

### 3 Techniken der Regionalisierung

Unter der Überschrift „Techniken der Regionalisierung“ werden zunächst in Abschnitt 3.1 die Typisierung des Raumes zu Regionstypen über unterschiedliche Typen von Merkmalen, in Abschnitt 3.2 die Techniken der Regionalisierung basierend auf Abgrenzungen nach unterschiedlichen Merkmalen und in Abschnitt 3.3 die Technik der Indexbildung beschrieben. Diese drei Abschnitte sind aus den „Regionalen Standards“ der Ausgabe von 2013 unverändert übernommen. Neu ist mit Abschnitt 3.4 ein grundlegender Beitrag über die Regionalisierung durch Georeferenzierung in den Sozialwissenschaften. Potentielle Datenquellen auf unterschiedlichen Ebenen und zu unterschiedlichen Themen einer Georeferenzierung werden in Kapitel 10 aufgelistet und kurz beschrieben.

#### 3.1 Die Typisierung des Raumes zu Regionstypen<sup>1</sup>

*Jürgen H.P. Hoffmeyer-Zlotnik*

Eine Regionalisierung von Umfragedaten strukturiert und untergliedert die Stichproben nach jenen wirtschaftlichen, baulichen, planerischen, demographischen und/oder sozialen Merkmalen des Raumes, die hinsichtlich der Forschungsfrage als Kontextmerkmale einen Sinn ergeben. Das heißt, eine befragte Person wird in einem Merkmalsraum verortet, welcher einen Kontext für Denken und Handeln der befragten Person darstellt.

Vorausgesetzt, die Forschungsfrage beschäftigt sich zum Beispiel mit der Einstellung der Befragten (Grundgesamtheit: Wohnbevölkerung in Privathaushalten im Alter ab 18 Jahren) im Gegensatz zu denen in der Bundesrepublik Deutschland lebenden Personen ohne deutsche Staatsbürgerschaft, dann muss zunächst hinterfragt werden, welche Kontextmerkmale Denken und Handeln der befragten Personen beeinflussen könnten. Um beispielhaft unterschiedliche Typen von Merkmalen anzuführen, werden im Folgenden drei Ebenen betrachtet:

1. eine Unterscheidung nach politischen oder kulturellen Abgrenzungen: zum Beispiel die Unterscheidung nach „West“, definiert über die alten Bundesländer, und „Ost“, bestehend aus den neuen Bundesländern. Die DDR unterschied sich von der alten Bundesrepublik durch eine andere politische Sozialisation. Auch war der Umgang mit „Ausländern“ in den vor 1990 bestehenden beiden deutschen Staaten jeweils ein anderer: Die sichtbarsten Gruppen von Ausländern in der DDR waren Vietnamesen, Kubaner und Mosambikaner, alles räumlich konzentrierte Gruppen, die einem im Alltag nur selten begegneten. Die Gruppen von Auslän-

---

1 Unveränderter Abdruck von Kapitel 3, Abschnitt 1 der Fassung aus den „Regionalen Standards, Ausgabe 2013“

dern in der Bundesrepublik waren die ehemaligen „Gastarbeiter“ und deren Familien, am sichtbarsten die Türken. Anzutreffen waren die Fremden in jeder mittleren bis größeren Stadt der Bundesrepublik, in der Regel in zentrumsnahen Wohnquartieren.

2. eine Unterscheidung nach einer siedlungsstrukturellen Typisierung: zum Beispiel die Berücksichtigung unterschiedlicher Kontaktmöglichkeiten mit „Ausländern“ in „Stadt“ und „Land“, bedingt schon über die gegenüber dem Land höhere Dichte an Personen mit nicht-deutscher Staatsbürgerschaft in der Stadt. Daneben zeichnen sich Städte durch eine höhere Konzentration von Bildungs- und Ausbildungsstätten aus, sodass in der Stadt, gegenüber dem Land, auch allgemein ein höheres Bildungsniveau anzunehmen ist. Geht man davon aus, dass Kontaktmöglichkeiten und Bildungsniveau die Einstellung zu „Ausländern“ beeinflussen, so muss diese Fragestellung einen Stadt-Land-Unterschied aufweisen.
3. eine Unterscheidung nach einem sozialstrukturellen Merkmal: zum Beispiel die Unterscheidung nach der Dichte an Personen mit Migrationshintergrund in der Wohnbevölkerung. Mit steigender Dichte an Personen mit Migrationshintergrund steigen auch die Kontaktmöglichkeiten. Wenn höhere Kontaktmöglichkeiten die Einstellung beeinflussen, dann muss die Dichte von Personen mit Migrationshintergrund in der Wohnbevölkerung einen Einfluss auf die Einstellung haben.

Betrachtet man diese drei Möglichkeiten der Abgrenzung (1) oder der Typisierung (2 und 3) des Raumes „Bundesrepublik Deutschland“, so entstehen über die Regionalisierung nach außen funktional oder strukturell abgegrenzte räumliche Einheiten, die, (bei 2 und 3) in der Regel vielfach vorkommend, einen „Typ“ darstellen. Im Folgenden soll nun gezeigt werden, wie eine Typisierung des Raumes durchgeführt werden kann.

### 3.2 Die Technik der Regionalisierung<sup>2</sup>

*Jürgen H.P. Hoffmeyer-Zlotnik*

Methodisch kann die Strukturierung des Raumes, das heißt die Regionalisierung (im Sinne der Sozialforschung) synthetisch oder analytisch vorgenommen werden. Die synthetische Regionalisierung fasst nach ihrer Ähnlichkeit hinsichtlich eines ausgewählten Merkmalssets (Ähnlichkeitsprinzip) diskrete räumliche Grundeinheiten zu möglichst homogenen Regionen zusammen (Sinz 1995, S. 806).

Technisch geht die synthetische Regionalisierung von einer kleinsten Gebietseinheit, einer „Zelle“, aus und gruppiert benachbarte Zellen nach ihrer Ähnlichkeit hinsichtlich eines ausgewählten Merkmalssets zu „Regionen“. Die verwendeten Merkmale sollten theoriegeleitet ausgewählt werden und können topographische, wirt-

---

2 Unveränderter Abdruck von Kapitel 3, Abschnitt 2 der Fassung aus den „Regionalen Standards, Ausgabe 2013“

schaftliche, soziale, städtebauliche, architektonische und/oder bauliche Strukturen abbilden. Mit der Gruppierung benachbarter Zellen gleichen Typs wird die Zahl der Bezugseinheiten reduziert (siehe auch oben: Beispiele 2 und 3). Diese Art der Typisierung dient dem Zweck, regionale Strukturunterschiede herauszuarbeiten, mit dem Ziel, unterschiedliche Räume über ein Ordnen und Gruppieren miteinander zu vergleichen.

Die analytische Regionalisierung geht von der Region als Untersuchungseinheit aus, unterteilt den gesamten Untersuchungsraum nach einer Reihe für die spezielle Analyse wesentlicher Faktoren und gliedert die Raumelemente nach dem „Verflechtungsprinzip“ unter Berücksichtigung von „Interaktionsbeziehungen zwischen den Systemelementen des betrachteten Raumes“ (Sinz 1995, S. 806).

Technisch basiert die Abgrenzung von Regionen auf drei Elementen,

- den räumlichen Grundeinheiten,
- einem Set von Merkmalen zur Abgrenzung homogener Regionen,

und für die Durchführung einer analytischen Regionalisierung:

- einem Verflechtungsmerkmal.

Die räumlichen Grundeinheiten für eine Typisierung sind nach administrativen, statistischen oder datenerhebungstechnischen Kriterien abgegrenzte Raumeinheiten. Diese Raumeinheiten können Zählbezirke der amtlichen Statistik, Stadtteile, ökologische Bereiche oder „Sichtbereiche“ wie bei „natürlichen Gebieten“ oder Wohnquartieren (siehe Hoffmeyer-Zlotnik 1984), „Gemeinden“ (Böltken 1997; Behrens und Marhenke 1997), „Kreise“ (Böltken und Irmen 1997) oder gar Länder oder Staaten sein. Für die Regionalisierung nationaler Umfragedaten bieten sich am ehesten die administrativen Raumeinheiten „Gemeinde“ und „Kreis“ an, da auf diesen Ebenen für die gesamte Bundesrepublik Deutschland Daten von den Statistischen Ämtern des Bundes und der Länder (siehe Abschnitt 9.1) angeboten werden.

