Mindestfallzahlregel schreibt vor, dass stets eine festgelegte Anzahl an Befragten zu einem veröffentlichten Ergebnis beiträgt. Nach der Randwertregel ist eine Tabellenzeile oder -spalte zu sperren, wenn alle Merkmalsträger die gleiche Ausprägung ausweisen. Ein Dominanzfall liegt vor, wenn ein oder zwei Beitragende einen zu großen Anteil der Gesamtsumme auf sich vereinen.

Überträgt man die Mindestfallzahlregel auf Rasterveröffentlichungen, so sollen in jeder Gitterzelle mehrere oder kein Merkmalsträger enthalten sein. Ist diese Bedingung nicht erfüllt, darf zu dieser Gitterzelle keine Angabe veröffentlicht werden. Der Thematik, welcher Schwellenwert bei der Mindestfallzahlregel bei der Auswertung von georeferenzierten Daten gewählt werden sollte, wird in Abschnitt 2b) nachgegangen.

Die Dominanzregel ist bei den Statistiken relevant, bei denen metrische Merkmale georeferenziert vorliegen. Das trifft beispielsweise oft bei den Wirtschaftsstatistiken zu. Die Konzentration von Kennzahlen wie Umsätze, Exporte oder Anzahl der Beschäftigten bei kleinen regionalen Gliederungen kann eine leichte Zuordnung zu den beitragenden Einheiten erlauben. Um zu überprüfen, ob ein Dominanzfall vorliegt, wird die $p\%$ -Regel empfohlen. Ein Tabellenfeld wird nach der $p\%$ -Regel dann gesperrt, wenn der zweitgrößte Beitragende den größten Beitragenden aufgrund der veröffentlichten Gesamtsumme und der Kenntnis seines eigenen Beitrags so genau schätzen könnte, dass die Schätzung um weniger als $p\%$ vom tatsächlichen Wert abweicht.

Auch bei der Auswertung von georeferenzierten Daten können Randwerte entstehen. Wenn alle Merkmalsträger in einer regionalen Einheit die gleiche Ausprägung aufweisen, so ist enthüllt, welche Ausprägung jeder einzelne Merkmalsträger bei diesem Merkmal aufweist. Als Extrembeispiel stelle man sich eine Rasterkarte vor, die angibt, wie viele Unternehmen wegen Umweltvergehen verurteilt wurden. In einer Rasterzelle wurden 23 Unternehmen verurteilt, es gibt in diesem Gebiet jedoch nur 23 Unternehmen, was aus einer anderen Quelle ermittelt werden kann. Dann weisen alle Unternehmen die gleiche Ausprägung auf: „Verurteilt: ja“ und die Leserinnen und Leser wissen, dass jedes einzelne Unternehmen ein Umweltvergehen begangen hat. Bei Randwert-Konstellationen bestehen Enthüllungsrisiken, auch wenn es sich nicht um zu kleine Fallzahlen handelt (Walla 2007).

Daten der amtlichen Statistik unterliegen dem BStatG. Jedoch ist eine Geheimhaltungsmaßnahme nach §16 Satz 1 Nr. 4 BStatG nur notwendig, wenn die Einzeldaten dem Merkmalsträger zugeordnet

werden können. Ein Gutachten von Karg (Karg 2008), das sich auf Daten bezieht, die dem BDSG unterliegen, kann hier eventuell zu Vergleichszwecken herangezogen werden. Demnach besteht ein Personenbezug bei Fachdaten zur Landnutzung (Karg 2008, S. 61) und Kriminalität (Karg 2008, S. 65).

b) Welche Mindestfallzahl soll angewendet werden?

In einem von der Arbeitsgruppe Geodatenschutz des Interministeriellen Ausschusses für Geoinformationswesen (2014, S. 13) erarbeiteten „Behördenleitfaden zum Datenschutz bei Geodaten und -diensten“ wird empfohlen, dass bei Auswertungen mit geografischem Bezug stets mindestens vier Einheiten zu einem Datum beitragen sollen. Es wird hier vor allem auf Haushalte abgestellt, aber erwähnt, dass dieses Kriterium auch bei Daten auf Personenebene angewandt werden kann. Dann sei kein Personenbezug mehr herstellbar. Allerdings enthält dieser Leitfaden außer dieser Handlungsempfehlung keine methodische Begründung dieses Schwellenwertes. Als Quelle wird vielmehr folgender Abschnitt aus Karg und Weichert (2007, S. 25) genannt:

„Eine Anonymisierung im Hinblick auf das Eigentum kann dadurch erreicht werden, dass mehrere Grundstücke zusammengefasst werden und deren Eigenschaft in gemeinsamer Form, z.B. durch einen Mittelwert, beschrieben wird. Diese Form des Zusammenfassens von personenbezogenen Daten wird auch Aggregieren genannt. Es gibt keine eindeutigen rechtlichen Vorgaben, wann von einer Anonymität hinreichend gewährleistenden Aggregation gesprochen werden kann. Unstreitig ist, dass eine Zusammenfassung von zwei Personeneinheiten zu einer Merkmalseinheit noch nicht ausreicht. Entsprechend der statistikrechtlichen Praxis kann man davon ausgehen, dass bei einer Zusammenführung von mindestens vier Personeneinheiten zu einem Datensatz der Personenbezug hinreichend verschleiert wird.“

Es ist allerdings festzustellen, dass in der Praxis und der Literatur unterschiedliche Schwellen genannt werden. Karg (2008, S. 23) dokumentiert, dass manchmal die „Aggregation“ von mindestens vier, von Aufsichtsbehörden aber manchmal bis zu zehn Grundstücken gefordert werde. Zur Begründung wird auf die erhöhte Sicherheit der Betroffenen verwiesen. Je mehr Angaben zusammengenommen werden, desto geringer sei die Wahrscheinlichkeit der Reidentifizierung. In seinem Gutachten empfiehlt der Rat für Sozial- und Wirtschaftsdaten (2012, S. 49) gar, den australischen Ansatz zu übernehmen. Dort werden die Raster so groß gewählt, dass sie mindestens 30 Wohneinheiten oder 60 Personen enthalten. Auch Szibalski (2007, S. 142) dokumentiert sehr unterschiedliche Mindestfallzahlen. Bei einer Umfrage unter nationalen Statistikämtern in Europa im Jahr 2006 schwankten diese zwischen zwei Beitragenden bis zu mindestens 20 und in einem Fall gar mindestens 31 Merkmalsträgern in einer Rasterzelle. In einem aktuelleren Beitrag auf der Konferenz des „European Forum for Geography and Statistics (EFGS)“ berichten Martin et al. (2013), dass

in Norwegen eine Mindestfallzahl von 0 (also keine), in Österreich 30 und in England 100 gilt. Auch sie stellen fest, dass es in Europa derzeit eine große Variation der Mindestfallzahlen gibt. Auf der Internetseite des EFGS selbst ist ein Leitfaden⁴ zu finden, der Rasterzellen mit einer Größe von 1 km² empfiehlt. Falls Geheimhaltung relevant wäre, so solle mit einer Mindestfallzahl von 10 geprüft werden.

Im Gutachten von Karg (2008, S. 24) werden weitere Randbedingungen für eine „zuverlässige Aggregation“, gemeint ist hier eine ausreichende Zusammenfassung, genannt. Diese entsprechen der üblichen Praxis in den statistischen Ämtern. Die Prüfung der Geheimhaltung bei zusammenhängenden Einheiten muss auf der hierarchischen Ebene durchgeführt werden, von der aus Eigentumsbeziehungen zu den unteren Ebenen bestehen. Um beispielsweise die Einzelangaben einer Pflegeeinrichtung zu schützen, müssen die Angaben von Einrichtungen verschiedener Träger zusammengefasst werden. Sollen Angaben über Betriebe veröffentlicht werden, müssen die Angaben von so vielen Betrieben zusammengefasst werden, dass mehrere Unternehmen als Eigentümer hinter diesen Betrieben stehen. Weichert (2009, S. 351) nennt als weiteres Kriterium, dass keine Zusatzinformationen vorhanden sein dürfen, dass beispielsweise alle Grundstücke gleich groß sind, wie es in manchen Neubausiedlungen der Fall ist. Dies würde in der amtlichen Statistik als Randwert identifiziert und sollte daher nicht veröffentlicht werden.

