

### Herausforderungen und mögliche Lösungsansätze bei der Kombination statistischer und räumlicher Daten aus Sicht der Bundesstatistik

Schnorr-Bäcker, Susanne; Etienne, Malte

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

#### Empfohlene Zitierung / Suggested Citation:

Schnorr-Bäcker, S., & Etienne, M. (2018). Herausforderungen und mögliche Lösungsansätze bei der Kombination statistischer und räumlicher Daten aus Sicht der Bundesstatistik. *Stadtforschung und Statistik : Zeitschrift des Verbandes Deutscher Städtestatistiker*, 31(1/2018), 63-69. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-56845-1>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

Susanne Schnorr-Bäcker, Malte Etienne

# Herausforderungen und mögliche Lösungsansätze bei der Kombination statistischer und räumlicher Daten aus Sicht der Bundesstatistik

*Der Bedarf an kleinräumigen Auswertungen aus der Bundesstatistik steigt stetig, besonders angetrieben durch ein datengestütztes Monitoring politischer Strategien und dem damit verbundenen Ziel, räumliche Ungleichheiten zu identifizieren. Die Weiterentwicklung des kleinräumigen Datenangebots in der Bundesstatistik ist im Rahmen der Integration statistischer und räumlicher Daten mit einigen Herausforderungen verbunden. Der Aufsatz zeigt ausgehend von einer Prozessbeschreibung der statistischen Datenproduktion mögliche Lösungsansätze hierfür auf. Im Fokus stehen dabei der Geodatenzugang, die statistische Geheimhaltung und Sicherstellung wichtiger Qualitätsanforderungen sowie die Berücksichtigung von Interoperabilitätsanforderungen.*

## Dr. Susanne Schnorr-Bäcker

Bis zum Eintritt in den Ruhestand am 01.01.2018 Leiterin des Referats Koordinierung, Regionalstatistik, Indikatoren in der Abteilung Strategie und Planung, Internationale Beziehungen, Forschung und Kommunikation im Statistischen Bundesamt. Themenschwerpunkte: Politikberatung für ein statistisches Monitoring, zuletzt die Regierungsstrategie „Gut leben“, „Europa 2020“ der Europäischen Kommission, „How's Life“ bzw. die Better Life-Initiative der OECD sowie bereichsübergreifende Aspekte des Ausbaus des kleinräumigen bundesstatistischen Datenangebots auf nationaler und internationaler Ebene.

✉ Schnorr-Baecker@gmx.de

## Malte Etienne

M. A. Sozialwissenschaft, Statistisches Bundesamt, seit 2017 Referent im Referat Europäische und internationale Koordinierung, davor wiss. Mitarbeiter im Referat Koordinierung, Regionalstatistik, Indikatoren.

✉ Malte.Etienne@destatis.de

## Schlüsselwörter:

Bundesstatistik – Georeferenzierung – Datenintegration – Qualität – Interoperabilität

## 1 Einleitung<sup>1</sup>

Kleinräumige Statistiken bekommen ein immer größeres Gewicht, vor allem wenn regionale Ungleichheiten identifiziert und diesen entgegengewirkt werden soll – so wie es das Ziel richtungsweisender politischer Strategien auf nationaler, europäischer und internationaler Ebene ist, etwa „Gut leben“ in Deutschland, „Europa 2020“ in der Europäischen Union oder den „Sustainable Development Goals“ der Vereinten Nationen (vgl. Schnorr-Bäcker 2017). Um den steigenden Anforderungen nach kleinräumigen Statistiken gerecht zu werden, muss die Bundesstatistik in den kommenden Jahren auf diese Entwicklungen reagieren (vgl. Thiel u. Meinke 2017: 14). Die dabei auftretenden Herausforderungen und möglichen Lösungsansätze werden im Folgenden näher beschrieben. Dazu wird zunächst das aktuelle Programm an kleinräumigen Daten aus der Bundesstatistik kurz präsentiert. Anschließend wird auf die zentralen Schritte bei der Integration statistischer und räumlicher Daten eingegangen. Die dabei auftretenden Herausforderungen sowie möglichen Lösungsansätze werden im nächsten Abschnitt näher beschrieben. Abschließend wird ein kurzer Ausblick vorgenommen.

## 2 Regionale Datenveröffentlichungen in der Bundesstatistik

Das Programm der Bundesstatistik, das für Deutschland von den Statistischen Ämtern des Bundes und der Länder veröffentlicht wird, umfasst knapp 400 Statistiken von denen ca. zwei Drittel EU-weit harmonisiert sind. Länderübergreifende Analysen europäischer Mitgliedsländer sind damit für weite Bereiche möglich. Die statistischen Produkte werden hauptsächlich in zwei grundlegenden Datenbanken präsentiert: in GENESIS-Online<sup>2</sup> können statistische Informationen (Tabellen, Grafiken, einige Karten) für über 220 Statistiken vor allem auf Ebene der Bundesländer (NUTS-1 Einheiten)<sup>3</sup> abgerufen werden. Zudem existiert die Regionaldatenbank Deutschland<sup>4</sup>, einem Angebot der Statistischen Ämter des Bundes und der Länder, in der über 90 Statistiken – mehrheitlich basierend auf administrativen Quellen – für die niedrigste NUTS-Ebene (NUTS-3: Kreise und kreisfreie Städte für Deutschland) geführt werden. Davon bieten 20 Statistiken statistische Nachweisungen sogar auf Gemeindeebene für Deutschlands mehr als 11.000 Gemeinden an. Die Regionaldatenbank ist zugleich die

Datenquelle für den Regionalatlas<sup>5</sup> mit seinen Choroplethenkarten für ca. 140 statistische Indikatoren. Für jeden Indikator sind Werte und Zeitreihen nach Jahren, Tabellen und Karten abrufbar. Diese Online-Anwendung, die nach Open-Government-Standards – so wie alle Datenveröffentlichungen aus der Bundesstatistik entwickelt wurde, ist grundsätzlich für jeden Nutzer frei verfügbar und gebührenfrei zugänglich. Daneben existieren weitere Atlanten (z. B. Außenhandel, Luftverkehr, Bundes- und Europawahlen, Zensus).

