

# RatSWD Working Paper Series

RatSWD.  
Rat für Sozial- und  
Wirtschaftsdaten

287

## Für einen besseren Datenzugang im Bereich Regionaldaten

Stellungnahme aus dem  
Verein für Socialpolitik

Thomas K. Bauer, Christina Gathmann,  
Ines Helm, Christian Raffer, Sebastian Siegloch,  
Sandra Schaffner, Andreas Lichter

Dezember 2023

[www.ratswd.de](http://www.ratswd.de)

# RatSWD Working Papers

## des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

---

Die *RatSWD Working Papers*-Reihe startete Ende 2007. In dieser Online-Publikationsreihe werden konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

*RatSWD Working Papers* sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autorinnen bzw. Autoren dar und nicht die des RatSWD. Die Zuwendungsgeber des RatSWD haben die Publikationen nicht beeinflusst.

**Herausgeberin oder Herausgeber der RatSWD Working Papers-Reihe ist die/der Vorsitzende des RatSWD:**

seit 2020 Monika Jungbauer-Gans

2014–2020 Regina T. Riphahn

2009–2014 Gert G. Wagner

2007–2008 Heike Solga

# Für einen besseren Datenzugang im Bereich Regionaldaten

Stellungnahme aus dem Verein für Socialpolitik

Thomas K. Bauer<sup>1</sup>, Christina Gathmann<sup>2</sup>, Ines Helm<sup>3</sup>, Christian Raffer<sup>4</sup>,  
Sebastian Siegloch<sup>5</sup>, Sandra Schaffner<sup>6</sup>, Andreas Lichter<sup>7</sup>

Dezember 2023

<https://doi.org/10.17620/02671.91>

Dieses Paper ist außerdem beim Verein für Socialpolitik erschienen:  
[https://www.socialpolitik.de/de/vfs\\_stellungnahmen\\_daten\\_regionaldaten](https://www.socialpolitik.de/de/vfs_stellungnahmen_daten_regionaldaten)

<sup>1</sup> RWI Essen und Ruhr-Universität Bochum

<sup>2</sup> LISER- Luxembourg Institute of Socio-Economic Research and University of Luxembourg

<sup>3</sup> Ludwig-Maximilians Universität München

<sup>4</sup> Deutsches Institut für Urbanistik

<sup>5</sup> Universität zu Köln und ECONtribute

<sup>6</sup> RWI Essen

<sup>7</sup> Heinrich Heine Universität Düsseldorf

Kontakt: Prof. Dr. Thomas K. Bauer | [bauer@rwi-essen.de](mailto:bauer@rwi-essen.de)

## Abstract

Regionale Ungleichheiten sind ausgeprägt und persistent. Qualitativ hochwertige Regionaldaten sind notwendig, um auf Basis empirischer Analysen ein besseres Verständnis über die Ursprünge und Konsequenzen regionaler Disparitäten sowie der Wirksamkeit regionalpolitischer Maßnahmen zu erlangen. Vor dem Hintergrund vorliegender Probleme im Zugang zu und der Bereitstellung von hochwertigen Regionaldaten werden in diesem Papier Handlungsempfehlungen zu einer Verbesserung des Angebots an Regionaldaten für wissenschaftliche Analysen abgeleitet

**Keywords:** Regionaldaten

## Inhalt

1. Einleitung.....	4
2. Problemlage bei existierenden Regionaldaten .....	4
3. Zusammenführung von Regionaldaten .....	8
4. Fehlende Regionaldaten .....	9
5. Handlungsvorschläge.....	10
Literaturverzeichnis.....	12

# 1. Einleitung

Regionale Ungleichheiten, etwa bei Einkommen oder Beschäftigung, sind ausgeprägt und persistent. Um ein besseres Verständnis über die Ursprünge und Konsequenzen regionaler Disparitäten sowie der Wirksamkeit regionalpolitischer Maßnahmen zu erlangen, sind insbesondere empirische Analysen einer Vielzahl politisch höchst relevanter Fragen notwendig. Welche Ursachen hat regionale Ungleichheit? Welche Faktoren behindern eine Angleichung regionaler Lebensverhältnisse und welche Faktoren können eine Angleichung befördern? Welche politischen Maßnahmen können regionale Wirtschaftsräume und Arbeitsmärkte stärken?