Eine alternative Handlungsempfehlung, die in diesem Beitrag nicht unterstützt werden kann, bezieht sich bei Auswertungen mit geografischem Bezug auf den Maßstab bzw. Detaillierungsgrad von Karten. In einem Beitrag wurde als „allgemeine Regel“ ein Maßstab von 1:10000 empfohlen (Karg 2008, S. 13). Allerdings merkt Weichert (2009, S. 350) korrekterweise an, dass es „keine maximalen Grundstücksgrößen gibt, und daher keine Flächengröße [bei Auswertungen von Flächendaten] genannt werden kann, ab der ein Personenbezug ausgeschlossen werden kann“. Ein Maßstab alleine kann deshalb als Prüfkriterium bei der Sicherstellung des Statistikgeheimnisses nicht empfohlen werden.

Es wird geraten, vor der Veröffentlichung von Karten oder Rasterzellen die Besetzungszahlen jeder geografischen Einheit anhand der aufgeführten und diskutierten Geheimhaltungsregeln zu prüfen. Eine einheitliche Empfehlung lässt sich aus der Literatur nicht ableiten. Der Datenschutzbeauftragte des Landes Schleswig-Holstein Thilo Weichert (2009, S. 351) empfiehlt: „In der Praxis bietet es sich an, eine größere Aggregation vorzunehmen. Bei einer Zusammenfassung von 10 Grundstücken kann in der Regel von einer hinreichend Anonymität herstellenden Aggregation ausgegangen werden.“ Je höher die Mindestfallzahl, desto eher ist die Besetzung für den Datenschutz ausreichend. Allerdings ist der Informationsverlust auch höher, da durch die Anwendung der Geheimhaltungsregel mehr Gitterzellen oder Kartenflächen als schützenswert markiert werden.

c) Traditionelle Geheimhaltungsmethoden:

Sperrung oder Vergrößerung

Zeigt eine der Geheimhaltungsregeln an, dass die Informationen für die regionale Einheit sensibel oder kritisch ist, so darf für diese regionale Einheit die Angabe nicht veröffentlicht werden. Sind in einer Rasterzelle laut Mindestfallzahlregel zu wenige Befragte enthalten oder dominiert ein Beitragender nach der Dominanzregel, kann als Maßnahme die Information in der Zelle unterdrückt werden. Liegt ein Randwertproblem vor, müssen mehrere Felder unterdrückt werden. Das Geheimhaltungsverfahren wird Zellspernung genannt und den traditionellen Geheimhaltungsmethoden zugeordnet.

In Tabellen werden gesperrte Felder durch den Punkt „.“ gekennzeichnet. In Karten wird in der Regel eine Grau- oder Weiß-Färbung verwendet und diese in der Legende erklärt. Beim Atlas „Agrarstatistik“ wird sie in der Legende mit „Kein Wert vorhanden oder geheim zu halten“ erläutert. Wird als Geheimhaltungsmethode die Sperrung verwendet, bleibt zu untersuchen, ob aufgrund von Summenbeziehungen zu veröffentlichten Randsummen weitere Gitterzellen zu sperren sind. Dieses Problem ergibt sich, wenn für größere Gebiete, die sich additiv aus den einzelnen Gitterzellen ergeben, ebenfalls Werte veröffentlicht werden. In diesem Fall ergibt sich bei Unterdrückung eines Wertes noch kein Schutz.

Alternativ kann das Raster lokal größer gezogen oder mehrere Zellen zusammengelegt werden. Es kann auch für die gesamte Karte ein breiteres Raster gewählt werden.

Jedes Geheimhaltungsverfahren führt zu einem Informationsverlust. Das Sperren von Werten in einzelnen Rasterzellen oder georeferenzierten Flächen in Kartendarstellungen bedeutet ein Zurückhalten von Informationen. Das Vergrößern der Flächen oder Zellen bringt einen Verlust der geografischen Genauigkeit (loss in spatial accuracy) mit sich. Eine Analyse zum Informationsverlust, der durch größere Gitterzellen entsteht, findet sich im Beitrag von Schmon in diesem Heft.

d) Datenverändernde Geheimhaltungsverfahren

Eine Alternative zu den traditionellen Geheimhaltungsverfahren, die Informationen in Veröffentlichungen unterdrücken, indem beispielweise Tabellenfelder oder Gitterraster gesperrt werden, ist die neuere Klasse der datenverändernden Geheimhaltungsverfahren. Bei diesen wird ein Rückschluss auf einzelne Befragte verhindert, indem entweder die Einzeldaten oder die statistischen Ergebnisse leicht verändert werden. Ein Überblick über klassische

⁴ Der Leitfaden (EFGS Standard for official grid statistics, all variables v.1.0, Wordformat) kann von folgender Seite abgerufen werden (Stand 29.04.2014): <http://www.efgs.info/geostat/1B/efgs-standard-for-official-statistics/view>. Dort findet sich folgende Textpassage: "The following statistical units, variables and divisions are recommended as official statistics;

free of charge and without disclosure to be reported on km² grid cells. If disclosure is still considered to be needed, the recommended threshold value is set to 10." Die im Leitfaden aufgeführten statistischen Auswertungen fallen in Deutschland alle unter § 16 BStatG, sobald Bundesstatistiken ausgewertet werden, um die Kennzahlen zu berechnen.

und datenverändernde Geheimhaltungsverfahren findet sich in Höhne (2010).

In manchen Ländern werden datenverändernde Geheimhaltungsverfahren zur Sicherung des Statistikgeheimnisses bei Veröffentlichungen, die aus georeferenzierten Daten erstellt werden, diskutiert oder verwendet. Einen Überblick über Vorschläge, datenverändernde Verfahren auf Geodaten anzuwenden, bietet Young et al. (2012). Armstrong et al. (1999) hat dabei einen neuen Begriff für die Anwendung der Datenveränderung auf die georeferenzierten Merkmale eingeführt: „geographical masking“ oder verkürzt zu „geomasking“ bezeichnet die prätabulare, datenverändernde Geheimhaltung durch Veränderung der geografischen Koordinaten.

In der Praxis wurde in Finnland ein selbst weiterentwickeltes Verfahren zur „restringierten lokalen Imputation“ eingesetzt (Statistics Finland 2010). In England wurde eine Variation des Record Swapping vorgeschlagen; bei diesem sogenannten „local density swapping“ wird in dichter besiedelten Regionen mit einer geringeren Wahrscheinlichkeit „geswappt“, also das Attribut der geografischen Verortung mit anderen Merkmalsträgern getauscht. Das Verfahren wurde an anonymisierten Zensusdaten von 1991 getestet und schnitt bei den Informationsverlustmaßen besser ab als ein zufälliges Record Swapping Verfahren. In der Gesamtpopulation einzigartige Ausprägungen können durch eine zufällige Veränderung der geografischen Verortung durch ein Swapping-Verfahren jedoch nicht ausreichend geschützt werden (Young et al. 2012).

In Deutschland wurde ein Mikroaggregationsverfahren, das sich auch als prätabulares Geheimhaltungsverfahren eignen würde, auf georeferenzierte Arbeitsmarktdaten angewandt. In erster Linie wurden die Gruppen von mindestens 15 Erwerbstätigen, jedoch nicht aus Geheimhaltungsgründen, gebildet, sondern um sogenannte Nachbarschaften zu bilden und dort Nachbarschaftseffekte zu untersuchen (Scholz et al. 2012).

Werden georeferenzierte Ergebnisse als detaillierter Input für weitere Modellrechnungen und nicht direkt zur Veröffentlichung benötigt, so besteht eine Möglichkeit darin, der Wissenschaft in geschützten Räumen und unter organisatorisch-rechtlichen Rahmenbedingungen über Forschungsdatenzentren Zugang zu „georeferenzierten Mikrodaten“ bzw. zu georeferenzierten Ergebnissen ohne Geheimhaltungsprüfung zu ermöglichen. Dann kann die Wissenschaft innerhalb des Forschungsdatenzentrums forschen, und die dort erzeugten Ergebnisse werden danach auf das Statistikgeheimnis geprüft.

3. Aspekte des Rasters aufgrund der Geheimhaltung

a) Gitterzellenbasierte Grundeinheiten

Um den gespeicherten Datensatz flexibel auswerten zu können, sollten die Gitterlänge, die in Veröffentlichungen verwendet wird, und die Basislänge des Gitterrasters, das an den Einzeldaten gespeichert ist, in einer Beziehung zueinander stehen. Aufgrund der Änderungen des §10 BStatG darf an den Einzeldaten von Bundesstatistiken die Zugehörigkeit zu einer Gitterzelle mit der Größe 100 x 100 Meter gespeichert werden. Wenn sich bei Auswertungen herausstellt, dass dieser Detailgrad häufig nicht für Veröffentlichungen verwendet werden kann, so sollte eine „übliche“ Gitterlänge für diese Statistik festgelegt werden.