Für den Agrar- und den Bevölkerungszensus stehen Kartenviewer online zur Verfügung, die statistische Informationen in Rasterkarten bereitstellen. Während gitterbasierte Darstellungen in der geographischen Community bereits weit verbreitet sind, sind diese zumindest in der deutschen amtlichen Statistik – seit dem Inkrafttreten einer Änderung im Bundesstatistikgesetz<sup>6</sup> aus dem Jahr 2013 – mittlerweile grundsätzlich erlaubt für ökonomische und ökologische Statistiken und mit strengeren Vorgaben bezüglich der Datengeheimhaltung auch für die Haushalts- und Personenstatistiken. Rasterbasierte Daten besitzen für die amtliche Statistik einige Vorteile im Vergleich zu administrativen Gebietseinheiten auf allen Ebenen. Bei Gebietsstandsänderungen auf subnationaler Ebene, beispielsweise durch den Zusammenschluss von Gemeinden, ist es kostenintensiv und aufwendig, die statistischen Daten entsprechend der neuen oder alten Gebietsabgrenzungen zu überarbeiten. Vergleiche von statistischen Daten durch die Verknüpfung administrativer oder nicht-administrativer Gebietseinheiten werden zunehmend schwieriger. Dem hingegen bleiben Geocodes für rasterbasierte Daten im Laufe der Zeit konstant.

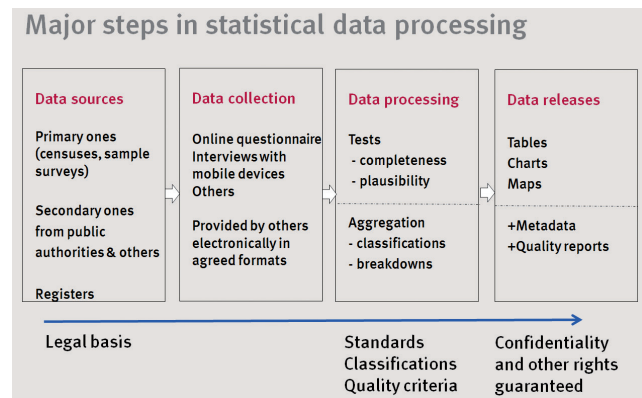
### 3 Integration räumlicher und statistischer Daten

In Deutschland existiert keine kodifizierte allgemeine Strategie für die Integration von statistischen und räumlichen Daten, wie dies z. B. in Australien, Mexiko oder Neuseeland der Fall ist (vgl. UN-GGIM 2016). Dennoch spielt die Strukturierung der dafür notwendigen Prozesse der Datenintegration für die laufenden Arbeiten eine große Rolle. Da das Statistische Bundesamt für die Methodik im statistischen Verbund in Deutschland zuständig ist, ist ein Bottom-up-Ansatz derzeit von besonderem Interesse. Nach dem Geschäftsprozessmodell Amtliche Statistik (GMAS), das als erweiterte praktische Anwendung des auf internationaler Ebene entwickelten Generischen Statistischen Geschäftsprozessmodell (GSBPM) in der amtlichen Statistik in Deutschland eingesetzt wird (Statistisches Bundesamt 2017: 14 f.), ist die Definition und Entwicklung von standardisierten Prozessen für die Integration von statistischen und räumlichen Daten derzeit von schwerpunktmäßiger Bedeutung.

Grundsätzlich können vier verschiedene Phasen unterschieden werden.

In einem ersten Schritt wird die Datensammlung und -gewinnung aus verschiedenen Quellen festgelegt. Es können die folgenden drei Hauptdatenquellen für die Bundesstatistik unterschieden werden: *Primärquellen* wie z. B. passgenaue

Abbildung 1: Wesentliche Schritte bei der Datenproduktion in der deutschen amtlichen Statistik



Quelle: Schnorr-Bäcker (2014a), Folie 4

selbst entwickelte Umfragen – oftmals in Übereinstimmung mit europäischen Verordnungen – wie etwa dem Labour Force Survey. Weitere Quellen sind *statistikinterne Register* wie z. B. das Unternehmensregister. Heutzutage sind *Sekundärquellen* bzw. Daten anderer öffentlicher Einrichtungen von steigender Bedeutung. Fast die Hälfte des statistischen Programms wird überwiegend durch solche Verwaltungsdaten abgedeckt (vgl. Statistisches Bundesamt 2017: 44). Zukünftig könnten zudem neue digitale Daten, auch als Big Data bezeichnet, beispielsweise Mobilfunkdaten für die Bevölkerungs-, Tourismus- und Transportstatistiken oder Satellitenbilder für die Umweltstatistiken zur Messung der Feinstaubbelastung, als Quellen an Bedeutung gewinnen (Schnorr-Bäcker 2016a: 8 ff.).

Die zweite Phase bezieht sich auf die konkrete Datensammlung. Nachdem die Basis-Informationen über die jeweiligen Einheiten für eine spezifische Statistik gewonnen wurden, müssen die Daten anschließend qualitativ aufbereitet werden. Die meisten Einzeldaten über Personen und Haushalte, Gebäude und Unternehmen besitzen eine Adresse als Hilfsmerkmal für die Vorbereitung und Durchführung der Erhebung. Für jede Erhebungseinheit wird geprüft, ob sie existiert und ob die postalische Adresse<sup>7</sup> korrekt ist. Die postalische Adresse stellt in all ihren Dimensionen (Name, Straßename und Nummer, Postleitzahl und Name der Gemeinde) die Grundlage für die Georeferenzierung dar. Ebenso für statistikinterne Register kann die Adresse geocodiert werden und permanent in statistikinternen Registern wie etwa dem Unternehmensregister gespeichert werden.<sup>8</sup> Für die Darstellung rasterbasierter Daten wird ein georäumliches Referenzsystem benötigt; für die EU ist dies in der INSPIRE-Richtlinie festgelegt. Diese wurde national umgesetzt in den sogenannten Geodatenzugangsgesetzen des Bundes<sup>9</sup> und der jeweiligen Länder. Für die Bereitstellung hochwertiger Geodaten, wie z. B. digitaler Landschaftsmodelle, Gebäudepolygonen oder georeferenzierter Adressdaten, ist auf Bundesebene das Bundesamt für Kartographie und Geodäsie die zentrale Einrichtung.