Grundlage derartiger empirischer Analysen sind qualitativ hochwertige Regionaldaten. Eine Umfrage unter ca. 500 Mitgliedern des Vereins für Socialpolitik zu den Problemen des Datenzugangs für die Wissenschaft offenbarte jedoch insbesondere bezüglich der Qualität und der Zugangsmöglichkeiten zu Regionaldaten erhebliche Defizite. Nahezu 80 Prozent der Mitglieder mit einem regionalökonomischen Forschungsinteresse sind mit dem Zugang zu Regionaldaten eher oder sehr unzufrieden und 73 Prozent gaben an, dass sie Projekte aufgrund von Defiziten beim Zugang zu Regionaldaten nicht durchführen konnten. Folgende Probleme beim Datenzugang wurden von VfS-Mitgliedern mit einem regionalökonomischen Forschungsschwerpunkt am häufigsten genannt: vorhandene Daten werden nicht oder nicht detailliert genug bereitgestellt (jeweils 70 Prozent), fehlende Verknüpfbarkeit von Daten (75 Prozent) und organisatorische Hürden beim Datenzugang (53 Prozent).

Vor diesem Hintergrund fasst der vorliegende Text die Ergebnisse der Unterarbeitsgruppe (UAG) Regionaldaten in drei inhaltlichen Feldern zusammen. Kapitel 2 stellt die Problemlage bei der Bereitstellung und Verwendung existierender Regionaldaten dar. Kapitel 3 beschreibt die besondere Problematik der Zusammenführung verschiedener Regionaldatenbanken. Kapitel 4 widmet sich schließlich fehlenden Daten, die insbesondere für evidenzbasierte Politikberatung von Bedeutung sind. Die Stellungnahme schließt mit konkreten Vorschlägen zur Verbesserung der Datensituation im Bereich der Regionalökonomik, die zum Teil im geplanten Forschungsdatengesetz der Bundesregierung umgesetzt werden können. Dabei fokussieren sich die Vorschläge auf die speziellen Problemlagen bei der Verfügbarkeit und Nutzung von Regionaldaten.

## 2. Problemlage bei existierenden Regionaldaten

Zweifelsohne existieren unzählige Regionaldatenquellen bzw. Daten, die potenziell für empirische regionalökonomische Analysen verwendet werden können. Neben den amtlichen Daten des Statistischen Bundesamts, den INKAR-Daten des Bundesinstituts für Bau-, Stadt- und Raumforschung oder den Daten der Urban-Audit-Initiative der statistischen Ämter der Städte<sup>1</sup> stehen umfangreiche Daten mit Regionalbezug bei Forschungseinrichtungen (bspw.

---

<sup>1</sup> Siehe <https://www.staedtestatistik.de/arbeitsgemeinschaften/kosis/urban-audit>

Immobilienscout24-Daten beim FDZ Ruhr am RWI oder das georeferenzierte Sozioökonomische Panel beim DIW) und privaten Anbietern (bspw. microm, Infas360) zur Verfügung oder können über die Verwendung geeigneter Algorithmen bspw. aus Internetangeboten (bspw. openstreetmap oder Google) oder Satellitendaten gewonnen werden.

### ***Intransparenz des Datenangebots***

So erfreulich der Umstand eines potenziell sehr umfangreichen Angebots von Regionaldatenquellen sein mag, besteht für Wissenschaftlerinnen und Wissenschaftler das Problem einer weitgehenden Intransparenz der zur Verfügung stehenden regionalen Daten. Insbesondere bei Befragungsdaten und administrativen Daten ist häufig nicht bekannt, ob und über welchen Zeitraum Regionalindikatoren zur Verfügung stehen. Da Metainformationen über Regionaldaten vielfach fehlen und kaum standardisiert sind, erfolgt die Identifikation geeigneter Regionalinformationen für eine spezifische Forschungsfrage üblicherweise dezentral und händisch. Die Entwicklung eines geeigneten Metadaten-systems und – darauf aufbauend – eines zentralen Repositoriums für Regionaldaten könnte diese Intransparenz erheblich verringern.