Durch das im Rahmen von INSPIRE festgelegte Euro-Grid, bei dem sowohl die Projektion (Lambert Azimuthal Equal Area), die Lage der Gitterzellen, die Gitterweite und Schlüssel als Identifikatoren jeder Zelle etabliert wurden, sind neue funktionale Raumbezüge geschaffen worden. Die eindeutigen Schlüssel können als RasterIDs verwendet werden. Dies ist hilfreich, um eine feste Gebietssystematik an gitterzellenbasierten Grundeinheiten und somit Regionsidentifikatoren zu dokumentieren. Dadurch können Raumstrukturdaten (z.B. Bebauung, Erreichbarkeit) in Forschungsvorhaben weiter verwendet werden. Auch können beispielsweise Wissenschaftler Geofachdaten über diese RasterIDs an Mikrodaten in den Forschungsdatenzentren anspielen.

Solch eine rasterbasierte Gebietssystematik sollte einerseits möglichst kleinräumig sein. Andererseits sollten die Gitterzellen jedoch so ausgewählt sein, dass aufgrund des Datenschutzes für möglichst viele Gitterzellen auch Informationen enthalten sein können (Sigismund 2014). Die Vorteile rasterbasierter Gebietssystematiken sind, dass sie zeitreihenrobust und themenneutral sind. Da sie weder topografische, noch siedlungsstrukturelle/lebensweltliche Faktoren oder verwaltungstechnische Gliederungen berücksichtigen, eignen sie sich sowohl für Sozial-, Wirtschafts- als auch Bevölkerungsdaten. Die gerade genannten themenbasierten Gliederungen eignen sich zwar besser für spezielle Planungs- und Analyse Zwecke im Themengebiet. Diese Gliederungen sind aber meistens zeitlichen Änderungen unterworfen. Rasterbasierte Gebietssystematiken sind dagegen zeitlich stabil. Damit sind sowohl zeitreihenorientierte aber auch themenübergreifende Analysen problemlos möglich. Bei anderen Gebietssystematiken bedingen die Pflege zeitlicher Veränderungen bei gleichen Systematiken bzw. die Abstimmung verschiedener Systematiken einen erheblichen Vorbereitungsaufwand bei der Analyse der geobasierten Daten.

b) Informationsverlust minimieren durch hierarchische Gitterstruktur

Grundsätzlich ist ein bundesweit einheitliches Raster wünschenswert. In allen Gebieten die gleiche Rastergröße zu verwenden, hat jedoch auch eine Reihe von Nachteilen. Bei kleinmaschiger Auswertung entstehen in dünn besiedelten Landgebieten oft viele Geheimhaltungsfälle; wählt man ein großes Raster, ist der Informationsgehalt für Städte sehr niedrig. Das Informationspotenzial der Daten wird dann nicht genutzt.

Fasst man benachbarte Gitterzellen bis zur Erreichung von Mindestbesetzungszahlen zu Nachbarschaftsarealen zusammen, entstehen unregelmäßige „Gitterklumpen“. Das Verfahren ist methodisch aufwändig und nicht zeitreihenrobust (Sigismund 2014).

Optimal wäre daher eine Gebietssystematik, deren Gitterbreite variiert. Hier würde sich eine hierarchische Aggregation anbieten. Eine solche wird vom Bundesamt für Statistik der Schweiz (Meyer 2011) und in Österreich (Strobl 2005) bereits verwendet. Anhand eines Datenbestands der Gebäude in Deutschland haben Behnisch et al. (2013) mit diesem Verfahren Mischrasterkarten für Hamburg erstellt. Bei der hierarchischen Aggregation wird von einer kleinen Basislänge der Gitterzellen gestartet. Sind die Besetzungszahlen nicht ausreichend, werden hierarchisch vier Gitterzellen zu einer neuen Zelle mit doppelter Gitterlänge zusammengefasst. Sollten auch in diesen größeren Zellen noch Mindestbesetzungszahlen unterschritten sein, so werden wiederum vier Zellen zu einer sehr großen zusammengelegt, die dann eine Gitterlänge vom Vierfachen der Basislänge hat.

Bei der Interpretation ist zu beachten, dass die Flächen der Gitterzellen nun unterschiedlich groß sind und die Absolutwerte daher nicht mehr zugleich auch Dichtewerte darstellen.

4. Geografische Differenzbildung

Die Mehrheit der Auswertungen der amtlichen Statistik bezieht sich auf administrative Gebiete (wie z.B. Länder, Regierungsbezirke, Kreise oder Gemeinden). Aber in einigen Statistiken gibt es bereits abweichende regionale Gliederungssysteme.

Die Reisegebiete in der Tourismusstatistik, die Wassereinzugsgebiete bei Wasserstatistiken, wie der öffentlichen Abwasserbehandlung, oder auch die EU-Orte (Siedlungsstruktur⁵) beim Zensus 2011 (Heidrich-Riske et al. 2013, S. 474) weichen von den administrativen Einheiten ab⁶. Bei diesen Statistiken wird eine andere, funktional besser passende räumliche Abgrenzung gewählt als die administrativen Einheiten. Wenn Statistiken wie die Wasserstatistiken oder die Tourismusstatistik zusätzlich zur funktionsräumlichen Gliederung auch für administrative Gebiete ausgewertet werden, können durch das „Übereinanderlegen“ verschiedener Gliederungssystematiken – das sogenannte Verschneiden – kleine Schnittmengen entstehen. Die Thematik wird auch als „disclosure by geographical differencing“ in der internationalen Geheimhaltungsliteratur diskutiert (Hundepool et al. 2010, S. 169).

Dieses Enthüllungsrisiko ist zusätzlich zu ausreichenden Besetzungszahlen zu prüfen. Durch das Verschneiden zweier Gebietsklassifikationen entsteht eine neue, deutlich detailliertere Gliederung. Diese können sich Leserinnen und Leser, denen beide Publikationen zur Verfügung stehen, selbst berechnen. In der Geheimhaltung wird in diesem Falle auch von Restkategorien gesprochen. Die Besetzungszahlen und die Verteilung von Einheiten müssen auch für diese neu entstehende detailliertere Gliederungssystematik ausreichend sein. Sonst dürfen aufgrund der Datenschutzerfordernisse die Ergebnisse nicht nach beiden Systematiken gleichzeitig veröffentlicht werden.

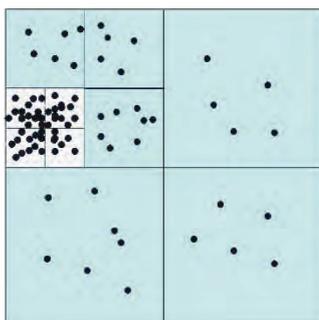
Das Risikoszenario ist in Abbildung 3 dargestellt. Werden gleichzeitig Ergebnisse für die Gebietseinheit der besiedelten Gemeindeflächen und für eine andere regionale Gliederung, hier Rasterzellen, veröffentlicht, so können bei Überlappungen dieser beiden Gliederungen durch die Differenzbildung Enthüllungsrisiken entstehen. Indem man vom

⁵ „Auf administrativen Grenzen basierende Zahlen zur Bevölkerungsdichte beziehen immer die gesamte Fläche einer regionalen Einheit mit ein, wodurch das Ergebnis von verfügbarem Raum je Einwohner unter Umständen von der Realität vor Ort abweichen kann. Das Kon-

zept der ‚EU-Orte‘ vermag somit die Realität ein Stück weit näher abzubilden, indem es die tatsächlich gewachsenen Siedlungsflächen in den Fokus rückt.“ (Heidrich-Riske et al. 2013, S. 474f.)

⁶ Die Aufzählung erhebt keinen Anspruch auf Vollständigkeit.

1 | Prinzip der hierarchischen Aggregation



2 | Ergebnis der hierarchischen Aggregation

6	5	5
14	15	
13	7	8
7		5

Quelle: Sigismund 2014

Wert der Gemeinde A den Wert der Gitterzelle 2 subtrahiert, so erhält man den Wert des Anteils der Gemeinde A in der Gitterzelle 1. In diesem Szenario würde bei zu geringer Besetzungszahl, einem Randwert oder einem dominierenden Beitragenden die Sensitivität dieses Gemeindeteils bereits auffallen, wenn die Geheimhaltungsprüfung für Rasterzelle 1 durchgeführt wird. Der Wert für Gitterzelle 1 würde bei der Verletzung einer der Geheimhaltungsregeln nicht veröffentlicht.