### 3.2.1 Beispiele für eine synthetische Regionalisierung

Analyseregionen aus Politik und Wirtschaft sind oft synthetische Regionen, die auf den räumlichen Grundeinheiten von „Gemeinde“ oder „Kreis“ aufsetzen und benachbarte Einheiten gemäß vorgegebener Kriterien zu Einheiten einer funktionalen, mittleren Größe aggregieren. Am stärksten orientiert sich hierbei die Europäische Statistik (siehe Kapitel 7) und regionale Förderung an den über die Hierarchie der nationalen Verwaltungseinheiten vorgegebenen Aggregatebenen. Aber auch die Marktforschung orientiert sich an Marktregionen mittlerer Größe, die oft aus Aggregaten von Kreisen bestehen. Im Sinne einer synthetischen Regionalisierung sind hier zwei Modelle möglich, die anhand der Nielsen-Gebiete demonstriert werden können:

Zunächst wird ein Gesamttraum in überschaubare regionale Grundeinheiten untergliedert. Hier bieten sich die Bundesländer und, innerhalb der Bundesländer, die Kreise an.

Die A.C. Nielsen Company (2018) fasst, ausgehend von den 16 Ländern der Bundesrepublik Deutschland als regionale Grundeinheiten, diese zu einer überschaubaren Anzahl von 7, nach Möglichkeit vergleichbaren, zusammenhängenden Beobachtungsregionen zusammen.

Innerhalb dieser Regionen werden Teilregionen zu (Nielsen-)Ballungsräumen aggregiert, indem, jetzt ausgehend von den Stadt- und Landkreisen als regionaler Grundeinheit, benachbarte Kreise nach dem Merkmal eines Mindestmaßes an Bevölkerungsgröße (von mindestens 1 Mio. Einwohnern) und hoher Bevölkerungsdichte (von 1.000 bis 1.500 Einwohnern pro qkm) zu 13 Ballungsräumen zusammengefasst werden (siehe Seeborn 2005). Hierzu werden benachbarte Stadt- und Landkreise mit bestimmten Merkmalsausprägungen synthetisch aggregiert.

Bei den drei anfangs genannten Beispielen für eine Regionalisierung im Hinblick auf die Forschungsfrage der „Einstellung der Deutschen zu Migranten“ trifft auf den Tatbestand der synthetischen Regionalisierung einerseits die Unterscheidung nach „Ost“ und „West“ beziehungsweise nach den Bundesländern zu. Andererseits lassen sich Aggregate benachbarter räumlicher Einheiten (Gemeinden, Kreise) ausweisen, in denen die Ausländerdichte einen festgesetzten Schwellenwert überschreitet.

### 3.2.2 Beispiele für eine analytische Regionalisierung

Die analytische Regionalisierung führt zu einer Abgrenzung von Verflechtungsgebieten. Verflechtungsgebiete bestehen in der Regel aus einem Kern und einem mit diesem in Beziehung stehendem Umland. Ideale Beispiele für Verflechtungsgebiete stellen die (Stadt-)Regionen von Christaller (1933) und Burgess (1925) bis Boustedt (1966, 1975) und BIK (Behrens und Marhenke 1997; Behrens und Wiese in Abschnitt 5.2) dar.

Alle Verflechtungsgebiete bestehen aus einem Zentrum oder Kerngebiet und einem mehr oder weniger abgestuften Umland.

- Christaller definiert zentrale Orte und betrachtet deren Einzugsbereiche. Das hierbei verwendete Maß für die Verflechtung ist die „Reichweite“, über die ein „Zentrum“ von den Nutzern aus dem Umland in Anspruch genommen wird.
- Burgess sieht die Stadtregion als ein auf ein Oberzentrum ausgerichtetes Modell konzentrischer Ringe, unterschieden über unterschiedliche Kriterien der Dichte und der Nutzung. Das Maß der Verflechtung ist die Ausrichtung auf das Zentrum.
- Boustedt, und in dessen Revision das BIK, grenzen die Stadtregion über die Kernstadt und deren Umland ab. Die Zonen des Modells werden über Dichtemaße mit unterschiedlicher Ausprägung unterschieden; als Verflechtungsmaß wird die Pendlerbeziehung zwischen Umland und Kern betrachtet.

Das in der deutschen Umfrageforschung heute gebräuchlichste Instrument einer analytischen Regionalisierung ist die Abgrenzung der Stadtregionen des BIK (siehe Abschnitt 5.2). Dieser Index wird über die ADM-Institute den dort erstellten Datensätzen sozialwissenschaftlicher Umfragen in der Regel zugespielt und ist damit über das GVI (Gemeindeverzeichnis) frei erhältlich (siehe Abschnitt 4.3). Das in der

administrativen Regionalforschung häufigste Instrument sind raumordnerische Analyseregionen, zum Beispiel die (siedlungsstrukturellen) Gebietstypen des BBSR (siehe Abschnitt 5.3), die ebenfalls nach Verflechtungsbeziehungen und Dichtemaßen abgegrenzt werden, dabei aber zusätzlich die politische und planerische Relevanz berücksichtigen.

Als Beispiel für eine analytische Regionalisierung sind auch die Raumordnungsregionen des Bundes zu sehen. Raumordnungsregionen, in der Größe zwischen Kreisen und Regierungsbezirken anzusiedeln (siehe Abschnitt 4.3.2), stellen als Beobachtungs- und Planungseinheiten räumliche Aggregate von Kreisen als räumliche Bausteine dar. Diese Planungsregionen, orientiert an bundesweit vergleichbaren Vorgaben der Länder, auf Verflechtungsanalysen basierend, müssen sowohl eine fundierte Datenbasis als auch eine eindeutig zugeordnete Planungshoheit und genügend Spielraum für die Umsetzung von Planung aufweisen. Die 96 Raumordnungsregionen bauen als funktionale Einheiten auf der räumlichen Grundeinheit „Kreis“ auf, berücksichtigen in der Regel die Abgrenzungen von Regierungsbezirken und respektieren generell die Grenzen der Bundesländer, selbst wenn dies in einigen Fällen der zugleich angestrebten funktionalräumlichen Abgrenzung widerspricht. Damit bilden Raumordnungsregionen ein räumliches Raster bundesweit vergleichbarer Analyseregionen für Zwecke der Raumbeobachtung und Politikberatung. Ziel der Raumordnung ist das Bemühen des Staates, eine den „sozialen, wirtschaftlichen und kulturellen Erfordernissen entsprechende Ordnung“ (Ley 1966, S. 1509) des Raumes herbeizuführen (siehe Bundesministerium für Raumordnung, Bauwesen und Städtebau 1994, S. 5ff.), damit zum Beispiel die Gleichwertigkeit der Lebensbedingungen in unterschiedlichen Regionen hergestellt werden kann.

### 3.3 Technik der Indexbildung<sup>3</sup>

*Jürgen H.P. Hoffmeyer-Zlotnik*

Für die Typisierung auf den unterschiedlichen Ebenen der räumlichen Betrachtung gibt es eine Reihe von Instrumenten, auf die der Forscher zurückgreifen kann. Auf der Ebene nationaler Umfragen bieten sich siedlungsstrukturelle Typisierungen (siehe Abschnitt 5.3) und die Typisierung von Stadtregionen (die Instrumente des BIK, siehe Abschnitt 5.2) an. Auf der Ebene der Stadt existiert ein Instrumentarium zur Klassifikation von Städten (Friedrichs 1995; Hoffmeyer-Zlotnik 2000a). Eine Untergliederung von Städten, aufbauend auf administrativen Abgrenzungen, ist nicht flächendeckend für die Bundesrepublik Deutschland, sondern nur in Kooperation mit einzelnen Kommunen (als Fallstudie) möglich. Für eine tiefe Untergliederung der Stadt in soziale Räume (unter anderen Friedrichs 1977) oder Wohnquartiere (siehe Hoffmeyer-Zlotnik 2000b) liegen Instrumente vor. Eine Sozialraumanalyse

3 Unveränderter Abdruck von Kapitel 3, Abschnitt 3 der Fassung aus den „Regionalen Standards, Ausgabe 2013“

(siehe Abschnitt 6.1.1) nutzt im Idealfall kleinräumig aufbereitete Zensusdaten, die zur Zeit nicht zur Verfügung stehen. Alternativ kann man aber entweder die bestehenden kommunalstatistischen Daten der kommunalen Datenbanken wie KOSTAT, IRB (siehe Abschnitt 6.2) oder Urban Audit (siehe Abschnitt 7.3) nutzen oder man erhebt eigene Daten für eine Wohnquartiersbeschreibung (siehe Abschnitt 6.3).