Sobald die Basisdaten für eine bestimmte Statistik vollständig verfügbar und plausibilisiert sind, kann die Datenaggregation

durchgeführt werden. Das bedeutet, dass die Basisdaten nach generell geltenden Klassifikationen der amtlichen Statistik klassifiziert und berechnet werden. Diese stimmen mit europäischen und internationalen Klassifikationen überein (z. B. bezogen auf wirtschaftliche Aktivitäten, Berufe, Krankheiten etc.). Ebenfalls von Bedeutung sind Gebietsklassifikationen. Neben nationalen (Fach-)Klassifikationen sind besonders die NUTS-Klassifikation für Europa und die Klassifikation des Grads der Urbanisierung DegUrba (degree of urbanization), die gemeinsam von der OECD und dem europäischen Statistikkamt entwickelt wurden, als regionale Klassifikationen zu nennen (vgl. für weitere Beispiele auch Schnorr-Bäcker 2014b).

Die letzte Prozessphase umfasst die Entwicklung von Tabellen, Grafiken und Karten. Hierbei müssen gründliche Prüfungen durchgeführt werden, um eine Geheimhaltung zu garantieren. Zur Prüfung, ob die Geheimhaltung verletzt werden könnte, werden verschiedene Methoden und Verfahren eingesetzt. In den entsprechenden Fällen müssen die statistischen Daten entfernt werden oder in geeigneter Weise bearbeitet werden, sodass eine Geheimhaltung bei der Veröffentlichung und Verbreitung der Ergebnisse garantiert ist. Die finalen Publikationen beinhalten nicht nur Informationen bezüglich der Metadaten, also Erklärungen über einzelne Items, die gezeigten Indikatoren oder die Form, in der die Datengewinnung erfolgte, sondern auch wichtige Hinweise über die Qualität der statistischen Nachweisungen. Die Produktion regionaler Bundesstatistiken wird gemeinsam vom Statistischen Bundesamt in Zusammenarbeit mit den statistischen Ämtern der Länder

bewerkstelligt. Dabei werden die verschiedenen Dimensionen von Qualität berücksichtigt. Eine klare und präzise Definition und Beschreibung von allen Arbeitsschritten in jeder Phase stellt sicher, dass der produzierte Output die erforderlichen Qualitätsstandards erfüllt.

#### 4 Herausforderungen und mögliche Lösungen

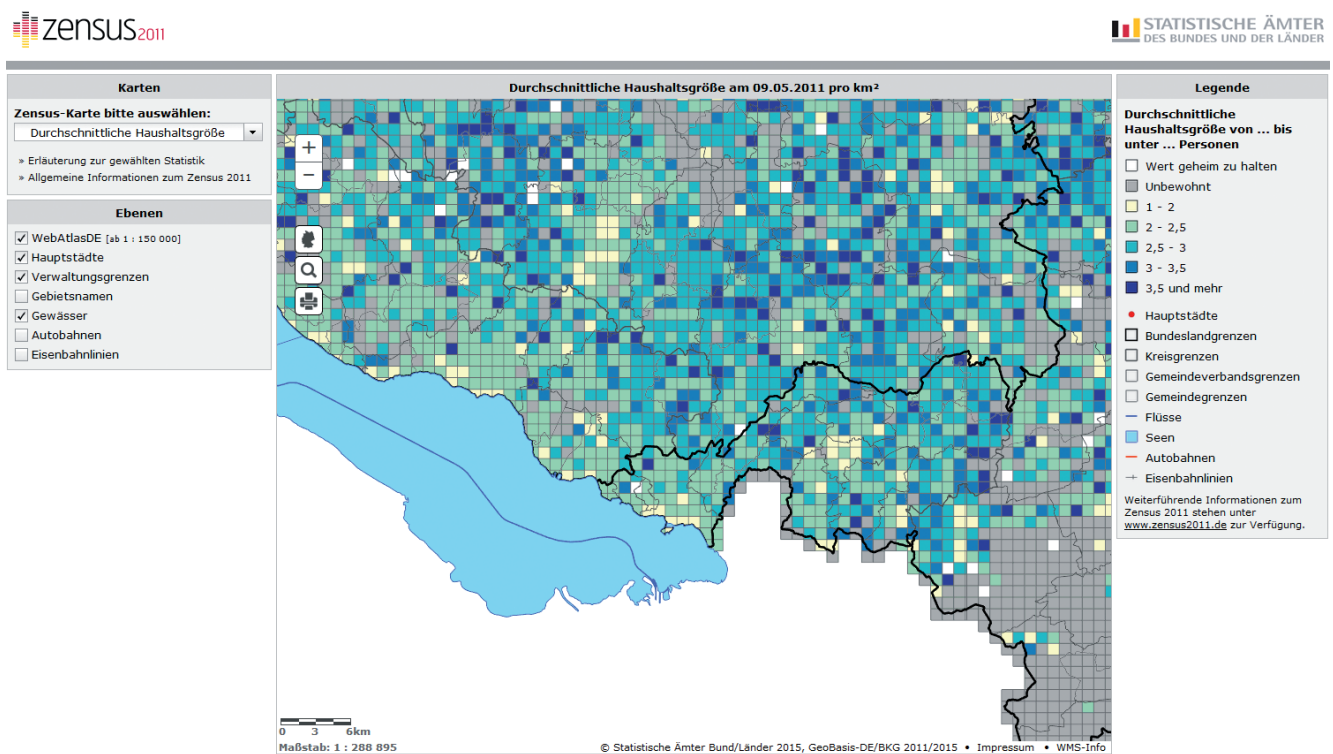
Aus Sicht der Produzenten der deutschen amtlichen Statistik sind die folgenden Aspekte von besonderer Bedeutung:

- (1) Zugänglichkeit und Nutzbarkeit
- (2) Geheimhaltung und Schutz der Privatsphäre/Datenschutz
- (3) Qualitätsanforderungen
- (4) Interoperabilität

##### ad (1) Zugänglichkeit und Nutzbarkeit

Die amtliche Statistik in Deutschland verbreitet statistische Daten nach Open-Government-Standards. Anders verhält es sich für räumliche Daten, die vom Bundesamt für Kartographie und Geodäsie und den zuständigen Vermessungsämtern auf Länderebene bereitgestellt werden. Vor allem, wenn Hauskoordinaten oder Geocodes in der amtlichen Statistik benötigt werden, z. B. im Rahmen der Georeferenzierung von statistischen Daten, wird für die Kooperation eine rechtliche Vereinbarung benötigt. Denn für die Datensammlung und -weiterverarbeitung sind die Länder in ihren Hoheitsgebieten zuständig. Hinzu kommen Lizenzierungsfragen in ihren

Abbildung 2: Territoriale Grenzen schneidende Rasterzellen



Quelle: <https://atlas.zensus2011.de/>



unterschiedlichen Ausprägungen, d. h. insbesondere auch bezüglich möglicher Gebühren. Allerdings kann dieses Problem auch länderübergreifend mit europäischen Nachbarländern auftreten. So ergibt sich bei auf Geocodes basierenden Statistiken das Problem von Rasterzellen, die nationale Grenzlinien schneiden.

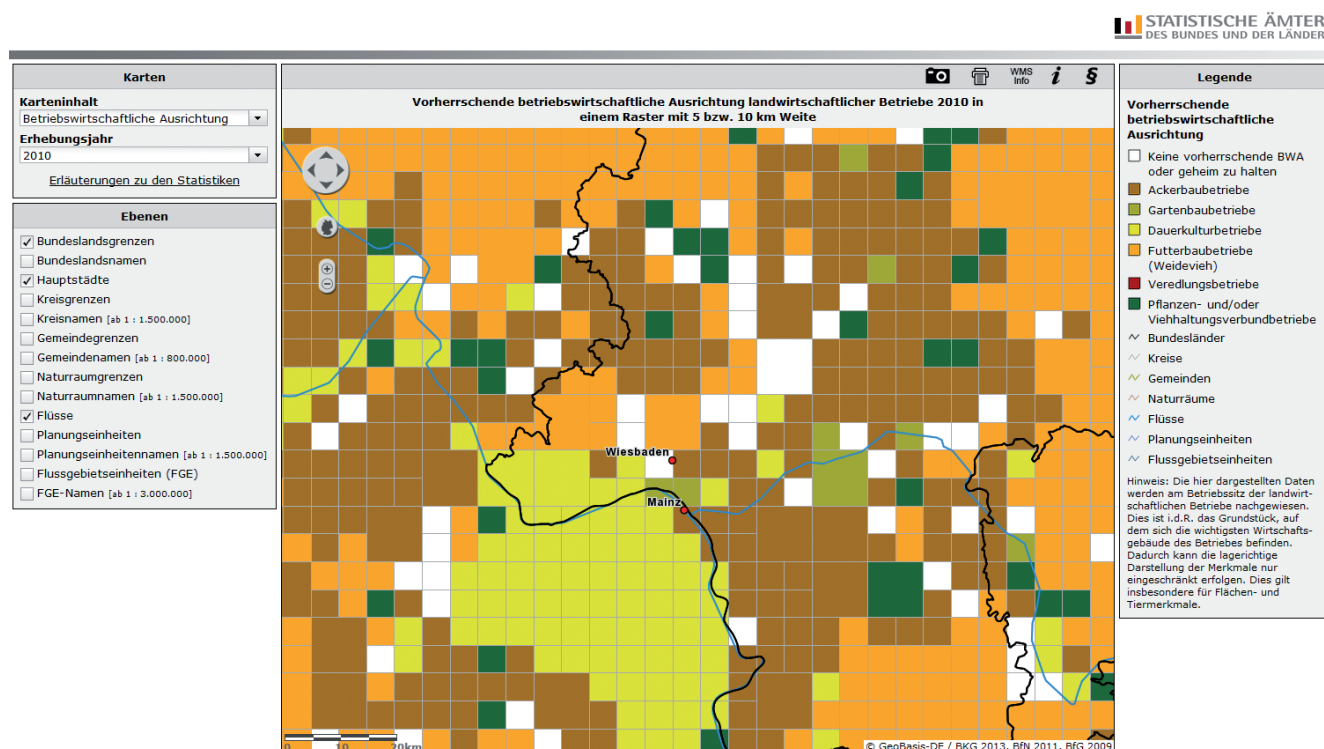
Abbildung 2 zeigt für einen südlichen Teil von Deutschland die durchschnittliche Haushaltsgröße pro Quadratkilometer aus dem Bevölkerungs-Zensus 2011. Auf der Abbildung erkennbar sind innerhalb Deutschlands der Grenzverlauf zwischen den Bundesländern Bayern und Baden-Württemberg sowie die Landesgrenzen zu Österreich und der Schweiz (inklusive eines Ausschnitts des Bodensees). Die im Hintergrund erkennbaren grauen Grenzverläufe repräsentieren die Gebietsgrenzen auf NUTS-3 Ebene (Kreise und kreisfreie Städte). Diese unterstützen die Nutzerinnen und Nutzer bei der Lokalisierung der Verwaltungsgrenzen. Für zusätzliche Analysen können auch die Gebietsnamen sowie Autobahnen oder Eisenbahnlinien eingeblendet werden. Zur Darstellung eines vollständig rasterbasierten Abbilds dieser Region ist bereits am Anfang des Produktionsprozesses ein Datenaustausch der georeferenzierten Rohdaten für die grenzüberschreitenden Rasterzellen notwendig. Sowohl auf nationaler als auch auf europäischer Ebene stellen z. B. Verwaltungsvereinbarungen zwischen den statistischen und den geodatenhaltenden Stellen einen Lösungsansatz dar. Hinzu kommt am Ende des Produktionsprozesses eine Prüfung auf Geheimhaltung, wiederum grenzüberschreitend.