Ein leicht verändertes Szenario ist in Abbildung 4 dargestellt. In diesem Fall kann das Ergebnis der unabhängigen Geheimhaltungsprüfung der Gitterzellen und der Gemeinden sein, dass keine Geheimhaltungsfälle auftreten. Wurden jedoch bereits Gemeindeergebnisse veröffentlicht, so kann durch die Subtraktion des Wertes der Gemeinde B vom Wert der Gitterzelle 1 auch hier der Wert für den Teil der Gemeinde A berechnet werden, der in Gitterzelle 1 liegt. Der Wert dieses Teils der Gemeinde A entsteht als Schnittmenge der beiden regionalen Gliederungen. Für diese Schnittmenge müssten wieder ebenfalls alle Geheimhaltungsregeln geprüft werden. Ergebnisse nach beiden regionalen Gliederungen dürfen zusammen nur veröffentlicht werden, wenn in allen möglichen Schnittmengen keine Geheimhaltungsfälle auftreten.

5. Darstellungen von Verhältniszahlen in Karten

Sollen Ergebnisse von Geofachdatenauswertungen in Karten dargestellt werden, so wird oft die Einfärbung in unterschiedlichen Farbtönen gewählt. Dies entspricht dem Nachweis eines Ergebnisses in verschiedenen Kategorien, beispielsweise niedrig, mittel, hoch. Aufgrund einer begrenzten Anzahl an Farbtönen oder -schattierungen werden keine genauen Werte, sondern Intervalle, von denen in der Regel die Grenzen veröffentlicht werden, publiziert. Dies gilt natürlich nicht bei interaktiven Karten, bei denen der exakte Wert für eine geografische Einheit in der Regel angezeigt wird, z.B. als Mouseover-Effekt, sobald man mit dem Mauszeiger über die Einheit fährt oder durch das zusätzliche Anzeigen von Grafiken oder Tabellen, wenn man diese anklickt.

Grundlage der Kartendarstellung sind meist Verhältniszahlen. Das liegt darin begründet, dass viele

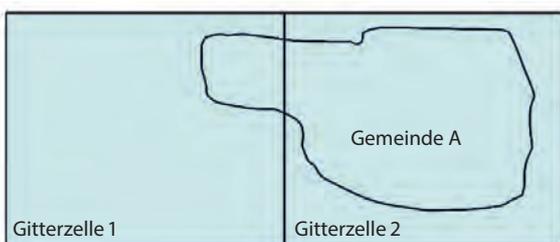
Merkmale stark korreliert sind, sodass eine Darstellung von Absolutwerten verschiedener Themen kaum einen Informationsgewinn ermöglicht. Wird beispielsweise die Anzahl an Einwohnern kartiert, so ergibt sich ebenfalls eine gleiche räumliche Struktur bei der Kartierung von Seniorinnen und Senioren, Kindern usw. Es wird letztendlich nur die Information „dünn oder dicht besiedeltes Gebiet“, „Dorf oder Stadt“ abgebildet. Die Kartierung des Anteils der Seniorinnen und Senioren an der Einwohneranzahl (Altenquotient) oder andere Anteilsinformationen sind dagegen viel aussagekräftigere Informationen. Verhältniszahlen bergen dabei als Quotient aus zwei Einzelwerten das Risiko, dass mit der Kenntnis eines Einzelwertes und einer hinreichenden Genauigkeit der Verhältniszahl auch der andere Einzelwert rückgeschlossen werden kann. Handelt es sich bei einem der Werte um eine geheim zu haltende Information, so besteht deshalb ein Geheimhaltungsrisiko. Diesem kann nur durch eine Vergrößerung oder Geheimhaltung der Verhältniszahl begegnet werden. Das Geheimhaltungsproblem für Verhältniszahlen kann auch bereits auftreten, wenn man nur sehr grobe Informationen über den einen Einzelwert hat.

Beispiel:

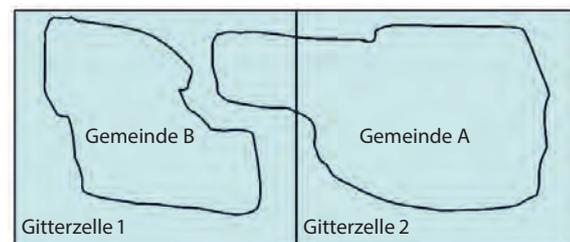
Region	Ärzte je 1000 Einwohner	Einwohner
Gemeinde C.....	0,8	ca. 1 000 bis 1 500

Wird für eine Gemeinde oder eine Rasterzelle als Kennzahl ausgewiesen, dass dort 0,8 Ärzte je 1000 Einwohner praktizieren, so kann man trotz nur sehr grober Kenntnis der Einwohnerzahl berechnen, wie viele Ärzte absolut in der Region ermittelt wurden. Obwohl mit der Veröffentlichung von Verhältniszahlen keine direkten Einzelangaben veröffentlicht werden, enthalten sie für kleine Einheiten ein großes Potenzial, einzelne Angaben zurückzuschließen. Das liegt darin begründet, dass neben der notwendigen Kenntnis von einem der beiden Werte bei vielen Merkmalen die Ganzzahligkeit der Einheiten unterstellt werden kann (z.B. Personen, Unternehmen usw.). Damit ergeben sich relativ große Intervalle, in denen noch ein Rückschluss auf die ganzzahligen Werte möglich ist. Selbst bei nicht ganzzahligen Einzelangaben ergeben sich noch hohe Risiken.

3 | Risikoszenario 1 der regionalen Differenzbildung



4 | Risikoszenario 2 der regionalen Differenzbildung



Quelle: eigene Darstellung

6. Zusammenfassung

Die Wahrung des Statistikgeheimnisses ist für die amtliche Statistik eine zentrale Aufgabe. Um den Schutz der für eine Bundesstatistik gemachten Einzelangaben zu wahren, werden die Geheimhaltungsregeln geprüft und bei Geheimhaltungsfällen ein Geheimhaltungsverfahren angewendet. Jedes Verfahren führt dabei immer zu einem Informationsverlust in den geplanten Veröffentlichungen, schützt aber dafür die Einzelangaben. Der vorliegende Beitrag stellte die wichtigsten Aspekte bei der Geheimhaltungsprüfung von Veröffentlichungen aus georeferenzierten Daten zusammen. Es wurde die Anwendung der Geheimhaltungsregeln auf geplante Publikationen von Geofachdaten beschrieben. Dabei wurde deutlich, dass es für die Mindestfallzahlregel in der Anwendung noch keine einheitliche Mindestfallzahl gibt. Neben den klassischen Geheimhaltungsverfahren der Sperrung und Vergrößerung werden in der Literatur auch datenverändernde Geheimhaltungsverfahren bei Geodaten vorgeschlagen, die in manchen Ländern bereits eingesetzt wurden. Die Besonderheiten von Verhältniszahlen in Karten und das Problem der geografischen Differenzbildung, das bei zusätzlichen Rasterkarten in noch stärkerem Maße auftritt, wurden erläutert. Dennoch wird deutlich, dass die statistische Geheimhaltung bei georeferenzierten Daten noch nicht abschließend untersucht und geklärt ist.

Weichert (2009, S. 352) stellt dabei die Forderung auf, dass „die Geodaten verarbeitenden Stellen in Verwaltung und Wirtschaft [...] durch einen sensiblen Umgang dazu beitragen [können], dass das Vertrauen in die Wahrung des Persönlichkeitsrechts nicht beschädigt wird.“ Im Bereich der wissenschaftlichen Diskussion des Datenschutzes wird des Öfteren nach Selbstverpflichtungen gerufen. Dieser bedarf es in Teilgebieten, in denen es keine rechtlichen Regelungen gibt. Im Bereich der Bundesstatistik ist diese Forderung nicht passend, da das Bundesstatistikgesetz (BStatG) für georeferenzierte Daten gleichermaßen wie bei allen anderen Daten auch gilt. Aber einheitliche und abgestimmte Handlungsempfehlungen auf Basis von methodischen Untersuchungen und rechtlichen Stellungnahmen wären für die praktische Arbeit hilfreich.

Da georeferenzierte Daten aber nicht nur in immer größeren Maße bei der amtlichen Statistik verfügbar sein werden, sondern auch von anderen Datenproduzenten erzeugt und genutzt werden, sollte die Diskussion um den „richtigen“ Datenschutz bei Auswertungen von georeferenzierten Daten auch im breiteren gesellschaftlichen Rahmen geführt werden. Gutman und Stern (2007) fordern, dass Universitäten und Fachgesellschaften sich engagieren sollten, dass Wissenschaftlerinnen und Wissenschaftler im verantwortungsvollen Umgang mit georeferenzierten Daten geschult werden. Es sollen Normen und Handlungsanweisungen zur ethisch korrekten Verwendung solcher Daten entwickelt werden.