Zusätzlich zur Nutzung existenter Instrumente räumlich struktureller Typisierungen, wie zum Beispiel von BIK oder BBSR, kann der Forscher aus ihm zugänglichen Daten im Sinne der Forschungsfrage eigene Maßzahlen und Indikatoren bilden. Das hierzu am häufigsten benutzte Verfahren ist die Generierung einer Maßzahl zur Kennzeichnung von Strukturen, Veränderungen oder Verflechtungen. Dieses geschieht mittels „Quotenbildung“, das heißt, es wird eine Teilmasse durch die Gesamtmasse geteilt und auf Prozentanteile normiert. Solche Maßzahlen zur Kennzeichnung von Strukturen dienen dazu, das innere Gefüge von Beobachtungs- oder Analyseeinheiten darzustellen. Es können folgende Typen von Kennziffern, Maßzahlen oder Indizes unterschieden werden:

- **Leistungskennzahlen:** Diese stellen die (vor allem wirtschaftliche) Leistungskraft einer Region dar. Leistungskennzahlen beschreiben Beziehungen.
- **Versorgungsziffern:** Diese zeigen den Grad der Versorgung von Personengruppen mit Gütern oder Dienstleistungen.
- **Maßzahlen zum Versorgungsgrad:** Diese geben Informationen über Strukturen von Bedarf und Verbrauch.
- **Maßzahlen zur Darstellung von Veränderungen:** Diese beschreiben eine Entwicklung zwischen zwei Zeitpunkten.
- **Maßzahlen zur Darstellung von Raumverflechtungen:** Diese zeigen Art, Intensität und Richtung von Verflechtungs- und Austauschbeziehungen unterschiedlicher (wie auch immer definierter) Regionen zueinander, wie zum Beispiel jegliche Art räumlicher Mobilität.

### **3.4 Regionalisierung durch Georeferenzierung in der Sozialforschung**

*Pascal Siegers, Stefan Müller & Julia Klinger<sup>4</sup>*

#### **3.4.1 Einleitung: Georeferenzierung in der Sozialforschung**

Die öffentliche Verfügbarkeit von Geodaten hat in den vergangenen Jahren erheblich zugenommen. Die Gründe dafür sind zum einen, dass durch technologische Fortschritte immer mehr und immer detaillierte Daten mit Raumbezug für die administrative und wirtschaftliche Verwertung generiert werden. Zum anderen hat der Gesetzgeber Maßnahmen getroffen, damit Daten, die von öffentlichen Verwaltungen

---

4 GESIS – Leibniz-Institut für Sozialwissenschaften, Datenarchiv für Sozialwissenschaften

erhoben werden, für Forschung und Wirtschaft genutzt werden können. Dazu wurden Datenportale geschaffen, die den Zugang zu Geodaten der öffentlichen Hand erleichtern sollen. Als Folge dieser Maßnahmen hat die Relevanz von Geodaten in Forschung, Verwaltung und Wirtschaft in den letzten 20 Jahren erheblich an Bedeutung gewonnen.

Auch in der Sozialforschung werden Geodaten zunehmend für die Anreicherung von Befragungsdaten verwendet. So können zum Beispiel den Daten führender Umfrageprogramme wie des Sozio-Oekonomischen Panels (SOEP), des Nationalen Bildungspanels (NEPS) oder der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) mit Hilfe von Georeferenzen auf kleinräumiger Ebene (Geo-) Informationen hinzugefügt werden. Anwendungen lassen sich in verschiedensten Teildisziplinen der Sozialwissenschaften finden, wie etwa der politischen Verhaltens- und Einstellungsforschung (Förster 2018; Klinger, Müller und Schaeffer 2017) oder der Bildungsforschung (Weßling 2016).

Die Vorteile bei der Verwendung von Geodaten liegen dabei auf der Hand. Erstens ermöglichen sie die kleinräumige Modellierung des Lebensumfeldes der Befragten. Zweitens kann das räumliche Lebensumfeld der Befragungsteilnehmenden unabhängig von geographisch eher grob skalierten administrativen Einheiten wie zum Beispiel Gemeinden betrachtet werden, soweit dafür ausreichend Daten zur Verfügung stehen. Und drittens können Zeitreihen über räumliche Referenzen einfacher harmonisiert und damit verglichen werden.

Mit der progressiven Umsetzung der Europäischen Richtlinie für Geodaten „Infrastructure for Spatial Information in Europe“ (INSPIRE) wird indessen das Angebot frei verfügbarer Geodaten weiter steigen und für die Forschungspraxis noch weiter an Bedeutung gewinnen.

Auch die Berechnung von Distanzen und Erreichbarkeiten ist für Fragestellungen aus der Sozialforschung sehr relevant, zum Beispiel wenn es um die Verfügbarkeit von Kinderbetreuungseinrichtungen, öffentlichem Personennahverkehr oder der Gesundheitsversorgung geht.

Der Verwendung von Geoinformationen in der Befragungsforschung sind durch Anforderungen des Datenschutzes hingegen auch Grenzen gesetzt. Zur Wahrung der Anonymität der Daten muss bei der Bearbeitung von Geoinformationen in Befragungsdaten besondere Sorgfalt Anwendung finden, weil Geodaten Zusatzinformationen beinhalten, die eine De-Anonymisierung der Befragten potentiell erleichtern (Blatt 2012). Nach der neuen Datenschutzgrundverordnung müssen die Teilnehmenden einer Befragung über die Verwendung ihrer personenbezogenen Daten umfassend aufgeklärt werden. Werden die Daten georeferenziert und mit weiteren Geodaten räumlich verknüpft, müssen diese Verfahren Bestandteil der informierten Einwilligung der Befragten sein.

Dieser Beitrag gibt eine Übersicht über die Grundlagen der Georeferenzierung und der räumlichen Verknüpfung, insbesondere mit Befragungsdaten.

Dazu startet der Beitrag mit einer kurzen Einführung in (1) die Grundbegriffe der Georeferenzierung und stellt daran anschließend (2) die wichtigsten Geobasisdaten für Deutschland und deren Bezugsquellen dar. Im Anschluss daran werden (3)

verschiedene Verfahren der räumlichen Verknüpfung illustriert und (4) Fragen des Datenschutzes bei der Georeferenzierung von Befragungsdaten angeschnitten.

### 3.4.2 Grundlagen der Georeferenzierung

Der Begriff Georeferenzierung bezeichnet sehr allgemein die Zuweisung von raumbezogenen Informationen zu Daten. So gesehen sind die klassischen Regionalisierungstechniken auch Verfahren der Georeferenzierung. Im engeren Sinn wird der Begriff Georeferenzierung hingegen für die Zuweisung direkter Raumbezüge von Daten verwendet (RatSWD 2012, S. 11). Dies erfasst auch klassische Raumbezüge, wie zum Beispiel die Verwaltungsebenen der Bundesrepublik (Hillmert et al. 2017, 270f), geht zum Teil allerdings deutlich darüber hinaus, weil sehr fein-granulare Raumbezüge abgebildet werden können. Zudem können anhand der räumlichen Verknüpfung Merkmale über ihren Raumbezug in Beziehung gesetzt werden. So können zum Beispiel Daten aus der amtlichen Lärmkartierung den Wohnorten von Teilnehmern an einer Befragung zugeordnet werden (Müller et al. 2017). Dieser Prozess der Zuweisung von Raumbezügen und Verarbeitung raumbezogener Daten soll in seinen Grundzügen in diesem Abschnitt vorgestellt werden.