### ad (2) Geheimhaltung und Schutz der Privatsphäre/Datenschutz

Für die amtliche Statistik sind Datenschutz und statistische Geheimhaltung von herausragender Bedeutung; sie sind z. B. im Verhaltenskodex für Europäische Statistiken festgelegt (vgl. Eurostat 2011). Im Rahmen der Verbreitung statistischer Informationen auf den traditionellen Kommunikationskanälen auf der Grundlage von Tabellen oder Grafiken müssen verschiedene Geheimhaltungsregelungen berücksichtigt werden. In der amtlichen Statistik wird hierbei zwischen primärer und sekundärer Geheimhaltung unterschieden. Bei der Integration statistischer und räumlicher Daten stellen diese Regeln nichts Neues dar, allerdings gilt der Satz: je kleinräumiger statistische Daten ausgewiesen und dargestellt werden, desto dringlicher wird das Problem des Aufdeckungsrisikos. Entsprechend wurden Richtlinien und Methoden für den Datenschutz entwickelt – allgemeine ebenso wie statistisch spezifische –, die in der amtlichen Statistik in Deutschland angewandt werden (vgl. z. B. Ronning et al. 2005). Im Hinblick auf gitterbasierte Darstellungen müssen allerdings weitere Aspekte berücksichtigt werden, wie im folgenden Beispiel in Abbildung 3 illustriert.

Abbildung 3 zeigt die vorherrschende betriebswirtschaftliche Ausrichtung landwirtschaftlicher Betriebe (z. B. Ackerbau, Gartenbau, Dauerkultur, etc.) aus dem Atlas Agrarstatistik in der Grenzregion zwischen den Bundesländern Hessen und Rheinland-Pfalz (die Grenzlinie ist in dieser Region nahezu identisch mit dem Verlauf des Rheins, siehe die sich überwiegend überschneidende blaue und schwarze Linie im

Abbildung 3: Vorherrschende betriebswirtschaftliche Ausrichtung landwirtschaftlicher Betriebe, 2010



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

unteren Drittel). Bezüglich der statistischen Geheimhaltung ist es entscheidend, dem Aufdeckungsrisiko von jedweden zusätzlichen Informationen, die das Betriebsgeheimnis eines landwirtschaftlichen Betriebs gefährden und somit zu einer Schädigung der Geschäftstätigkeiten führen könnten, vorbeugend entgegenzuwirken.

Die Herausforderung liegt darin, das Potential gitterbasierter Darstellungen bestmöglich auszuschöpfen, d. h. einen vertieften Einblick auf kleinräumige Gebiete und die sich dort vollziehenden Prozesse und Entwicklungen zu ermöglichen. Hierfür existieren einige Ansatzpunkte. Zum einen ist es möglich, variierende Rastergrößen innerhalb eines rasterbasierten Atlas einzusetzen. Das bedeutet, dass der Atlas Agrarstatistik primär Gitterzellen mit einer 5 km-Weite ausweist. Eine generelle Verwendung dieser Skala würde einige Geheimhaltungsfälle hervorrufen, d. h. die Rohdaten in den jeweiligen Gitterzellen müssten geheim gehalten werden. Die Lösung des Problems besteht in der Aggregation von Rasterzellen, sodass in manchen Fällen eine Gitterweite von 10 km angemessen ist. Insgesamt bedeutet dies, dass der Detaillierungsgrad des Atlas grundsätzlich höher ist, auch wenn dafür einige Verluste durch die variierenden Rastergrößen hinsichtlich der Vergleichbarkeit innerhalb der Karte hinzunehmen sind. Hinsichtlich der Visualisierung ist es bei variierenden Rastergrößen allerdings wichtig, die Umrisse der jeweiligen Zelle einzublenden, da in der Karte sonst in manchen Fällen die Besetzungszahlen einer Zelle oder eines Zellverbundes nicht zu unterscheiden sind.

Neben diesem Visualisierungsansatz variierender Gitterzellen können einige eher grundsätzliche Aspekte in den Fokus gerückt werden. Erstens kann durch die Verwendung von *statistischen Indikatoren*, die z. B. auf Quotienten basieren, das Aufdeckungsrisiko (d. h. die Möglichkeit, Werte wieder einzelnen Auskunft gebenden Unternehmen zuzuweisen) bereits entscheidend reduziert werden. Zweitens nimmt das Risiko durch die vornehmliche Verwendung *breiter Größenklassen* für Wertbereiche von statistischen Indikatoren weiter ab. Drittens ist es ratsam, nur solche statistischen Kennzahlen darzustellen, für die sichergestellt ist, dass sie *nicht bereits anderweitig* in Tabellen oder ähnlichen Darstellungsformaten veröffentlicht wurden.

### ad (3) Qualitätsanforderungen

Qualität, in all ihren Dimensionen, ist für die amtliche Statistik traditionell nicht nur in Deutschland, sondern auch weltweit bedeutsam, auch wenn Qualitätsanforderungen lange Zeit nicht detailliert in schriftlicher Form und formal niedergelegt wurden (vgl. Schnorr-Bäcker 1986). Im Jahr 1999 etablierte das Statistische Bundesamt eine „Qualitätsoffensive Statistik“ in umfassender und ganzheitlicher Weise in Form eines „Total Quality Management Systems“ und entwickelte in den Folgejahren nach diesem Startpunkt gemeinsam mit den statistischen Ämtern der Länder Maßnahmen und Instrumente für die deutsche amtliche Statistik.<sup>10</sup> Auch sind seitdem mit Blick auf den oben genannten Verhaltenskodex für Europäische Statistiken in das Europäische Statistische System Qualitätskriterien in einer stärker formalisierten Weise eingeführt worden, die zugleich eine Basis für Peer Review-Verfahren darstellen.