Dr. Jörg Höhne leitet die Abteilung *Gesamtwirtschaft* im Amt für Statistik Berlin-Brandenburg. Er studierte Statistik und Wirtschaftsmathematik in Berlin und Moskau und promovierte 2009 an der Universität Tübingen mit einer Arbeit über „Verfahren zur Anonymisierung von Einzeldaten“.

Julia Höninger, Diplom-Volkswirtin, leitet das Referat *Volkswirtschaftliche Gesamtrechnungen, Erwerbstätigkeit*. Zuvor arbeitete sie als wissenschaftliche Mitarbeiterin im Referat *Mikrodaten, Analysen, Forschungsdatenzentrum* des Amtes für Statistik Berlin-Brandenburg.

7. Literatur

- Behnisch, Martin; Meinel, Gotthard; Tramsen, Sebastian; Diesselmann, Markus (2013): Using quadtree representations in building stock visualization and analysis. *Erdkunde* Vol. 67 No.2, S. 151-166.
- Dorer, Peter; Mainusch, Helmut; Tubies, Helga (1988): Bundesstatistikgesetz. Verlag C. H. Beck.
- Gutman, Myron P.; Stern, Paul C. (2007, Hrsg.): Putting people on the map – protecting confidentiality with linked social-spatial data. Issues arising from the integration of remotely sensed and self-identifying data. National Research Council of the National Academies – The National Academic Press, Washington D.C.
- Heidrich-Riske, Holger; Scholz, Bettina; Stepien, Halina (2013): GIS-gestützte Ermittlung der „EU-Orte“ im Rahmen des Zensus 2011 für die Datenlieferung an Eurostat. Statistisches Bundesamt, Wirtschaft und Statistik, Juli 2013, S. 467-475.
- Höhne, Jörg (2010): Überblick über Anonymisierungsverfahren für Mikrodaten. Kapitel 2, in: Verfahren zur Anonymisierung von Einzeldaten. Statistik und Wissenschaft, Band 16, Statistisches Bundesamt, Wiesbaden, S. 22-45. Verfügbar unter: https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band16_Anonymisierung-Einzeldaten_1030816109004.pdf?__blob=publicationFile
- Hundepool, Anco; Domingo-Ferrer, Josep; Franconi, Luisa; Giessing, Sarah; Lenz, Rainer; Naylor, Jane; Schulte Nordholt, Eric; Seri, Giovanni; de Wolf, Peter-Paul (2010): Handbook on Statistical Disclosure Control. Version 1.2, available from http://neon.vb.cbs.nl/casc/SDC_Handbook.pdf
- Interministerieller Ausschuss für Geoinformationswesen (2014): Behördenleitfaden zum Datenschutz bei Geodaten und -diensten, http://www.imagi.de/SharedDocs/Downloads/IMAGI/DE/Imagi/behoerdenleitfaden.pdf?__blob=publicationFile

- Karg, Moritz (2008): Datenschutzrechtliche Rahmenbedingungen für die Bereitstellung von Geodaten für die Wirtschaft. Gutachten im Auftrag der GIW-Kommission. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD).
- Karg, Moritz; Weichert, Thilo (2007): Datenschutz und Geoinformationen. Eine Studie im Auftrag des Bundesministeriums für Wirtschaft und Technologie (BMWi). Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD).
- Martin, David; Steinnocher, Klaus; Petri, Ekkehard (2013): Sensitivity analysis of disclosure control measures. EFGS conference, Sofia, verfügbar unter: http://www.efgs.info/geostat/workshops/efgs-2013-sofia-bulgaria/efgs-2013-conference-1/day2_ws1_8_presentation_martin
- Meyer, Werner (2011): Arbeitsergebnisse des Bundesamts für Statistik der Schweiz. Vortrag auf dem 3. Dresdner Flächensymposium 2011.
- Rat für Sozial- und Wirtschaftsdaten (2012): Endbericht der AG „Georeferenzierung von Daten“ des RatSWD – Bericht der Arbeitsgruppe und Empfehlung des Rates für Sozial- und Wirtschaftsdaten (RatSWD) http://www.ratswd.de/Geodaten/downloads/RatSWD_Endbericht_Geo-AG.pdf
- Scholz, Theresa; Rauscher, Cerstin; Reiher, Jörg; Bachteler, Tobias (2012): Geocoding of German Administrative Data – The Case of the Institute for Employment Research, FDZ-Methodenreport 9/2012, verfügbar unter: http://doku.iab.de/fdz/reporte/2012/MR_09-12_EN.pdf
- Statistics Finland (2010): Production and dissemination of grid data since the 1970 Census in Finland. Conference of European Statisticians, Fifty-eighth plenary session, Paris, 8-10 June 2010.
- Statistische Ämter des Bundes und der Länder (2006): <http://www.statistikportal.de/Statistik-Portal/qualistandards.pdf>
- Strobl, Josef (2005): Hierarchische Aggregation: Detailinformation versus Datenschutz am Beispiel adressbezogener georeferenzierter Datensätze. Salzburger Geographische Arbeiten 38, Salzburg, S. 163-171.
- Szibalski, Martin (2005): Anonymität von Erhebungseinheiten und statistische Geheimhaltung in digitalen Karten amtlicher Statistikdaten. Methoden – Verfahren – Entwicklungen Heft 2/2005, Statistisches Bundesamt, S. 5-7.
- Szibalski, Martin (2007): Kleinräumige Bevölkerungs- und Wirtschaftsdaten in der amtlichen Statistik Europas – Ergebnisse einer Umfrage zur Speicherung, Analyse und Publikation. Wirtschaft und Statistik 2/2007, Statistisches Bundesamt, S. 137-144.
- Walla, Wolfgang (2007): Standpunkt: Was ist daran geheim? Statistisches Monatsheft Baden-Württemberg 8/2007, S. 51-53.
- Weichert, Thilo (2009): Geodaten – datenschutzrechtliche Erfahrungen, Erwartungen und Empfehlungen. Datenschutz und Datensicherheit, Heft 6/2009, S. 347-252.
- Young, Caroline; Martin, David; Skinner, Chris (2009) Geographically intelligent disclosure control for flexible aggregation of census data, International Journal of Geographical Information Science, 23:4, S. 457-482, verfügbar unter: <http://dx.doi.org/10.1080/13658810801949835>

Historisches

100 Jahre Erster Weltkrieg – Berliner Statistik im Jahr 1914

von **Jürgen Paffhausen**

Im aktuellen Jahr 2014 wird auf vielfältige Art und Weise des Ausbruchs des Ersten Weltkrieges vor 100 Jahren gedacht. Der folgende Beitrag soll keine weitere Aufarbeitung dieses grausamen Ereignisses sein, das die Zeitschrift *DER SPIEGEL* in einer Spezialausgabe aus dem Jahr 2004 als die „Ur-Katastrophe des 20. Jahrhunderts“ betitelte [1]. Im Folgenden wird vielmehr kurz geschildert, was die Statistiker für das damalige Berlin trotz des entbrannten Kampfes im ersten Kriegsjahr 1914 an statistischem Material zusammengetragen und veröffentlicht haben.

Das umfassendste Werk jener Zeit ist der 33. Jahrgang des *Statistischen Jahrbuches der Stadt Berlin* [2]. Das Buch, dessen Deckblatt Abbildung a zeigt, erschien im Jahr 1916, also etwa zur Mitte des Krieges. Es enthält die Statistik der Jahre 1912 bis 1914. Auf den Krieg ging der damalige Direktor des Statistischen Amtes der Stadt Berlin, Herr Prof. Dr. Heinrich Silbergleit, im Vorwort des Jahrbuches folgendermaßen ein: „Der vorliegende Jahrgang umfaßt die Jahre 1912 bis 1914, führt aber darüber hinaus teilweise auch mitten in die Zeit des Krieges hinein... Wenn die vorliegende Veröffentlichung auch unter den schwierigen Verhältnissen des Krieges, von denen das Statistische Amt durch

militärische Einziehungen von Mitarbeitern schwer betroffen wurde, zum Abschluß gebracht werden und zu diesem beträchtlichen Umfange gedeihen konnte, so danke ich dies der Mitwirkung des Herrn Direktorialassistenten Dr. Meinerich, des wissenschaftlichen Assistenten Herrn Dr. Guradze und der Herren Statistischen Sekretäre Dr. Gehrke, Dr. Koch, Faber und Hoffmann.“ Der Umfang des Werkes ist mit 1 196 Seiten tatsächlich beachtlich. Es umfasst Daten aus allen Gesellschaftsbereichen. Die Berichterstattung beginnt mit einer umfangreichen Darstellung des Bevölkerungsbestandes und der Bevölkerungsbewegung. Die letzte Bevölkerungszahl liegt in diesem Band für den Dezember 1913 vor. Weiterhin sind Ergebnisse der Volkszählung vom 1. Dezember 1910 zu finden.