Wie bereits erwähnt, bezeichnet *Georeferenzierung* die Zuweisung von Raumbezügen zu Daten. In der Sozialforschung handelt es sich dabei in einem ersten Schritt meistens um *indirekte Georeferenzen*, weil der Raumbezug in den Daten ohne Bezug auf ein bestimmtes Koordinatenreferenzsystem vorliegt. Solche Georeferenzen liegen häufig bereits aus dem Stichprobenverfahren vor, wenn Adressen der Befragungspersonen aus den Melderegistern gezogen werden, oder sie werden eigens erhoben, zum Beispiel, wenn in Telefon- und Onlineerhebungen die Postleitzahlen ermittelt werden. Indirekte Georeferenzen sind folglich Namen von Kreisen, Gemeinden oder Ortsteilen genauso wie Adressen oder Postleitzahlen.

Die Verarbeitung von indirekten Georeferenzen in Geoinformationssystemen (GIS), die zur Darstellung und Verarbeitung von Geodaten genutzt werden, ist jedoch nicht möglich. Dafür müssen die indirekten Georeferenzen in *direkte Georeferenzen*, das heißt *Geokoordinaten*, umgewandelt werden. Geokoordinaten drücken den Raumbezug innerhalb eines spezifischen *Koordinatenreferenzsystems* aus, enthalten also die Positionsdaten und ermöglichen die Verortung der Daten im Raum, die dann mit Hilfe von Geoinformationssystemen dargestellt werden können. Koordinatenreferenzsysteme definieren den Ursprung eines Koordinatensystems und bestimmen den Bezug zur Erdkrümmung. Jede Koordinate besteht aus einem x- und y-Wert, der – in Metern oder Grad – die Distanz zum Ursprung des Koordinatensystems enthält. Koordinaten ermöglichen so eine eindeutige Verortung der Punkte (Koordinaten) auf der Erdoberfläche. Sie werden in Geoinformationssystemen für die Visualisierung und Bearbeitung von Geodaten verwendet. Es existieren viele verschiedene Koordinatenreferenzsysteme. In Europa gängige Koordinatenreferenzsysteme sind UTM und ETRS89. Zusammengefasst kann gesagt werden, dass im Allgemeinen Geokoordinaten gemeint sind, wenn von georeferenzierten Daten gesprochen wird.

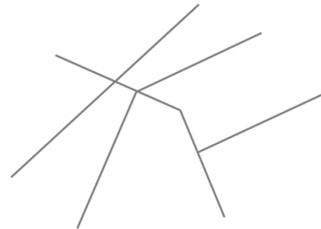
Da in der Sozialforschung in der Regel keine direkten Georeferenzen vorhanden sind, müssen die indirekten Georeferenzen in Geokoordinaten umgewandelt werden. Dieses Verfahren wird als *Geokodierung* bezeichnet (Zandbergen 2014, S. 2). Einfache Beispiele für die Geokodierung sind die Zuweisung einer Punktkoordinate zu einer Adresse oder die Zuweisung eines Polygons zu einer Gemeinde (entsprechend der Gemeindegrenze). Die Attribute der Adresse (zum Beispiel ein bestimmter Haushalt) oder der Gemeinde können dann über den Raumbezug kartographisch visualisiert oder mit anderen Attributen des Raums verknüpft werden. Der einfachste Fall dafür liegt vor, wenn zwei Attributen Punktkoordinaten zugewiesen werden. Dann können mit Hilfe von Geoinformationssystemen etwa die *Distanzen* (in der Luftlinie) berechnet werden (Meyer und Bruderer Enzler 2013, S. 327). So kann zum Beispiel die Entfernung vom Wohnort eines Befragten zu den Kindergärten im Wohnumfeld oder zur Arbeitsstätte der Studienteilnehmenden bestimmt werden.

Auf Geokoordinaten basierende Daten liegen entweder als Vektor- oder Rasterdaten vor, die beispielhaft in Abbildung 3.4.2-01 dargestellt sind. Bei Vektordaten werden Punkt- (z.B. die Position einer Adresse), Linien- (z.B. die Straße, in der eine Adresse liegt) und *Polygondaten* (z.B. die Grenzen der Gemeinde, in der eine Adresse liegt) unterschieden. *Rasterdaten* unterteilen ein bestimmtes Gebiet in gleichgroße Flächenabschnitte (in der Regel Quadrate).

Punkte, z.B. Adressen von Befragten



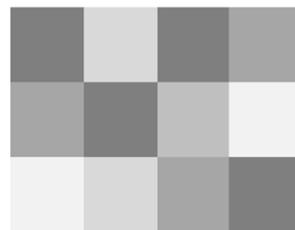
Linien, z.B. Straßen



Polygone, z.B. Gemeindegrenze



Raster, z.B. gleichgroße Flächenabschnitte



Quelle: Müller 2019.

Abbildung 3.4.2-01 Illustration verschiedener Formen von Georeferenzen

### 3.4.3 Geobasisdaten für die Bundesrepublik Deutschland

Die Visualisierung und Bearbeitung von Geodaten in der Sozialforschung knüpft in der Regel an die Verwaltungsebenen und andere Gebietseinheiten an. Zwei Gründe liegen dafür auf der Hand. Zum einen werden die Stichproben in der Umfrageforschung häufig stratifiziert nach Bundesländern und Gemeinden gezogen, sodass die Befragungsinformationen geklumpt für bestimmte Gebietseinheiten vorliegen. Zum anderen, weil georeferenzierte Daten (z.B. zu Bevölkerung, Wirtschaft etc.) für die Gemeinden oder Kreise vorliegen, werden diese auch für die Auswertung der Befragungsdaten verwendet.

Die Informationen über Verwaltungs- und Gebietseinheiten sind schon lange als Hintergrundinformation für die Auswertung von Befragungsdaten erschlossen. Neu ist, dass die Regionalinformationen als direkte Georeferenzen mit den Befragungsdaten verbunden werden können. Dazu wird jedoch für alle Einheiten die Information benötigt, welche Flächen in die Grenzen der jeweiligen Einheiten fallen, zum Beispiel zu einer Gemeinde oder einem Landkreis gehören. Solche Daten werden als *Geobasisdaten* bezeichnet. Geobasisdaten liegen in Deutschland für alle relevanten Regional- und administrativen sowie nicht-administrativen Gebietseinheiten vor. Die höchste subnationale Ebene der Regionaleinheiten bilden die Bundesländer gefolgt von Regierungsbezirken (die jedoch nicht in allen Bundesländern existieren), Landkreisen und kreisfreien Städten und Gemeinden. Als Gebietseinheiten liegen u.a. Raumordnungsregionen und Wahlkreise vor<sup>5</sup>. Für alle diese „klassischen“ Regional- und Gebietseinheiten liegen Geobasisdaten vor, die eine Projektion in Koordinatenreferenzsysteme ermöglichen<sup>6</sup>.

Abbildung 3.4.3-01 zeigt die Darstellung der Geobasisdaten für Bundesländer, Kreise und Gemeinden in der Bundesrepublik, die über das Bundesamt für Kartographie und Geodäsie erhältlich sind.

Geodaten sind jedoch besonders attraktiv, wenn noch kleinräumigere Informationen verarbeitet werden können. Eine Möglichkeit dazu war bislang die Verwendung von Postleitzahlbereichen. Da die Postleitzahlen von der Deutschen Post AG jedoch zu kommerziellen Zwecken erstellt werden, sind offizielle Geobasisdaten für die Postleitzahlbereiche nicht kostenfrei zu erhalten. Gleiches gilt auch für die sogenannten PLZ8-Gebiete von der microm AG. Diese achtstelligen Codes unterteilen die Bundesrepublik in Flächensegmente, in denen jeweils etwa 500 Haushalte leben<sup>7</sup>. Die einzelnen PLZ8-Gebiete werden dabei zu Zwecken des Geomarketings anhand verschiedener Variablen typologisiert.