### ***Aggregationsebene und Zugriff auf existierender Regionalindikatoren***

Nicht zuletzt hat die breit angelegte Umfrage unter den Mitgliedern des Vereins für Socialpolitik gezeigt, dass derzeit zur Verfügung stehende Regionaldaten häufig auf einer zu hohen regionalen Aggregationsebene bereitgestellt werden. Die Mehrzahl der Regionalindikatoren der amtlichen Statistiken stehen lediglich auf Kreisebene zur Verfügung, vergleichsweise wenige auf Gemeindeebene. Kleinräumige Analysen, wie bspw. zur Entwicklung wirtschaftlicher und gesellschaftlicher Ungleichheiten in den Lebensverhältnissen innerhalb einzelner Städte, sind damit häufig nicht möglich. Vor allem aus datenschutzrechtlichen Gründen werden Regionalinformationen in Befragungsdaten von Personen oder Unternehmen häufig nicht oder nur auf einer relativ stark aggregierten regionalen Ebene zur Verfügung gestellt, da eine Verknüpfung selbst faktisch anonymisierter Befragungsdaten mit Regionaldaten aus anderen Quellen die Identifikation einzelner Individuen oder Unternehmen stark vereinfachen könnte (siehe Kapitel 3 zu den speziellen Problemen der Verknüpfung von Regionaldaten). So ist bspw. die für die Wissenschaft zur Verfügung stehende niedrigste Aggregationsebene im Regionalfile des Mikrozensus die sogenannte Mikrozensuskreisregion (MZKR), d.h. konstruierte Raumeinheiten mit mindestens 100.000 Einwohnern. Zwar kann man dem Mikrozensus prinzipiell auf Gemeindeebene Information zuspähen, jedoch keine Auswertungen auf Gemeindeebene durchführen. Regionalökonomische Analysen schließen sich damit letztlich fast aus.

Neben dem bereits erwähnten Metadaten-system, das eine erhöhte Transparenz hinsichtlich der Frage schaffen könnte, welche regionale Ebene theoretisch abbildbar wäre, sind wir überzeugt, dass eine Reihe an Lösungsmöglichkeiten für die skizzierten Probleme existieren (siehe im Detail: Kapitel 5). So könnte die Bereitstellung möglichst kleinräumiger Daten bspw.

im Rahmen von Gastaufenthalten oder einer Ausweitung der Möglichkeiten des Datenzugriffs über (geprüfte) Remote Job Submission oder Remote Desktop erfolgen. Darüber hinaus sollte sichergestellt werden, dass kleinräumige Regionalinformationen nicht bereits bei der Datenaufbereitung gelöscht werden, sondern zumindest gespeichert und einem Datentreuhänder übergeben werden.

### ***Fehlender Zugriff auf existierender Regionalindikatoren***

Es gibt prinzipiell eine Vielzahl von Befragungsdaten und administrativen Daten, für die auch sehr kleinräumige Regionalinformationen zur Verfügung stehen. Auf viele dieser Daten hat die Wissenschaft jedoch keinen oder nur sehr eingeschränkten Zugriff. Einige Beispiele:

- Am Institut für Arbeitsmarkt- und Berufsforschung (IAB) existieren für jeden sozialversicherungspflichtigen Beschäftigten kleinräumige Information zum Wohn- und Arbeitsort. Auf diese Daten kann jedoch nur für interne Zwecke oder über Forschungsk Kooperationen mit Beschäftigten des IABs zugegriffen werden. Die Analyse regionaler Arbeitsmärkte, der regionalen Mobilität von Beschäftigten oder der Determinanten des Pendlerverhaltens wird damit stark eingeschränkt.
- Der AfiD-Datensatz wird nicht von allen Bundesländern zur Verfügung gestellt.
- Trotz seiner erheblichen wirtschaftlichen und gesellschaftlichen Bedeutung gibt es in Deutschland nur sehr eingeschränkte regional detaillierte Informationen zu Immobilienmärkten. Viele Wissenschaftlerinnen und Wissenschaftler greifen bisher auf Daten von Immobilienscout24 zurück, die vom FDZ Ruhr am RWI externen Wissenschaftlerinnen und Wissenschaftlern zur Verfügung gestellt werden. Obwohl Immobilienscout24 eine sehr hohe Marktabdeckung aufweist, ist unklar, wie repräsentativ diese Daten für den gesamten Immobilienmarkt in Deutschland sind. Zudem werden in diesem Datensatz nur Angebotspreise jedoch keine Transaktionspreise erfasst. Diese Probleme könnten mit der Bereitstellung der Verkaufsdaten der Gutachterausschüsse von Immobilien, die auf Adressenebene existieren, behoben werden. Die Daten könnten somit georeferenziert und über ein Forschungsdatenzentrum zur Verfügung gestellt werden.
- Ein ähnliches Informationsdefizit existiert für (lokale) Mieten. Die Daten von Immobilienscout24 liefern lediglich Informationen zu Mietpreisen neuvermieteter Immobilien. Über Bestandsmieten existieren hingegen nur wenige Informationen. Im Mikrozensus werden Mieten nur alle vier Jahre erfasst und nicht mit kleinräumigen Identifikatoren zur Verfügung gestellt. Die georeferenzierte Version des Sozioökonomischen Panels als auch das EU-SILC sind in ihrem Befragungsumfang zu klein, um lokale Wohnungsmärkte adäquat analysieren zu können. EU-SILC stellt zudem keine kleinräumigen Regionalindikatoren zur Verfügung. Informationen im Rahmen der Wohnungs- und Gebäudezählung des Zensus werden der Wissenschaft nicht zur Verfügung gestellt und zudem vergleichsweise zeitnah gelöscht. Die den Statistischen Landesämtern zur Verfügung stehenden Individualdaten von Wohngeldempfängern