Der Verhaltenskodex wurde erstmals im Jahr 2005 eingeführt und im Jahr 2011 erneuert. Er umfasst 15 Grundsätze, von denen einige direkt auf den statistischen Produktionsprozess und weitere auf die Datenangebote aus der amtlichen Statistik bezogen sind. Mit Blick auf die Integration statistischer und räumlicher Daten erscheint dabei eine nähere Betrachtung folgender Grundsätze besonders relevant

- Geeignete Methoden und Verfahren (Grundsätze 7–10)
- Relevanz (Grundsatz 11)
- Genauigkeit und Zuverlässigkeit (Grundsatz 12)
- Aktualität und Pünktlichkeit (Grundsatz 13)
- Kohärenz und Vergleichbarkeit (Grundsatz 14)
- Zugänglichkeit und Klarheit (Grundsatz 15)

Grundsätzlich müssen **geeignete Methoden und Verfahren** im statistischen Produktionsprozess zum Einsatz kommen. Das bedeutet zusammengefasst, dass alle statistischen Ämter bei der Vorbereitung, Sammlung, Weiterverarbeitung und Verbreitung von Daten geeignete Werkzeuge einsetzen, die in effizienter und effektiver Weise eine Bereitstellung qualitativ hochwertiger statistischer Daten sicherstellen.

Der Grundsatz **Relevanz** sieht vor, dass europäische Statistiken den Nutzerbedarfen entsprechen. Neben weiteren Aspekten ist diese Zielsetzung eng mit der Aufgabe verknüpft, die Bandbreite an statistischen Produkten inhaltlich aktuell zu halten, d. h. die informationelle Übersicht über wesentliche Neuentwicklungen in Gesellschaft, Wirtschaft und Umwelt zu behalten. Damit ist auch gemeint, den steigenden Bedarf unterschiedlicher gesellschaftlicher Interessensgruppen sowohl in Bezug auf Regionalstatistiken in ihrer Gesamtheit als auch kleinräumigen Daten im Besonderen in Betracht zu ziehen. Darüber hinaus ist die amtliche Statistik in der Pflicht, zu umfassenden Geodateninfrastrukturen (GDI) auf nationaler, europäischer und internationaler Ebene, wie z. B. INSPIRE auf europäischer Ebene, GDI-DE für Deutschland oder international zur Initiative der Vereinten Nationen im Bereich Globales Geoinformationsmanagement (UN-GGIM)<sup>11</sup> beizutragen. Im Hinblick auf Geobasisdaten bedeutet „Relevanz“, dass diese bezüglich ihrer Granularität in geeigneter Weise bereitgestellt werden, sodass sie mit statistischen Daten verknüpft werden können.

Damit kann zugleich der Grundsatz **Genauigkeit und Zuverlässigkeit** erfüllt werden. Vor allem, wenn keine weiteren Aspekte von Belang sind, können beide Punkte eng miteinander verknüpft werden. Zur näheren Klärung der Bedeutung muss der zu erfüllende Zweck, d. h. müssen die beabsichtigte Verwendung und/oder die Nutzerbedürfnisse ebenfalls einbezogen werden. Einige Statistiken müssen in einer höheren Qualität, größeren Detailliertheit oder höheren Granularität als andere produziert werden. Beispielsweise werden im Rahmen von Tarifverhandlungen höchst genaue statistische Daten benötigt. Dagegen wird für auf Basis von Makrodaten vorgenommener Anpassungen des verfügbaren Einkommens die nötige Genauigkeit eher etwas niedriger ausfallen (z. B. Schätzungen basierend auf Makrodaten), denn der Einsatzzweck und die eingesetzten Methoden erlauben dies (vgl. Statistische Ämter des Bundes und der Länder 2005).

Im Grundsatz **Aktualität und Pünktlichkeit** bezieht sich Erstere eher auf den Produktionsprozess und Letztere eher auf den regulären Veröffentlichungstermin, der in der Regel im Voraus feststeht. Bei der Integration statistischer und räumlicher Daten ist ein weiterer Aspekt ebenfalls von Belang. Es muss festgelegt werden, welche Geometrien, i. d. R. Karten, gezeigt werden sollen, vor allem bei Zeitreihen mit sich verändernden administrativen Grenzverläufen im Zeitverlauf. In diesem Fall erscheinen die Nutzerbedürfnisse besonders relevant, wie im Falle der EU, die auf Grundlage von statistischen Daten europäische Fördermittel verteilt sowie deren lokale und regionale Auswirkungen misst. Die regulären NUTS-Veränderungen alle drei Jahre, die mit einer gewissen Zeitverzögerung zur Anwendung kommen, können hierfür als Beispiel genannt werden. Zumindest für Deutschland kann festgehalten werden, dass die auf europäischer Ebene verwendeten Geometrien oftmals nicht den aktuellen Gebietsstand berücksichtigen, vor allem aufgrund verschiedener zwischenzeitlicher Änderungen auf administrativer Ebene. Im Gegensatz dazu werden im Rahmen der EU-Kohäsionsförderung und ihres Monitoring die NUTS-Grenzen über die Zeit hinweg beibehalten um vergleichbare Informationen zu erhalten. Ein geeigneterer Ansatz könnte in der Verwendung von rasterbasierten Daten liegen, mit im Zeitablauf gleichbleibenden Rasterweiten. Diese könnten flexibel für verschiedene Vergleichszwecke genutzt werden und würden somit auch eine wichtige Anforderung im Bereich der **Kohärenz und Vergleichbarkeit** weiter stärken. Der Grundsatz sieht im Zeitverlauf konsistente und zwischen Regionen und Ländern vergleichbare europäische Statistiken vor (Eurostat 2011: 14).