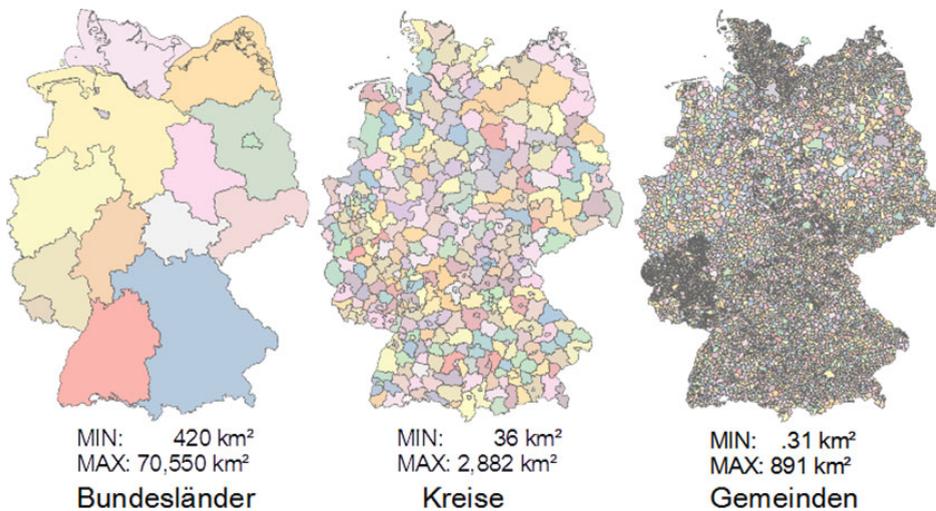
Unabhängig von den Grenzen der Verwaltungsebenen und Gebietskörperschaften wurde für Europa eine Rastereinteilung definiert, die sogenannten INSPIRE-Grids.

5 Siehe auch Gemeinde-Informationssystem GV-ISys der Statistischen Ämter des Bundes und der Länder.

6 Eine sehr synthetische Übersicht über die verschiedenen Verwaltungsebenen inkl. der Regierungsbezirke und Raumordnungsregionen findet sich bei Klinger (2018).

7 Für eine kurze Beschreibung siehe <https://www.microm.de/marktdaten/lage-und-standortdaten/plz8/>, Link geprüft am 14. Juni 2018.

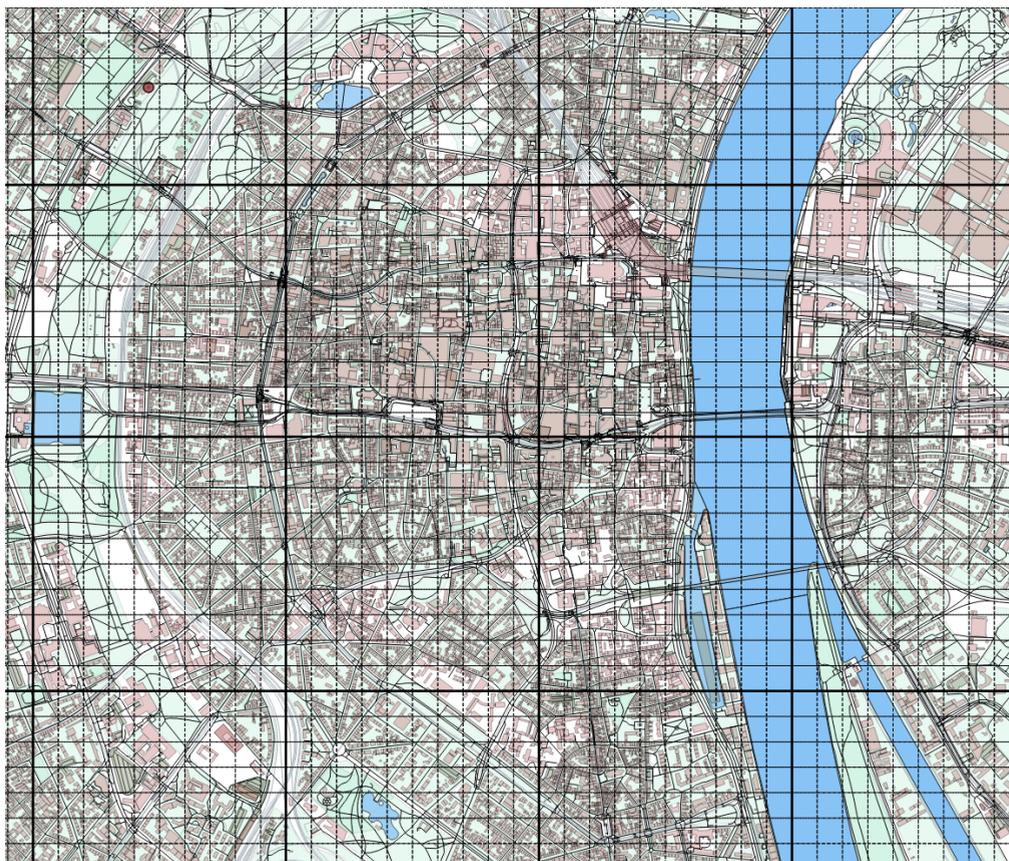
Die Grids unterteilen das Staatsgebiet Deutschlands in gleich große quadratische Rasterzellen. Jede Rasterzelle erhält eine eindeutige Identifikationsnummer (Grid ID), die systematisch aus der Koordinate der linken unteren Ecke der Rasterzelle und der Kantenlänge gebildet wird. INSPIRE-Grids liegen für unterschiedliche Kantenlängen der Rasterzellen vor: 100 Meter (Zellenfläche 10.000 m<sup>2</sup>), 250 Meter (Zellenfläche 62.500 m<sup>2</sup>), 500 Meter (Zellenfläche 250.000 m<sup>2</sup>), 1.000 Meter (Zellenfläche 1 km<sup>2</sup>), 10.000 Meter (Zellenfläche 100 km<sup>2</sup>) und 100.000 Meter (Zellenfläche 10.000 km<sup>2</sup>). Für die kleinräumige Beschreibung sind vor allem die Grids mit den Kantenlängen von 1 km und kleiner sinnvoll anwendbar. Zu beachten ist allerdings bei der Verwendung der kleineren Raster (100 Meter und weniger), dass diese Daten sehr viele Datenpunkte enthalten und entsprechend große Speicher- und Rechenkapazitäten für die Visualisierung und Bearbeitung erforderlich sind.



Eigene Darstellung mit Shapefiles vom Bundesamt für Kartographie und Geodäsie ([http://www.geodatenzentrum.de/download/archiv/vektor/vg250\\_ebenen/2015/](http://www.geodatenzentrum.de/download/archiv/vektor/vg250_ebenen/2015/))

Abbildung 3.4.3-01 Darstellung der Geobasisdaten für Bundesländer, Kreise und Gemeinden in Deutschland

Abbildung 3.4.3-02 zeigt die INSPIRE-Grids mit 1 km und 100 m Kantenlänge am Beispiel der Kölner Innenstadt. Die größeren Rasterzellen, die von durchgezogenen Linien eingegrenzt werden, sind die 1 km<sup>2</sup> Zellen. Die von den gestrichelten Linien eingegrenzten kleineren Zellen haben eine Fläche von jeweils 1 ha (der Kölner Dom, gut sichtbar südlich des Hauptbahnhofes, belegt das ca. 1,5-fache der 1 ha Rasterzellen).



Eigene Darstellung mit Shapefiles von Geofabrik GmbH and OpenStreetMap Contributors (Straßenkarte) sowie Bundesamt für Kartographie und Geodäsie (Rasterzellen). Maßstab 1:13.000.

*Abbildung 3.4.3-02* Visualisierung der INSPIRE-Grids mit 1 Km und 100 Meter Kantenlänge am Beispiel der Kölner Innenstadt

Tabelle 3.4.3-01 gibt eine Übersicht über Geobasisdaten, die für die Verwendung in der Sozialforschung relevant sein können und flächendeckend für ganz Deutschland vorliegen.