Es wurden nicht nur Daten für die Stadt Berlin zusammengetragen, sondern es wurden auch Zahlenangaben für 41 Nachbargemeinden, wie z. B. die Gemeinden Charlottenburg oder Neukölln, nachgewiesen, welche seinerzeit noch nicht zu Berlin gehörten.

Der Inhalt des Jahrbuches ist in folgende Abschnitte gegliedert:

I. Abschnitt: Stand und Bewegung der Bevölkerung.

II. Abschnitt: Naturverhältnisse.



Prof. Dr. Heinrich Silbergleit
Direktor des Statistischen Amtes der Stadt Berlin

... leitete das Amt von 1907 bis 1923 und trug wesentlich zur Vervollständigung des statistischen Arbeitsprogramms und zur Verfeinerung der Erhebungsverfahren bei. Er bemühte sich tatkräftig um das Zustandekommen der Groß-Berliner Statistik.

- III. Abschnitt: Grundbesitz und Gebäude.
 IV. Abschnitt: Öffentliche Fürsorge für Straßen und Gebäude.
 V. Abschnitt: Gewerbeverhältnisse.
 VI. Abschnitt: Märkte, Preise, Verbrauch, Verkehr.
 VII. Abschnitt: Versicherungswesen und Anstalten für Selbsthilfe.
 VIII. Abschnitt: Armenwesen, Wohltätigkeit, Kranken- und Gesundheitspflege.
 IX. Abschnitt: Polizei, Rechtspflege, Gefängnisse.
 X. Abschnitt: Anstalten und Vereine für Unterricht, Bildung und Leibesübung.
 XI. Abschnitt: Religionsverbände.
 XII. Abschnitt: Öffentliche Lasten und Rechte.

Eine weitere umfassende Veröffentlichung jener Zeit waren die Statistischen Monatshefte für Groß Berlin [3], (Abbildung b). Diese wurden unter Mitwirkung der damaligen Statistischen Ämter Charlottenburg, Neukölln, Berlin-Schöneberg und Berlin-Wilmersdorf vom Statistischen Amt der Stadt Berlin herausgegeben. Der 5. Jahrgang, Heft I/II enthält unter anderem die Bewegung der Bevölkerung Groß Berlins im Jahr 1914. Die Schrift erschien wie das Jahrbuch im Jahr 1916. In der Vorbemerkung zu dem Heft schreibt Direktor Prof. Dr. Heinrich Silbergleit: „Nach längerer, durch den Krieg herbeigeführter Unterbrechung wird die Herausgabe der Statistischen Monatsberichte mit vorliegendem Heft wieder aufgenommen. Die Berichterstattung fängt da an, wo sie im letzten Heft aufgehört hatte, d. i. in dem der Bevölkerung gewidmeten Abschnitt mit Juni 1914“.

Der Umfang der als Heft I/II zusammengefassten Monatsberichte ergibt insgesamt 99 Seiten, von denen mehr als ein Drittel der Bevölkerung gewidmet ist. Vorangestellt sind kurze Textdokumentationen zu ausgewählten Themen. Die Tabellen enthalten Angaben für das gesamte Jahr 1914 und dann beginnend mit Juni bis zum Jahresende auch für jeden einzelnen Monat. Für einige Themen, so beispielsweise für die „Arbeitslosenstatistik der Berliner Gewerkschaften“, enthalten die Tabellen auch Zahlen für Monate des Jahres 1915.

Der komplette Bericht hat folgenden Inhalt:

- I. Bewegung der Bevölkerung Groß Berlins im Jahre 1914.
- II. Einige Ergebnisse der Grundstücksaufnahme in Berlin und in den Vororten vom 15. Oktober 1910, verglichen mit den Erhebungen von Ende Oktober 1900 und 1905.

a | Deckblatt Statistisches Jahrbuch



b | Deckblatt Statistische Monatsberichte



- A. Bevölkerungsentwicklung und Bevölkerungsvorgänge.
- B. Beschäftigungsgrad und Arbeitslosigkeit.
- C. Verkehr.
- D. Preise.

Die in Abschnitt A des Tabellenteils wiedergegebenen Angaben zu den Sterbefällen enthalten allerdings keine Zahlen zu Kriegssterbefällen. Eine Ausnahme stellt die Gemeinde Charlottenburg dar. Bei ihr sind in den ausgewiesenen Gesamtzahlen der Sterbefälle, wie durch eine Fußnote angemerkt wird, auch Militärpersonen enthalten.

Zusätzlich zu den Statistischen Monatsheften hat das Statistische Amt

der Stadt Berlin auch noch Wochenberichte veröffentlicht [4], (Abbildung c). Diese mehr als DIN A4 großen Berichte umfassen zwei Seiten und sind auf die Stadt Berlin beschränkt. Der Schwerpunkt liegt auf dem Nachweis der Sterbefälle der jeweiligen Woche nach Todesursachen. Weitere Themen sind gemeldete Infektionserkrankungen, Säuglingssterbefälle, Eheschließungen und Geburten sowie Witterung, Grund- und Spreewasserstand und Erdbodentemperatur. Die Berichte erschienen für jede Woche des Jahres und wurden recht zeitnah veröffentlicht. Der erste Bericht bezieht sich auf die Woche vom 28. Dezember 1913 bis 3. Januar 1914 und wurde am 18. Januar 1914 herausgegeben. Der letzte Bericht des Jahres 1914, Deutschland beteiligte sich damals schon einige Monate am Krieg, enthält die Angaben der Woche vom 20. bis zum 26. Dezember 1914. Herausgegeben wurde er am 10. Januar 1915.

c | Erste Seite Statistischer Wochenbericht

The image shows the first page of a historical statistical report. At the top, it is titled 'Wochenberichte des Statistischen Amtes der Stadt Berlin' and specifies the week from December 28, 1913, to January 3, 1914. Below the title, there are sections for weather conditions ('Witterungsverhältnisse') and a list of deaths ('a. Sterbefälle'). The main part of the page is a large table with multiple columns representing different categories of data, likely related to infectious diseases ('b. Grundraster Infektionserkrankungen'). The table contains numerous numerical entries, some with small text annotations. At the bottom, there are additional notes and a reference to 'Statistisches Jahrbuch der Stadt Berlin, 33. Jahrgang'.

- [1] SPIEGEL Special, Heft 1/2004, SPIEGEL-Verlag Rudolf Augstein GmbH & Co. KG.
- [2] Silbergleit, Heinrich, Direktor des Statistischen Amtes der Stadt Berlin (Hrsg.): Statistisches Jahrbuch der Stadt Berlin, 33. Jahrgang, Verlag P. Stankiewicz' Buchdruckerei G. m. b. H., 1916.
- [3] Statistisches Amt der Stadt Berlin (Hrsg.): GROSS BERLIN, Statistische Monatsberichte, 5. Jahrgang, Heft I/II, 1916.
- [4] Statistisches Amt der Stadt Berlin (Hrsg.): Wochenberichte des Statistischen Amtes der Stadt Berlin, 43. Jahrgang, Nr. 1 bis Nr. 52, 1914.

Statistik erklärt

▣ „Geheimnisvolle“ Begriffe in statistischen Tabellen – Was bedeuten davon, darunter, und zwar sowie nachrichtlich?

In vielen Tabellen der Statistik stoßen Leserinnen und Leser im Tabellenkopf oder in der Vorspalte auf die Begriffe

davon, darunter, und zwar sowie nachrichtlich,

wobei davon und darunter sicher am häufigsten in Tabellen zu entdecken sind.

Doch was beinhalten diese Worte und wie sind sie einzusetzen? Laut dem Deutschen Institut für Normung empfiehlt sich die Nutzung bestimmter Begriffe in Abhängigkeit davon, ob es sich um eine Aufgliederung, Ausgliederung oder Zergliederung handelt: Hinweise hierzu lassen sich in der DIN 55301 finden. Die beiden Begriffe davon und darunter finden vor allem dann ihre Anwendung, wenn die Auf- bzw. Ausgliederungen nicht eindeutig zu erkennen sind.

Das Wort davon findet seinen Einsatz bei Aufgliederungen. Bei einer Aufgliederung (oft auch Auf- oder

Beispiele:

Jahresabwassermenge
davon
Schmutzwasser.....
Fremdwasser.....
Niederschlagswasser.....

Einwohner von Deutschland
davon
weiblich.....
männlich.....

davon ist hier fakultativ, da die Aufgliederung eindeutig ist.