könnten das Problem der fehlenden Informationen zu Bestandsmieten zumindest teilweise – insbesondere im unteren Preissegment -- lösen. Zwar stellen diese Daten keine repräsentative Stichprobe aller Bestandsmieten dar. Die Panelstruktur des Datensatzes würde es jedoch erlauben, zumindest eine Analyse der Entwicklung der Bestandsmieten im unteren Bereich der Verteilung der Mieten durchzuführen.

- Beim Taxpayer Panel, der wichtigsten Mikro-Steuerstatistik des Landes, sind Gemeindeinformationen erst ab 2012 verfügbar, für Bayern allerdings nur in pseudonomisierter Form.<sup>2</sup> In den Jahren von 2001 bis 2011 ist die tiefste regionale Auswertungsebene das Bundesland.

Es existiert eine Vielzahl weiterer Datensätze mit (kleinräumigen) Regionalinformationen, die der Wissenschaft nicht oder lediglich mit aggregierten Regionalinformationen zur Verfügung gestellt werden, was wiederum regionalökonomische Analysen in einer Vielzahl von Fällen unmöglich macht. Beispiele umfassen Daten der statistischen Ämter größerer Gemeinden, die teilweise Informationen auf Stadtteilbezirken umfassen, Grundsteuerdaten, Daten der Bundesnetzagentur, eine Vielzahl von Befragungsdaten wie bspw. das Nationale Bildungspanel oder der ALLBUS. Die Erfassung bzw. Beibehaltung dieser Informationen, die Abkehr von regionaler Aggregation, die Bereitstellung von Informationen in einem Metadatenystem und die datenschutzadäquate Bereitstellung der Daten mit Regionalidentifikatoren auf der jeweils kleinstmöglichen regionalen Ebene für Wissenschaftler könnten die evidenzbasierte Politikberatung im Bereich der Regionalökonomik erheblich fördern (siehe Kapitel 5).

### ***Kosten des Datenzugangs***

Der Bezug von regionalen amtlichen Daten aus mehr als einem Bundesland, die nicht bereits über [www.regionalstatistik.de](http://www.regionalstatistik.de) für alle Bundesländer angeboten werden, ist sehr aufwendig und mit erheblichen Kosten verbunden. In der Regel erfolgt die Bereitstellung derartiger Daten über eine "Koordinierte Abfrage", die vom Statistischen Landesamt des Bundeslandes der jeweiligen Wissenschaftlerin bzw. des jeweiligen Wissenschaftlers durchgeführt wird. Dieses fragt dann die Möglichkeit und die Kosten der Bereitstellung der Daten bei allen anderen Landesämtern an und erstellt ein Angebot. Dieses Prozedere kann dabei mehrere Wochen dauern.<sup>3</sup> Ebenso problematisch sind die Kosten der Bereitstellung dieser Daten. Insbesondere bei Paneldaten können diese Kosten sehr hoch sein, da die Rechnungsgrundlage der Datenbereitstellung nicht nur die Anzahl der Einheiten umfasst, sondern für jedes bezogene Jahr gesondert gezahlt werden muss. Insbesondere für Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler in der Postdoc-Phase stellen diese Kosten der Datenbereitstellung eine prohibitive Hürde dar. Vor dem Hintergrund diese Problemlage wäre eine standardmäßige Übermittlung und Harmonisierung aller Datenbestände der Landesämter beim Forschungsdatenzentrum des Statistischen Bundesamts sinnvoll. Mit dieser Zentralisierung der regionalen Datenbestände