Im Grundsatz **Zugänglichkeit und Klarheit** werden schließlich mehrere Kriterien hinsichtlich der Präsentation und Verbreitung statistischer Informationen festgelegt. Dies betrifft sowohl verfahrenstechnische als auch inhaltliche Aspekte. Für die Presse- und Öffentlichkeitsarbeit bedeutet dies, dass moderne Informations- und Kommunikationstechnologien für alle Arten von statistischen Daten der amtlichen Statistik verwendet werden müssen. Die mit der amtlichen Statistik korrespondierenden Metadaten müssen in standardisierten Metadaten-Systemen dargestellt und archiviert werden, um adäquate Interpretationen und umfassende Vergleiche dauerhaft zu ermöglichen. Ebenfalls zu veröffentlichen sind Informationen über die Qualität der dargestellten Statistiken. Auch bei der Darstellung kleinräumiger statistischer Daten müssen diese Anforderungen erfüllt werden, da für diese Ergebnisse aus der Bundesstatistik die genutzten räumlichen Daten stets miteinbezogen sind.

#### ad (4) Interoperabilität

Interoperabilität ist heutzutage ebenfalls ein wichtiges Thema im Rahmen der Integration statistischer und räumlicher Daten. Dies gilt gleichermaßen für beide Datentypen im Einzelnen als auch im Hinblick auf Daten-Kombinationen. Es geht im Wesentlichen darum, technische Schnittstellen einzurichten, um unterschiedliche Daten zu zeigen, die von einer Vielzahl öffentlicher Stellen auf nationaler oder supranationaler Ebene stammen. In der Regel ist Interoperabilität für räumliche Informationssysteme (GIS) erforderlich. Diese bestehen

grundlegend aus Modulen zum Darstellen und Abrufen der Daten, die auch Datenspezifikationen und Metadaten mit einbeziehen. Ebenen übergreifend existieren für Deutschland (unter Einbezug des Bundes, der Länder und Kreise) aktuell drei öffentlich zugängliche GIS, die Daten des Statistischen Bundesamts nutzen:

- INSPIRE und die nationale Anwendung GDI-DE
- GovData, das Datenportal für Deutschland
- Zentrum für Satellitengestützte Kriseninformation (ZKI)

Im Rahmen der Anwendung der europäischen INSPIRE-Richtlinie wurde von den Statistischen Ämtern des Bundes und der Länder nicht nur gemeinsam geprüft, ob und inwiefern sie rechtlich von der Richtlinie betroffen sind, sondern zugleich auch, inwiefern die Inhalte bestimmter Statistiken einen relevanten inhaltlichen Bezug auf die im Anhang III der INSPIRE-Richtlinie genannten Themenfelder aufweisen. Grundsätzlich sind für kleinräumige Daten für die genannten Themenfelder grundlegende Vorgaben vor allem aus Anhang I (mit Blick auf Adressen und des Art der Projektion) zu berücksichtigen. Die amtliche Statistik in Deutschland hat mindestens sieben thematische Felder identifiziert, zu denen sie beitragen kann. Hierzu wurde die Entscheidung getroffen, dass die Regionaldatenbank Deutschland die grundsätzliche Quelle für INSPIRE-relevante statistische Daten ist (vgl. Schnorr-Bäcker 2016b: 532 f.).

## 5 Schlussfolgerungen

Besonders die Regionalstatistik – auf Basis kartographischer Darstellungen – hat in der jüngeren Vergangenheit an Bedeutung gewonnen, zuletzt oft eng verbunden mit der Diskussion über räumliche Ungleichheiten. Für diese Arbeiten ist eine enge Kooperation mit Informationsanbietern von Geobasisdaten, vor allem den kartographischen Ämtern, essentiell. Es gibt bereits Prozess-Modelle zur Integration statistischer und räumlicher Daten, die ebenso Top-down-Ansätzen wie auch Bottom-up-Ansätzen folgen. Auf der operativen Ebene, die sich stärker darauf konzentriert, generelle Regelungen in praktische Anwendungen bei der Produktion statistisch-räumlicher Informationen in der einen oder anderen Weise zu übersetzen, sind die großen Herausforderungen in Deutschland derzeit: Datenzugriff auf Geobasisdaten, Gewährleistung der Geheimhaltungsregelungen der amtlichen Statistik, Sicherstellung der verschiedenen Qualitätsanforderungen sowie die Berücksichtigung von Interoperabilitätsanforderungen. Die Probleme sind mittlerweile zumindest identifiziert, mögliche Lösungen werden bereits diskutiert und einige konnten davon schon umgesetzt werden.

Neue Entwicklungen finden statt, d. h. vor allem treten neue Datenquellen und neue Ansätze der Datensammlung und -verarbeitung hervor, und beide beeinflussen die Prozesse und Produkte, die von statistischen Ämtern und kartographischen Einrichtungen gemeinsam durchgeführt bzw. bereitgestellt werden. Aus Sicht der amtlichen Statistik ist dies nichts grundlegend Neues, denn Kooperationen mit anderen Datenhaltern, d. h. vor allem anderen öffentlichen Einrichtungen,



haben eine lange Tradition. Allerdings betreten gegenwärtig verschiedene neue Akteure, wie z. B. private Unternehmen, die Wissenschaft und einzelne Akteure den Markt und bieten Informationen oder Methoden an, die auch für die amtliche Statistik nützlich sein könnten. Hieraus folgt, dass neue Wege der Zusammenarbeit gefunden werden müssen. Die größte Herausforderung liegt hierbei darin, sicherzustellen, dass die Prinzipien der amtlichen Statistik, die seit Jahrzehnten gültig und allgemein anerkannt sind, auch zukünftig eingehalten werden. Sie stellen das Rückgrat der Reputation und Akzeptanz der amtlichen Statistik auf nationaler Ebene, europäischer Ebene und weltweit dar. Wenn es gelingt, die verbleibenden Herausforderungen in einer zufriedenstellenden Weise zu steuern, können verschiedene Synergie-Effekte zum Wohle aller hergestellt werden.