Unterteilung genannt) sind alle Teilmengen des Gesamtwertes in der Tabelle genannt. Das bedeutet, dass die Summe aller Teilmengen dem Gesamtwert entspricht. Davon kommt dann zur Anwendung, wenn die Aufgliederung nicht eindeutig ist.

Der Begriff darunter ist bei Ausgliederungen in der Vorspalte bzw. im Tabellenkopf zu verwenden.

Von einer Ausgliederung wird gesprochen, wenn nur einzelne Teilmengen des Gesamtwertes aufgeführt werden. Da nicht alle Teilmengen bzw. die Differenz zum Gesamtwert nicht ausgewiesen werden, ist die Summe der aufgeführten

Teilmengen kleiner als der Gesamtwert.

Beispiele:

Nichtärztliches Personal insgesamt
darunter
Pflegedienst.....
medizinisch-technischer Dienst....
Verunglückte.....
darunter
in Personenkraftwagen.....
als Fußgänger.....

Die Worte und zwar sind bei einer sogenannten Zergliederung zu verwenden. Bei einer Zergliederung werden Teilmengen (z. B. hilfebedürftige Personen), die aus verschiedenen Gliederungen (z. B. unterschiedliche Beratungen und Hilfen) einer Gesamtmenge hervorgehen, aufgeführt. Da sich diese Teilmengen überschneiden können, ist eine Addition dieser Werte zur Ermittlung des Ingesamtwertes nicht möglich.

Ein Beispiel für eine Zergliederung wäre die Anzahl der Hilfen/Beratungen für junge Menschen, die sich unter anderem nach ambulanten Hilfen, stationären Hilfen und familienorientierten Hilfen zergliedern lassen. Da eine Person mehrere Hilfen in Anspruch nehmen kann, führt hier die Addition nicht zur Gesamtzahl der Personen.

In manchen Tabellen ist auch der Begriff nachrichtlich zu lesen. Stark vereinfacht stellt nachrichtlich eine Zusatzinformation dar.

Beispiel:

Wahlbeteiligung für alle zwölf Berliner Bezirke. Die Wahlbeteiligung für Berlin-Ost und Berlin-West lässt sich nachrichtlich ausweisen.

Mit nachrichtlich können aber auch Inhalte abgebildet werden, die beispielsweise nicht zu einer bestimmten Gruppe zählen, doch im Rahmen einer Erhebung erfasst werden.

Beispiel Personalstandstatistik:

Ohne Bezüge beurlaubte Beschäftigte und geringfügig Beschäftigte gehören nicht zum Personal im öffentlichen Dienst. Da diese Personengruppen im Rahmen der Personalstandstatistik dennoch erfasst werden, können diese nachrichtlich ausgewiesen werden.

Beim Einsatz von den Begriffen davon, darunter sowie und zwar und nachrichtlich ist zu bedenken, dass die genauen inhaltlichen Unterschiede meist nur den Nutzerinnen und Nutzern gegenwärtig sind, die sich häufiger mit Statistiken befassen. Werden in einer Tabelle gleich mehrere dieser Worte verwendet, kann dies auch verwirren, vor allem dann, wenn sich die unterschiedliche Bedeutung nicht sofort ableiten lässt.

Neuerscheinung

▣ Zweiter Kernindikatorenbericht zur nachhaltigen Entwicklung Berlins

Mit dem zweiten Datenbericht *Kernindikatoren zur nachhaltigen Entwicklung Berlins 2014* hat das Amt für Statistik Berlin-Brandenburg (AfS) in Kooperation mit der Berliner Senatsverwaltung für Stadtentwicklung und Umwelt eine aktuelle Bestandsaufnahme vorgelegt.

Nachhaltigkeit hat viele Dimensionen. Für eine umfassende Darstellung bildet der Bericht mit 16 ausgewählten Kernindikatoren 16 verschiedene Zielbereiche nachhaltiger Stadtentwicklung ab. Die Zielbereiche des Berliner Nachhaltigkeitsberichts sind Bildung, Vorschulische Erziehung, Sicherheit, Gesundheit, Sozialer Zusammenhalt, Bodenschutz, Lärmbelastung und Luftqualität, Klimaschutz, Gewässerqualität, Artenvielfalt, Wirtschaftsleistung, Beschäftigung, Innovation, Ressourcenschonung, Öffentlicher Haushalt und Chancengleichheit. Mit Hilfe der Indikatoren kann der Prozess der Zielerreichung verfolgt werden. Der Entwicklungsstand der Zielbereiche wird über eine fünfstufige Farbskala veranschaulicht. Die Befunde reichen von Rot: „Der Indikator bewegt sich in die falsche Richtung“ über Gelb: „Indikator unverändert“ bis Grün: „Der Zielwert des Indikators wird erreicht oder übertroffen“.

Zielgruppen des Datenberichts sind die interessierte Öffentlichkeit, die Fachressorts der zuständigen Berliner Senatsverwaltungen und andere Expertengruppen.

Dem Zielbereich Bildung wurde in allen seinen drei Teilindikatoren eine positive Entwicklung bzw. die Zielerreichung attestiert. Bei zwei Indikatoren der Zielbereiche Vorschulische Erziehung und Gesundheit führte die Korrektur der Bevölkerungszahlen durch den Zensus 2011 zu Brüchen in der Zeitreihe, sodass die Methodik der Trendermittlung nicht angewandt werden konnte. In den Zielbereichen Sicherheit, Sozialer Zusammenhalt, Bodenschutz, Wirtschaftsleistung, Beschäftigung, Forschungsausgaben und Ressourcenschonung wurde ein positiver Trend festgestellt. Während für den Teilindikator Verkehrslärm aus dem Zielbereich Lärmbelastung und Luftqualität eine positive Entwicklung abgeleitet werden konnte, ist die Belastung der Berliner Luft mit Stickstoffdioxid seit dem Jahr 2008 wieder im Ansteigen begriffen. Im Zielbereich Klimaschutz zeigen beide Teilindikatoren zu den CO₂-Emissionen eine Stagnation an. Im Zielbereich Chancengleichheit – gemessen als Anteil der vollzeiterwerbstätigen Frauen im Vergleich zu Männern – gibt es keinen Anhaltspunkt für zunehmende Chancengleichheit zwischen den Geschlechtern. Für den Zielbereich Artenvielfalt ergibt sich ein negativer Trend, allerdings standen keine aktuellen Daten für den Indikator zur Verfügung. Der Zielbereich Gewässerqualität bleibt ohne Trendbewertung, weil nur ein Erhebungszeitpunkt vorliegt. Gleichermassen kann der Zielbereich Öffentlicher Haushalt nach der festgelegten Methodik nicht bewertet werden, da die Zeit-

reihe innerhalb des Beurteilungszeitraumes von fünf Jahren eine deutliche Entwicklung von der negativen in die positive Richtung zeigt, die von der Berechnungsmethode nicht adäquat erfasst werden kann.

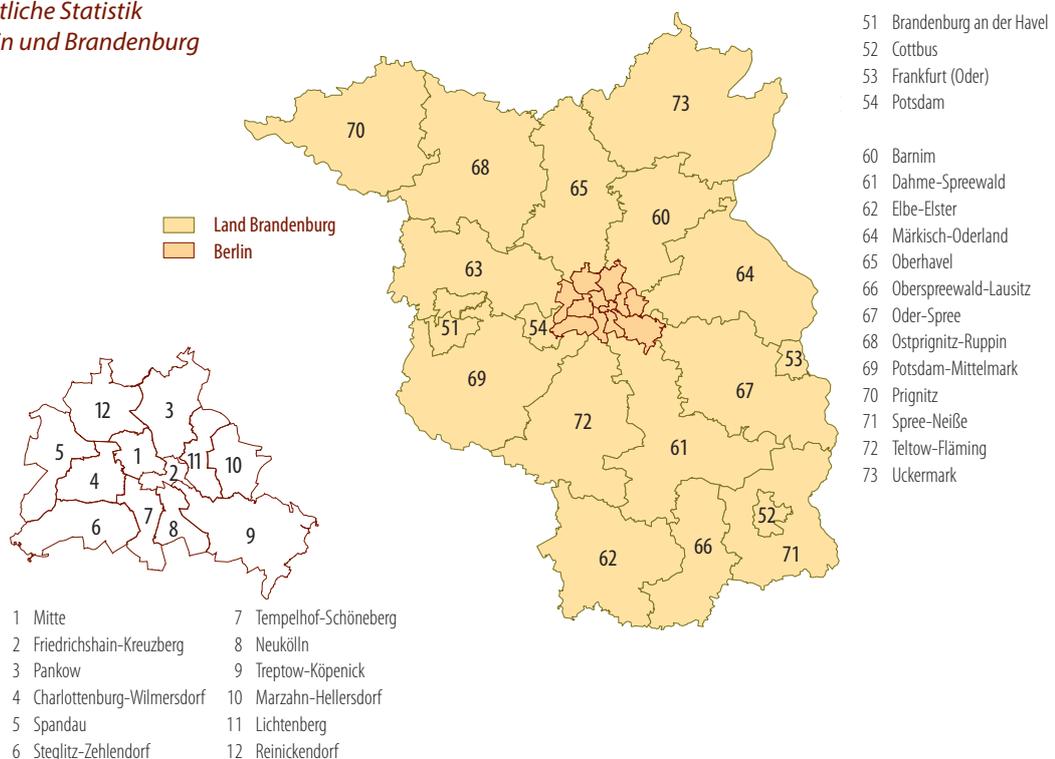


Der Kernindikatorenbericht zur nachhaltigen Entwicklung Berlins 2014 kann unter <https://www.statistik-berlin-brandenburg.de/home/kernindikatoren.asp> abgerufen werden.