---

<sup>2</sup> Siehe <https://www.forschungsdatenzentrum.de/de/steuern/tpp>

<sup>3</sup> In Frankreich kann man bspw. einen Antrag auf Datenbereitstellung digital auf einer zentralen Plattform einreichen. Die verantwortlichen Datenbereitsteller prüfen den Antrag und stellen gegebenenfalls Nachfragen. Üblicherweise wird der Antrag innerhalb von wenigen Tagen genehmigt.



würde der kostenintensive Aufwand für koordinierte Abfragen entfallen. Entsprechend könnten die Daten der Wissenschaft zu geringeren Kosten, bestenfalls aber kostenlos zur Verfügung gestellt werden.

### 3. Zusammenführung von Regionaldaten

Existierende Regionalinformationen liegen in vielen Formen vor. Vorwiegend werden regionale Daten auf Basis der Grenzen administrativer Verwaltungseinheiten (Gemeinden, Kreise, Regierungsbezirke, Bundesländer) oder unterschiedlich abgegrenzter Gebietseinheiten (wie beispielsweise auf Ebene der Arbeitsamtsbezirke, Mikrozensuskreisregion oder Raumordnungsregionen) zur Verfügung gestellt, wobei letztere nicht in allen Fällen mit den Grenzen administrativer Verwaltungseinheiten kompatibel sind. In Einzelfällen werden Daten in Form von Shape-Files zur Verfügung gestellt, wobei aber für die meisten Gemeinden in Deutschland keine Shapefiles einzelner Stadtteile existieren. Daten privater Anbieter werden häufig in Straßenabschnitten gruppiert, bei landwirtschaftlichen Daten und dem Zensus kommen hingegen regionale Raster zum Einsatz, wobei bspw. die einzelnen Beobachtungseinheiten in Raster der Größe 500mx500m aggregiert werden. Diese Raster sind europaweit verfügbar, da sie der europäischen INSPIRE Richtlinie folgen. Schließlich stehen für eine steigende Zahl von Regionalinformationen exakte Geopositionen (mit Angabe des Breiten- und Längengrads) zur Verfügung. Diese Daten werden üblicherweise aus Satellitendaten, Kartenangeboten aus dem Internet (bspw. via openstreetmap) oder durch Georeferenzierung von Adressen gewonnen. Relativ wenige amtliche Daten und Befragungsdaten stellen exakte Geopositionen der erfassten Beobachtungseinheiten zur Verfügung (Ausnahmen sind die georeferenzierte Version des SOEP, Immobilienscout24, der Unfallatlas der statistischen Ämter), wobei Wissenschaftlerinnen und Wissenschaftler zu einigen dieser Daten nur einen eingeschränkten Zugang haben (bspw. die georeferenzierten Beschäftigungsdaten des IAB). Für amtliche Daten und Befragungsdaten sollte die exakte Geoposition der Beobachtungseinheiten prinzipiell zur Verfügung stehen. Diese werden aber zumeist aus datenschutzrechtlichen Gründen der Wissenschaft nicht zur Verfügung gestellt.

Aufgrund der skizzierten heterogenen Formen der zur Verfügung stehenden Regionaldaten ergeben sich für Wissenschaftlerinnen und Wissenschaftler erhebliche Herausforderungen. Stehen unterschiedliche Datensätze oder verschiedene relevante Variablen nur auf unterschiedlichen regionalen Ebenen zur Verfügung, müssen die Daten auf die jeweils höchste zur Verfügung stehende Ebene aggregiert werden. Will man bspw. Befragungsdaten, in denen für die einzelnen Beobachtungseinheiten lediglich Regionalidentifikatoren auf der Bundeslandebene zur Verfügung stehen, mit Regionalinformationen aus anderen Datenquellen verknüpfen, müssen letztere auf die Bundeslandebene aggregiert werden. Häufig ist dies aufgrund der fehlenden Kompatibilität verschiedener Raumabgrenzungen aber nicht eindeutig möglich. Darüber hinaus werden aufgrund der Notwendigkeit der Aggregation der Regionalinformationen auf die jeweils höchste zur Verfügung stehenden regionalen

Abgrenzung die adäquat zu analysierenden regionalökonomischen Fragestellungen erheblich eingeschränkt.