Tabelle 3.4.3-01 Übersicht über Geobasisdaten

Gebietseinheiten	Beschreibung	Quelle für Basisdaten
Regierungsbezirke <sup>b</sup>	19 Bezirke (NUTS Ebene 2) <sup>a</sup>	Bundesamt für Kartographie und Geodäsie oder GV-ISys
Raumordnungsregionen	96 Raumordnungsregionen (Aktualisierung jährlich)	Bundesinstitut für Bau-, Stadt- und Raumforschung, GV-ISys
Kreise und kreisfreie Städte	412 Kreise und kreisfreie Städte Aktualisierung jährlich (NUTS Ebene 3) <sup>a</sup>	Bundesamt für Kartographie und Geodäsie oder GV-ISys
Wahlkreise	299 Wahlkreise (Aktualisierung 4-jährig)	Statistisches Bundesamt: Bundeswahlleiter, GV-ISys
Politische Gemeinde	>11.000 Gemeinden in Deutschland (Aktualisierung jährlich)	Bundesamt für Kartographie und Geodäsie oder in GV-ISys
Postleitzahl 5	Aus GV-ISys sind die Zustellpostleitzahlen der Deutschen Post für die Verwaltungsadresse jeder Regionaleinheit abrufbar (s. Link Onlineabfrage GV-ISys).	GV-ISys, Deutsche Post (Basisdaten nicht kostenfrei verfügbar)
Postleitzahl 8 (PLZ8)	Angebot von microm; PLZ8-Bereiche umfassen jeweils ca. 500 Haushalte und sind damit sehr kleinteilig. Die Gebiete werden mit jeder Aktualisierung neu zugeschnitten und sind deshalb nicht durch die Zeit vergleichbar.	microm AG (Basisdaten nicht kostenfrei verfügbar)
INSPIRE-Grids 1 km <sup>2</sup>	Raster aus Zellen mit 1.000 Meter Kantenlänge	Bundesamt für Kartographie und Geodäsie
INSPIRE-Grids 250 ha	Raster aus Zellen mit 500 Meter Kantenlänge	Bundesamt für Kartographie und Geodäsie
INSPIRE-Grids 6,25 ha	Raster aus Zellen mit 250 Meter Kantenlänge	Bundesamt für Kartographie und Geodäsie
INSPIRE-Grids 1 ha	Raster aus Zellen mit 100 Meter Kantenlänge	Bundesamt für Kartographie und Geodäsie

*Anmerkung:* <sup>a</sup> Nomenclature des unités territoriales statistiques; <sup>b</sup> Seit 2005 gibt es Regierungsbezirke nur noch in den Bundesländern Bayern, Baden-Württemberg, Hessen und Nordrhein-Westfalen



gungen des sozialen Handelns in die Auswertung einbeziehen, wenn Adresdaten für die Probanden vorliegen (Klinger et al. 2017; Förster 2018).

Ein Vorteil der Verwendung von direkten Georeferenzen bei der Erstellung von Regionalinformationen für Befragungsdaten ist die große Effizienz des Verfahrens. Liegen Punktkoordinaten für die Befragten vor, kann mittels einer einfachen Routine eine räumliche Verknüpfung mit den Geobasisdaten und den darin enthaltenen Attributen erstellt werden. So kann ein Datensatz erzeugt werden, der alle in Tabelle 3.4.3-01 erwähnten Verwaltungs-, Gebiets-, und Flächeneinheiten enthält. Für Verwaltungseinheiten weisen die Daten auch Namen und Kennnummern auf (wie den Amtlichen Gemeindegemeinschaften).

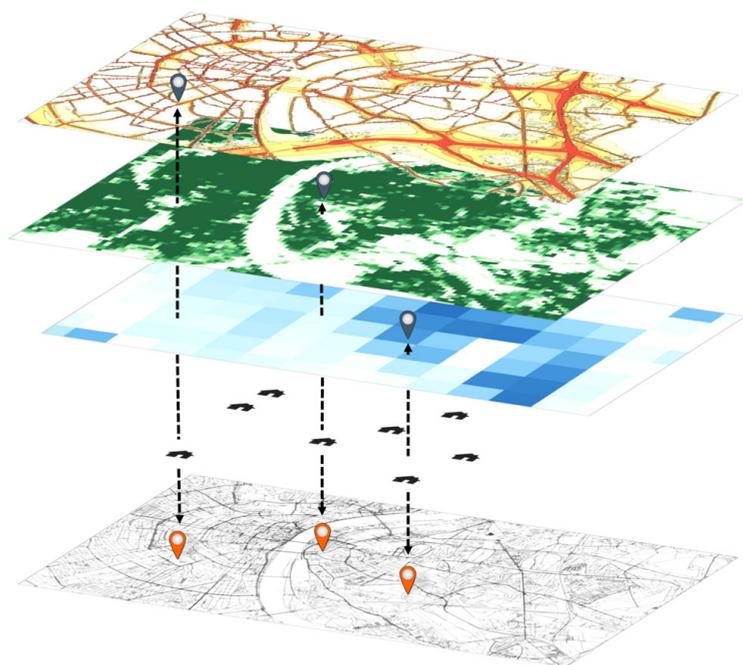
Die größte Herausforderung liegt dann darin, die indirekten Georeferenzen – bei hochwertigen Befragungen in der Regel Adressen – zu geokodieren. Das BKG bietet dafür einen Geokodierungsdienst, den BKG GeoCoder, an. Der Dienst funktioniert über eine Webschnittstelle (API) und gibt neben den Ergebnissen der Geokodierung auch Schätzer für die Qualität der Geokodierung aus, sodass für die Nutzenden unmittelbar ersichtlich ist, ob die Kodierung einer Adresse erfolgreich war. Der Vorteil des BKG GeoCoders gegenüber kostenfreien, aber meist kommerziell orientierten Angeboten im Netz ist, dass die zu konvertierenden Daten nicht gespeichert werden, sondern nur „on-the-fly“ verarbeitet werden. Aus datenschutzrechtlicher Perspektive ist dies eine notwendige Bedingung für die Verarbeitung von personenbezogenen Daten. Die technische Umsetzung der Geokodierung wurde zudem vom Bundesdatenschutzbeauftragten begutachtet und nicht beanstandet. Ein weiterer Dienst des BKG ermöglicht Erreichbarkeitsschätzungen über einen sogenannten Routingdienst. Damit kann – anders als bei Distanzen, die in der Regel die Entfernung in der Luftlinie messen – die Zeit ermittelt werden, die für die Überwindung einer Strecke gebraucht wird. Ein Nachteil des BKG GeoCoders und anderer BKG Dienste ist, dass diese nur für Behörden des Bundes frei zugänglich sind. GESIS bietet wissenschaftlichen Forschungsprojekten die Konvertierung von Adressen in Koordinaten über den BKG GeoCoder jedoch auf Anfrage kostenfrei an.

#### **3.4.4 Räumliche Verknüpfung und Modellierung von räumlichen Zusammenhängen**

Über direkte Georeferenzen (insbesondere wenn Punktkoordinaten vorliegen) ist eine räumliche Verknüpfung von Daten mit Raumbezug möglich. Dabei werden die Attribute einer räumlichen Ebene (d.h. eines Layers) auf eine andere Ebene übertragen und können dann gemeinsam ausgewertet, zum Beispiel korreliert, werden. Anstelle der Verwendung von Korrespondenztabelle für räumliche Einheiten (z.B. den Amtlichen Gemeindegemeinschaften) kann die Verknüpfung über einfache Routinen sichergestellt werden. Dafür stehen in Geoinformationssystemen entsprechende Funktionen zur Verfügung. Neben kostenpflichtigen und proprietären Programmen, wie zum Beispiel ArcGIS (ESRI 2015), sind kostenfreie Alternativen für die Bearbeitung von Geodaten verfügbar. Besonders bekannt und einfach in der Handhabung ist die freie Software QGIS (QGIS Development Team 2018). Für die räumliche Ver-

knüpfung und Auswertung von raumbezogenen Daten sind auch verschiedene Pakete der freien Statistiksoftware R verfügbar, zum Beispiel das Paket `sf` (simple features., <https://cran.r-project.org/web/packages/sf/index.html>). Dabei handelt es sich um eine Implementierung des ISO 19125 Geodatenstandards (<https://www.iso.org/standard/40114.html>), der auch vom Open Geospatial Consortium (OGC) geteilt wird und somit breite Unterstützung in der Community erfährt.