statistik Berlin Brandenburg

- | Wir berichten fachlich unabhängig, neutral und objektiv über die Ergebnisse der amtlichen Statistik.
- | Wir haben den gesetzlichen Auftrag zur Datenerhebung mit der Möglichkeit zur Auskunftspflicht.
- | Wir garantieren die Einhaltung des Datenschutzes.
- | Wir wenden adäquate statistische Methoden und Verfahren an und erhöhen kontinuierlich das erreichte Qualitätsniveau.
- | Wir gewährleisten regionale und zeitliche Vergleichbarkeit unserer Statistiken durch überregionale Kooperation.
- | Wir ermöglichen jedermann Zugang zu statistischen Ergebnissen.

Wir sind der führende Informationsdienstleister für amtliche Statistik in Berlin und Brandenburg



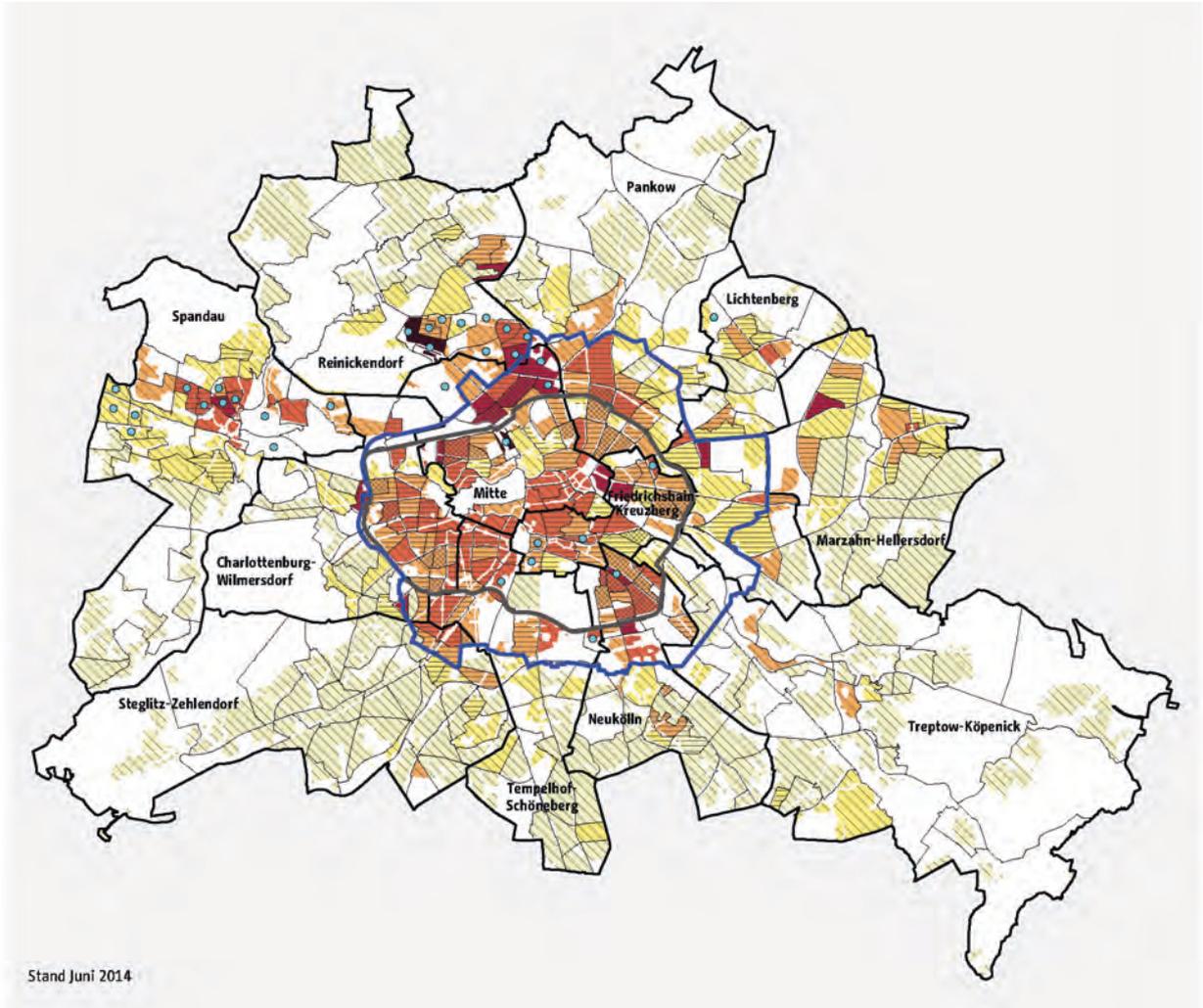
Unter

www.statistik-berlin-brandenburg.de

finden Sie einen Überblick über das gesamte Leistungsspektrum des Amtes mit aktuellen Daten, Pressemitteilungen, Statistischen Berichten, regionalstatistischen Informationen, Wahlstatistiken und -analysen.

Umweltgerechtigkeit im Land Berlin 2014

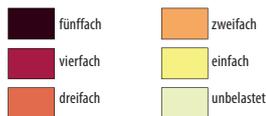
Integrierte Umweltbelastung einschließlich Sozialer Problematik



Senatsverwaltung für Stadtentwicklung und Umwelt



Mehrfachbelastung durch die Kernindikatoren Lärmbelastung, Luftschadstoffe, Grünflächenversorgung, Bioklimatische Belastung, Soziale Problematik (Statusindex nach Monitoring Soziale Stadtentwicklung 2013)



Die Kernindikatoren fließen als Einzelbelastung in die Bewertung ein, wenn sie im Hinblick auf die planungsraumbezogene Belastung nach der jeweiligen 3er-Klassifikation der schlechtesten Kategorie zugeordnet werden.

Planungsräume mit überwiegender einfacher Wohnlage (mehr als 66% der Adressen) sowie sehr hoher Lärm- und/oder sehr hoher Luftbelastung



weitgehend unbewohnte Fläche

Grenze Planungsraum (PLR)

Grenze Bezirk/Land

Schwerpunktbereich Innenstadt (Vorranggebiet Luftreinhalteung gemäß FNP)

Grenze Umweltzone

Die Auswertung der Daten erfolgte auf der räumlichen Ebene der 447 Planungsräume Berlins (kleinräumigste Ebene der „Lebensweltlich orientierten Räume“ - LOR).



Herausgeber: Senatsverwaltung für Stadtentwicklung und Umwelt

Konzeption Inhalte und Bearbeitung: Senatsverwaltung für Stadtentwicklung und Umwelt, Abteilung IX Umweltpolitik, Abfallwirtschaft und Immissionsschutz

In Zusammenarbeit mit: Planergemeinschaft Kohlbrenner eG

Datengrundlage: -TU Dresden (Lärm, Stand 2012)
-Helmholtz-Zentrum für Umweltforschung GmbH (Luftbelastung, Stand 2009)
-SRP GmbH (Grünversorgung, Stand 2012)
-Universität Kassel (Thermische Belastung, Stand 2012)
-Monitoring Soziale Stadtentwicklung 2013 (MSS), Senatsverwaltung für Stadtentwicklung und Umwelt, Datenstand 31.12.2013
-Amt für Statistik Berlin-Brandenburg (Einwohnerdaten 31.12.2013)

Kartengrundlage: -Lebensweltlich orientierte Räume (LOR), Stand Oktober 2006 (Ausgabe 2011)
-Übersichtskarte 1:50 000, Ausgabe 5/2013