Darüber hinaus existiert das Problem, dass sich verschiedene regionale Abgrenzungen über die Zeit ändern können, bspw. aufgrund von Eingemeindungen oder Kreisreformen. Damit wird die Generierung von Paneldaten, die gerade für regionalökonomische Analysen notwendig sind, erheblich erschwert. Zwar haben verschiedene Wissenschaftlerinnen und Wissenschaftler und Institutionen Routinen entwickelt, die versuchen, veränderte Raumabgrenzungen über die Zeit zu harmonisieren.<sup>4</sup> Diese stehen aber lediglich für einen begrenzten Zeitraum (bspw. seit 1990) und nicht für alle in verschiedenen Daten verwendeten Raumabgrenzungen zur Verfügung.

Aus diesen Problemfeldern ergeben sich folgende zentrale Handlungsempfehlungen (siehe auch Kapitel 5). Zum einen sollten regionale Daten bzw. Regionalidentifikatoren in Individual- und Unternehmensdaten grundsätzlich auf der jeweils niedrigsten verfügbaren regionalen Aggregationsebene – vorzugsweise als Geoposition – nachhaltig gespeichert und der Wissenschaft in einer datenschutzkonformen Version zur Verfügung gestellt werden. So sollte man bspw. die exakten Geopositionen des Wohnorts der Befragten des Mikrozensus gesichert (bspw. bei einem Datentreuhänder) speichern. Dies würde es ermöglichen, Daten jedweder Raumabgrenzung (bspw. räumliche Raster oder Gemeindeebene) dem Mikrozensus zuzuspielen, wobei sichergestellt werden kann, dass in jeder möglichen Raumabgrenzung eine Mindestzahl von Mikrozensusbeobachtungen (bspw. 20 Beobachtungen, jedoch nicht 100.000 Beobachtungen wie bisher) vorliegt, damit die resultierenden Daten faktisch anonymisiert werden. Raumabgrenzungen, die diese Mindestzahl unterschreiten, können gelöscht oder mit benachbarten Räumen zusammengespielt werden. Darüber hinaus könnte man die Datenbereitstellung über Remote Access oder Gastwissenschaftleraufenthalte organisieren, wobei die Analyseergebnisse der Wissenschaftler vor der Herausgabe geprüft werden. Ist die Ermittlung der Geoposition nicht möglich, sollte die ausgewiesene administrative Einheit zumindest dem jeweiligen Gebietsstand entsprechen und Routinen bereitgestellt werden, die eine Übertragung auf ältere oder höher aggregierte Gebietsstände ermöglichen.

## 4. Fehlende Regionaldaten

Für viele regionalökonomische Fragestellungen, wie beispielsweise die Analyse regionaler Ungleichheiten, wären Informationen über räumliche Preisniveauunterschiede zentral, nicht zuletzt da regionale Unterschiede in den nominalen Einkommen nur unzureichende Aussagekraft über die regionale reale Kaufkraft zulassen. Deshalb wird von der Wissenschaft und der Politik seit langem die Erhebung regionaler Preisindizes gefordert. Konzeptionen zur Machbarkeit der Erhebung regionaler Preisindizes liegen vor.<sup>5</sup> Das Bundesinstitut für

---

<sup>4</sup> Siehe bspw.

<https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/umstiegsschlüssel/umsteigeschlüssel.html>

<sup>5</sup> Siehe bspw. Möller, Hohmann und Huschka (Hrsg.) (2010) und Bundesinstitut für Bau-, Stadt- und Raumforschung (2009).

Bau-, Stadt- und Raumforschung (BBSR) und das IW Köln planen im Herbst dieses Jahres einen regionalen Preisindex vorzustellen. Nichtsdestotrotz sehen wir in diesem Bereich dringenden Handlungsbedarf, um eine jährliche Erhebung regionaler Preisindizes schnellstmöglich umzusetzen. Zudem wäre eine flächendeckende Bereitstellung der Daten der regionalen Gutachterausschüsse wünschenswert, um das erhebliche Informationsdefizit über den deutschen Immobilienmarkt zumindest teilweise zu beheben. Nicht zuletzt wären Daten zum Handel von Gütern und Dienstleistungen zwischen Regionen innerhalb Deutschlands bedeutsam, um regionale wirtschaftliche Verflechtungen analysieren zu können.