Der einfachste Fall einer räumlichen Verknüpfung ist die Zusammenführung mehrerer Attribute eines Punktes im Raum. So können zum Beispiel dem Wohnort eines Befragten Informationen zu Umwelt- oder Lärmbelastungen zugewiesen werden, um Ungleichheiten in der Belastung nach sozio-ökonomischem Status oder Migrationsstatus zu untersuchen (Diekmann und Mayer 2010). Ein Beispiel für diesen Fall der räumlichen Verknüpfung zeigt Abbildung 3.4.4-01.

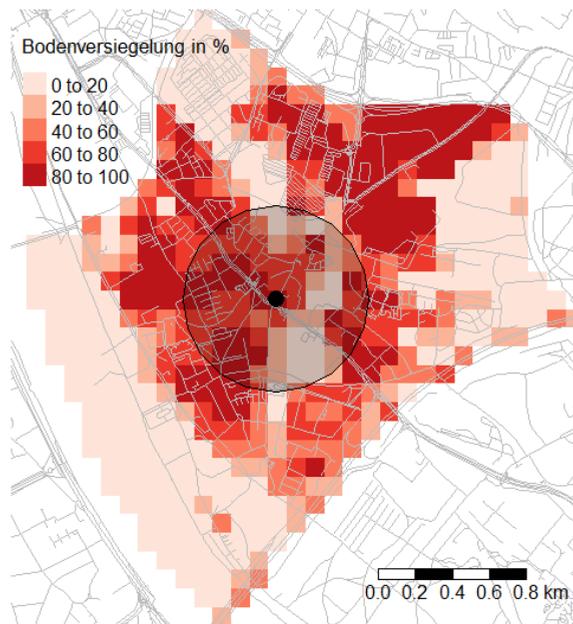


Eigene Darstellung. Quellen: Umweltbundesamt (Straßenlärm), Leibniz-Institut für Ökologische Raumentwicklung (Bodenversiegelung 1 ha INSPIRE-Grids) und Zensus 2011 (Ausländeranteil 1 Km<sup>2</sup> INSPIRE-Grids).

Abbildung 3.4.4-01 Darstellung einer räumlichen Verknüpfung von Positionsdaten, Geobasisdaten und Geofachdaten am Beispiel der Kölner Innenstadt

Die schwarzen Haussymbole stehen in der Abbildung für die Wohnorte der Befragten. Diese werden für die räumliche Verknüpfung in Koordinaten umgerechnet (orange-weiße Pfeile) und können in Geobasisdaten visualisiert werden. Die Geofachdaten werden als eigene Schichten den Punkten zugefügt. In unserem Beispiel

werden drei Attribute aus Geofachdaten verknüpft. Erstens Daten aus der amtlichen Lärmkartierung (oberste Schicht), die Auskunft darüber geben, wie stark die Befragten an ihrer Wohnadresse Lärmbelastungen ausgesetzt sind, zweitens das Ausmaß der Bodenversiegelung aggregiert für die 1 ha INSPIRE-Grids (zweite Schicht von oben) und drittens der Ausländeranteil aggregiert auf die 1 km<sup>2</sup> INSPIRE-Grids (dritte Schicht von oben). Der Datensatz, der aus der räumlichen Verknüpfung resultiert enthält alle Attribute des Raumes und kann mit den Attributen aus der Umfrage verknüpft werden, zum Beispiel um der Frage nachzugehen, ob ein höheres Einkommen tatsächlich mit geringeren Umweltbelastungen im Wohnumfeld einhergeht.



Eigene Darstellung. Quellen: Leibniz-Institut für Ökologische Raumentwicklung (Bodenversiegelung), OpenStreetMap/Stadt Köln (Geobasisdaten)

*Abbildung 3.4.4-02* Darstellung einer Puffer- bzw. Zonenbildung zur Aggregation von Geofachdaten (hier: Bodenversiegelung) am Beispiel des Kölner Stadtteils Poll

Über die Verknüpfung mittels Punktdaten hinaus sind vor allem drei allgemeine GIS-Funktionen zur räumlichen Verknüpfung relevant:

a. **Puffer** (auch Zonen genannt) sind Flächen, die dadurch gebildet werden, dass ihr Rand in einem bestimmten vorgegebenen Abstand um gegebene Punkte oder Linien (durch die Pufferfunktion) gebildet wird. Damit können Analysen durchgeführt werden, welche anderen Objekte in einem bestimmten Abstand um gegebene

Objekte vorhanden sind, beispielsweise mit welchem Lärmpegel Haushalte in einem bestimmten Abstand von einer Straße belastet werden.

b. Mit **Verschneidungen** kann festgestellt werden, ob und wie sich raumbezogene Objekte in Punkt-, in Linien- oder in Flächenform (die z.B. Umfragen repräsentieren) räumlich überlappen. Dies erlaubt auch, dass die Umfragedaten zueinander in Beziehung gebracht bzw. auf die jeweils andere geometrische Struktur aggregiert werden können.

c. Mit Hilfe von **Clusterfunktionen** kann – auch multivariat – analysiert werden, ob Umfragedaten bzgl. ihrer geometrischen Lage und anderer Parameter (aus Umfrageergebnissen abgeleitet) Cluster bilden, die sich signifikant von anderen abheben. Die dazu verwendeten Clusteralgorithmen wie z.B. k-means oder single- bzw. complete-linkage entsprechen denen aus der Umfragestatistik.

### 3.4.5 Verfügbarkeit von Geofachdaten für die Forschung

Neben den Geobasisdaten sind vor allem Geofachdaten für die Forschung relevant. Geofachdaten enthalten Attributdaten mit Raumbezug (zum Beispiel Informationen über die Flächennutzung). Beispiele für Geofachdaten wurden bereits erwähnt: aggregierte Merkmale aus dem Zensus 2011 auf INSPIRE-Grid-Ebene, Daten der amtlichen Lärmkartierung, Daten von Anbietern für Geomarketing wie der microm AG, der GfK oder infas 360°, Daten des FDZ Ruhr am RWI usw.

Gerade die öffentlichen Verwaltungen wurden durch den Gesetzgeber angehalten, Daten aus Verwaltungsprozessen öffentlich zur Verfügung zu stellen. Dazu wurden eigene Kataloge erstellt, insbesondere das Suchportal der Geodateninfrastruktur für Deutschland (GDI-DE, [www.geoportal.de](http://www.geoportal.de)). Eine Übersicht über Bezugsquellen für Geodaten aus öffentlichen und kommerziellen Quellen findet sich in Kapitel 9 der „Regionalen Standards“ oder im Anhang zu Müller et al. (2017).

Das Gros der Geodaten wird von Ländern oder Gemeinden erhoben. Die Zuständigkeit ergibt sich aus der föderalen Regierungs- und Verwaltungsstruktur der Bundesrepublik, in deren Folge die Datenerhebung stark fragmentiert erfolgt, auch wenn ein bundesgesetzlicher Auftrag vorliegt. Ein Beispiel dafür sind die Daten der amtlichen Lärmkartierung, die von den erhebenden Stellen (meist Kommunen mit mehr als 100.000 Einwohnern) als Karten publiziert werden. Die Rohdaten werden von jeder erhebenden Stelle über das Daten-Repositoryum European Environment Information and Observation Network (EIONET) zum Download angeboten (<http://cdr.eionet.europa.eu/>), ohne jedoch harmonisiert worden zu sein. Die Verwendung der Daten für die Bundesrepublik setzt deshalb eine zeitaufwendige Harmonisierung aller einzelnen Datensätze (ca. 500) voraus. Für einige Gebiete liegen die Daten dennoch nur als simple PDF-Karten vor und können deshalb nicht genutzt werden oder werden nur unter restriktiven Nutzungslizenzen freigegeben (Schweers et al. 2016; Müller et al. 2017). Ein ähnliches Bild zeigt sich bei Geodaten zu Versorgungs- und Verkehrsinfrastrukturen, die in den öffentlichen Katalogen ausgewiesen werden. Das Erhebungsgebiet ist meist auf einzelne Städte oder Kreise begrenzt. Daten, die eine flächendeckende Analyse der Bundesrepublik ermöglichen, sind die Aus-

nahme. Auch Daten, die nach gesetzlichem Auftrag für Gesamtdeutschland erhoben werden, sind für die Forschung nicht ohne Weiteres zugänglich. Zum Beispiel erhebt der TÜV-Rheinland im Auftrag des Bundesverkehrsministeriums Daten zur Verfügbarkeit von Breitbandinternetanschlüssen in Deutschland. Diese Daten werden zwar als graphische Kartenanwendungen im Internet publiziert, allerdings fehlt ein Dienst für Abfragen von Rohdaten oder einzelnen Datenpunkten für Zwecke wissenschaftlicher Auswertungen. Die Zugangsmodalitäten zu vielen Geodaten der öffentlichen Verwaltung sind weiterhin unregelt.