- 1 Eine ausführlichere Fassung in Englisch wurde bereits veröffentlicht unter dem Titel: "Combining statistical and geospatial data – challenges and possible solutions from German official statistics' perspective", Conference-Paper for the European Forum for Geography and Statistics, 2016 conference, Paris/ France, November 15–17, 2016.
- 2 <https://www-genesis.destatis.de>
- 3 Die Systematik der Gebietseinheiten für die Statistik (Nomenclature des Unités territoriales statistiques – NUTS) ist ein geographisches System, nach dem das Gebiet der Europäischen Union in hierarchische Ebenen unterteilt ist.
- 4 <https://www.regionalstatistik.de>
- 5 <https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Regionalatlas/RegionalatlasAktuell.html>
- 6 Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG) vom 22.01.1987 (BGBl. I S. 462,565), in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I S. 2394).
- 7 Sie wird auch definiert durch die Richtlinie des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur (GDI) in der Europäischen Gemeinschaft (INSPIRE, 2007/2/EG), siehe INSPIRE, Anhang 1, Nr. 5.
- 8 Zudem ist für weitere durch Bundesgesetz geregelte und von öffentlichen Stellen geführte Register eine Georeferenzierung entsprechend genereller rechtlicher Änderungen im sogenannten E-Government-Gesetz vorgesehen, siehe hierzu das Gesetz zur Förderung der elektronischen Verwaltung sowie zur Änderung weiterer Vorschriften, Paragraph 14 (Bundesgesetzblatt Jahrgang 2013 Teil I Nr. 43, ausgegeben zu Bonn am 31. Juli 2013, S. 2749–2760).
- 9 Gesetz über den Zugang zu digitalen Geodaten (Geodatenzugangsgesetz - GeoZG) vom 10. Februar 2009 (BGBl. I S. 278), das durch Artikel 1 des Gesetzes vom 7. November 2012 geändert worden ist.
- 10 Siehe „Qualitätsoffensive Statistik“, <https://www.destatis.de/DE/Methoden/Qualitaet/Sicherung/Offensive/QualitaetsoffensiveStatistik.html>, zuletzt aufgerufen am 18.10.2017; Statistisches Bundesamt: Die 10 Grundprinzipien der amtlichen Statistik; <https://www.destatis.de/DE/Methoden/Qualitaet/Leitlinien/10Grundprinzipien.html>, zuletzt aufgerufen am 18.10.2017.
- 11 <http://ggim.un.org/>

## Literatur

Eurostat (2011): Verhaltenskodex für Europäische Statistiken für die nationalen und gemeinschaftlichen statistischen Stellen. Angenommen vom Ausschuss für das Europäische Statistische System, 28. September 2011.

Ronning, Gerd et al. (2005): Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. In: Statistisches Bundesamt (Hrsg.): Statistik und Wissenschaft, Band 4, Wiesbaden.

Schnorr-Bäcker, Susanne (1986): ISI "Declaration on Professional Ethics", Internationaler Berufskodex für Statistiker aus der Sicht der Bundesstatistik. In: *Wirtschaft und Statistik*, Ausgabe 8/1986, S. 573–581.

Schnorr-Bäcker, Susanne (2014a): Common issues on benefits and challenges of Big Data sources. Presentation for the UNSD and NBS China International Conference on Big Data for Official Statistics, 28–30 October 2014 in Peking, China. [http://unstats.un.org/unsd/trade/events/2014/beijing/presentations/day3/morning/3.%20Panel%20Discus-](http://unstats.un.org/unsd/trade/events/2014/beijing/presentations/day3/morning/3.%20Panel%20Discussion)

[sion,%20Common%20issues--Susanne%20Schnorr-Baecker.pdf](http://unstats.un.org/unsd/trade/events/2014/beijing/presentations/day3/morning/3.%20Panel%20Discussion). Zuletzt aufgerufen am 18.10.2017.

Schnorr-Bäcker, Susanne (2014b): Regionalstatistische Daten für Deutschland, Europa und weltweit. In: *Zeitschrift für amtliche Statistik Berlin Brandenburg*, Heft 3/2014, S. 12–27.

Schnorr-Bäcker, Susanne (2016a): Statistical Monitoring Systems to Inform Policy Decision-Making, and New Data Sources. Programme of the IARIW conference 2016 in Dresden/Germany, Thursday, August 25. <http://www.iariw.org/c2016dresden.php>. Zuletzt aufgerufen am 18.10.2017.

Schnorr-Bäcker, Susanne (2016b): Regional statistical data for Germany, Europe and the world. In: *Statistical Journal of the IAOS*, 32 (2016), S. 515–535.

Schnorr-Bäcker, Susanne (2017): Statistical monitoring systems to inform policy decision-making, and new data sources. In: *Statistical Journal of the IAOS*, 33 (2017), S. 407–421.

Statistische Ämter des Bundes und der Länder (2005): Gesellschaftlicher Nutzen der amtlichen Statistik, Mehr als 100 gute Gründe für die amtliche Statistik, 2005. <http://www.statistikportal.de/Statistik-Portal/100Grunde.pdf>. Zuletzt aufgerufen am 18.10.2017.

Statistisches Bundesamt (2017): Strategie- und Programmplan, Für die Jahre 2017 bis 2021, Wiesbaden. [https://www.destatis.de/DE/UEberUns/UnsereZiele/Strategieplan2021.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/UEberUns/UnsereZiele/Strategieplan2021.pdf?__blob=publicationFile). Zuletzt aufgerufen am 18.10.2017.

Thiel, Georg; Meinke, Irina (2017): Gut aufgestellt für die Zukunft – ein Dank an Dieter Sarreither. In: *WISTA Wirtschaft und Statistik*, Ausgabe 5/2017, S. 9–18.

UN-GGIM (2016): Integration of geospatial, statistical and other related information. Report on the sixth session of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), Report 2016, 6/107, S. 11–12.