Der Zugang zu Geofachdaten aus wissenschaftlichen Quellen ist in der Regel einfacher. Neben den Beständen des FDZ Ruhr am RWI bietet auch das Leibniz-Institut für Ökologische Raumentwicklung (IÖR) Geodaten für die wissenschaftliche Nutzung an, insbesondere Daten zur Flächennutzung (den IÖR-Monitor, <http://www.ioer-monitor.de>). Die angebotenen Daten bieten etwa eine große Bandbreite an Indikatoren zu Siedlungsstrukturen, Landschaftsqualität oder Verkehr, die auch für die sozialwissenschaftliche Forschung sehr interessant sind.

Die Bereitstellung von Geodaten zum Download ist zum Teil dadurch eingeschränkt, dass die Datenvolumen sehr groß sein können. Das Open Geospatial Consortium (OGC) hat aus diesem Grund ISO Normen für Geo-Web-Dienste entwickelt (Bundesamt für Kartographie und Geodäsie 2015, S. 18ff). Web Map Services (WMS) zum Beispiel generieren Karten aus Geodaten und stellen diese in Vektor- oder Rasterform zur Darstellung bereit. Web Feature Service (WFS) wiederum ermöglichen das Herunterladen der Daten oder einzelner Datenpunkte im Geography Markup Language (GML) Format.

### 3.4.6 Verwendung von Geodaten und Anonymisierung

Kritisch sind bei der räumlichen Verknüpfung von Befragungsdaten mit Geodaten Fragen des Datenschutzes. Geoinformationen können als Zusatzinformationen zu den Befragungsdaten die De-Anonymisierung der Befragungsteilnehmer erleichtern. Zwei Problemfelder stehen hier im Vordergrund.

Zum einen sind die Georeferenzen selbst nicht zwingend personenbezogene Daten. Wohnadressen hingegen, die als Geokoordinaten vorliegen, sind als personenbezogene Daten zu *behandeln* und müssen deshalb besonders geschützt werden. Der Zugang zu Geokoordinaten für die Sekundärnutzung der Daten wird in der Folge stark eingeschränkt sein, was mit erheblichen Einschränkungen für das Analysepotential der Daten einhergeht. Die Punktkoordinaten werden zum Beispiel für die Berechnung von Distanzmatrizen (die Distanzen zwischen den Wohnadressen der Teilnehmenden an einer Befragung) gebraucht. Distanzmatrizen sind für die Schätzung von Modellen aus der räumlichen Ökonometrie, insbesondere der spatial lag Modelle, notwendig. Sie dienen der Überprüfung, ob räumlich näher gelegene Probanden sich im Hinblick auf bestimmte Attribute ähnlicher sind als Probanden, die weiter auseinander liegen. Distanzmatrizen sind allerdings besonders kritisch, weil sie zur Rekonstruktion der exakten Positionsangaben verwendet werden können, auch wenn die Koordinaten nicht mehr im Datensatz enthalten sind. Auf diese zu

verzichten, stellt jedoch einen substantiellen Informationsverlust der Geodaten dar (LeSage et al. 2014). Um Datensätze mit Distanzmatrizen dennoch anonymisieren zu können, werden Verfahren der Verfremdung von Datenmatrizen entwickelt, die keine substantiellen Verzerrungen in den Ergebnissen der statistischen Auswertung einführen (Kroll und Schnell 2016; Goldenholz et al. 2018).

Ein Verfahren für die datenschutzkonforme Bearbeitung der Georeferenzen von Umfragedaten schlagen Schweers et al. (2016) vor. Das Verfahren beruht darauf, Georeferenzen und Befragungsdaten systematisch getrennt zu halten und statistische Auswertungen lediglich mit den Geoattributen durchzuführen. Den Befragten werden dazu Geoattribute als Merkmale zugewiesen und die Geokoordinaten aus den Daten entfernt.

Auch die Aggregation der Adresskoordinaten auf kleinflächige INSPIRE-Grids (insbesondere mit den Kantenlängen  $\leq 100$  Meter) sind kritisch, weil nur wenige Personen in den Grids verortet werden können. Vor allem bei geringer Bevölkerungsdichte sind die Daten unter Umständen als personenbezogen zu bewerten. Empirische Auswertungen zum Deanonymisierungspotential der INSPIRE-Grids bei Befragungsdaten wurden bislang jedoch noch nicht veröffentlicht.

Aus Gründen des Datenschutzes werden auch die Merkmale aus dem Zensus 2011, die für die INSPIRE-Grids publiziert werden, zum Teil geheim gehalten, wenn nur wenige Merkmalsträger in den Gitterzellen verortet werden. Für Nutzende ist in der Folge nicht ersichtlich, ob das Merkmal im Grid nicht vorliegt oder nur schwach besetzt ist. Je kleiner die Gitterzellen sind, umso akzentuierter sind potentiell daraus resultierende Verzerrungen (Statistische Ämter des Bundes und der Länder 2015, S. 67ff.).

Zuletzt muss bei der Verwendung von Geodaten die Verteilung der Merkmale berücksichtigt werden. Gerade an den Rändern der Verteilung kann anhand extremer Werte der Geoattribute auf die Gebietseinheit oder Fläche geschlossen und damit können unter Umständen auch Personen re-identifiziert werden. Mit steigender Anzahl von öffentlich zugänglichen Geoattributen steigt das Risiko, bestimmte Flächen zu identifizieren, weil die Merkmalskombinationen singular sein können.

### 3.4.7 Schlussbemerkungen

Geodaten gewinnen an Bedeutung für die Sozialforschung, weil sie Befragungsdaten ergänzen und deren Potential erweitern. Geoinformationssysteme steigern die Effizienz beim Management von Regionalinformationen, und Methoden der räumlichen Verknüpfung bieten neue Möglichkeiten bei der Modellierung der Wohnumfelder der Befragungsteilnehmenden. Mit der progressiven Umsetzung der europäischen INSPIRE Richtlinie wird die Verfügbarkeit von Geodaten noch weiter zunehmen. Aufgrund der föderalen Regierungs- und Verwaltungsstruktur in der Bundesrepublik liegen die meisten Geodaten jedoch nur für einzelne Kreise oder Gemeinden vor. Eine flächendeckende Erhebung oder Harmonisierung und Kumulation der bestehenden Daten ist vorrangig Aufgabe der Wissenschaft. Gerade Daten über Verkehrs- und Versorgungsinfrastrukturen könnten in Zukunft noch an Wert gewinnen. Ein neuer

Trend in der Sozialforschung ist die Verwendung von Positionsdaten aus Smartphones, um Bewegungsprofile zu erstellen und auszuwerten (Bluemke et al. 2017). Alle Instrumente dazu liegen vor und es ist nur noch eine Frage der Zeit, bis diese Geoinformationen von Sozialforschenden verwendet werden, um klassische Umfragen zu ergänzen (Stanley et al. 2018). Für eine sinnvolle Auswertung und Interpretation dieser Positions- und Bewegungsdaten werden erweiterte Geodaten über Infrastrukturen (z.B. Verkehr, Handel, Bildungswesen, Gesundheitswesen, Wirtschaft etc.) des alltäglichen Lebens benötigt, um die Logik der Bewegungsprofile besser zu verstehen. Solche Daten liegen flächendeckend für Deutschland noch nicht vor und deren Harmonisierung und Kumulation ist ein sehr aufwendiger Prozess.