Die Entwicklung regionaler Disparitäten ist ein langfristiger Prozess. Daher erfordert die empirische Analyse der Determinanten und der Dynamik regionaler Ungleichheiten die Verfügbarkeit sehr langer Beobachtungszeiträume. Nicht zuletzt deshalb versuchen Wissenschaftlerinnen und Wissenschaftler verstärkt, historische Bestände an regionalen Indikatoren zu identifizieren und maschinenlesbar zu digitalisieren.<sup>6</sup> Trotz erheblicher technischer Fortschritte in der automatisierten Erkennung und Verarbeitung von Texten (Textmining), ist die Übertragung und Georeferenzierung digitalisierter Bibliotheksbestände in ein Format, das von statistischen Softwarepaketen gelesen werden kann, nach wie vor personal- und zeitintensiv. Dies liegt neben den bereits erwähnten Veränderungen von Raumidentifikatoren (Gemeinde- oder Kreisgrenzen) über die Zeit bspw. auch an zeitlichen Veränderungen der Schreibweisen von Gemeindenamen (Constanz vs. Konstanz). Darüber hinaus wurde der Bestand der in Bibliotheken zur Verfügung stehenden historischen Veröffentlichungen der Statistischen Ämter (wie bspw. dem Statistischen Amt des Deutschen Reiches oder der Weimarer Republik) bisher noch nicht vollständig digitalisiert. Hier liegt aus unserer Sicht ein weitgehend ungehobener Schatz von Regionaldaten, deren Erfassung die Forschung zu den Determinanten und der Entwicklung regionaler Disparitäten erheblich befördern würde. Die Erfassung dieser Daten erfordert die Bereitstellung von Mitteln zur Digitalisierung der in Bibliotheken zur Verfügung stehenden historischen statistischen Informationen und Übertragung in maschinenlesbare Formate und Dokumentation. Bereits existierende historische Regionalinformationen müssten darüber hinaus in das bereits in Kapitel 2 geforderten Metadatensystem eingepflegt werden.

## 5. Handlungsvorschläge

Aus der Analyse der bestehenden Situation des Zugangs zur Regionaldaten in Deutschland ergeben sich aus Sicht des Vereins für Socialpolitik folgende Bedarfe:

1. Eine qualitative Verbesserung des Angebots an Regionaldaten in Deutschland erfordert die Entwicklung eines geeigneten Metadatensystems und darauf aufbauend die systematische Erfassung, Speicherung und Bereitstellung bereits existierender amtlicher und nicht-amtlicher Regionalinformation in einem einheitlichen

---

<sup>6</sup> Siehe bspw. die in histat (<https://histat.safe-frankfurt.de/>) zur Verfügung gestellten Daten oder die ifo Prussian Economic History Database iPEHD (<https://www.ifo.de/willkommen-bei-ipehd-der-ifo-prussian-economic-history-database>).

Repository, um die derzeit bestehende Intransparenz des Bestands an Regionalinformationen aufzulösen. Das Repository könnte in einer zentralen Informationsstelle für Forschungsdaten angesiedelt werden. Die Erfassung sollte dabei interdisziplinär erfolgen (bspw. unter Einbeziehung existierender Daten der Raumforschung) und die Ämter der Städtestatistik sowie historische Datenbestände einbeziehen. Darüber hinaus sollte das regionale Metadatensystem die potenziell niedrigste Ebene administrativer Daten und mit öffentlichen Mitteln finanzierter Befragungsdaten erfassen.

2. Die Löschung von Regionalidentifikatoren in Datensätzen sollte mittelfristig ausgeschlossen werden. Regelungen sind so anzupassen, dass die möglichst niedrigste räumliche Ebene einer Beobachtungseinheit nachhaltig als exakter Geocode gespeichert wird. Bei Individualdaten wäre dies die Wohnadresse, bei Unternehmen und Betrieben deren Standortadressen. Damit könnte man eine flexible Aggregation der Daten sowie eine adäquate Verknüpfung verschiedener Regionaldaten sicherstellen.
3. Die Verknüpfung von Mikrodaten mit Regionalinformationen sollte auf der niedrigsten regionalen Aggregationsebene angestrebt werden. Die datenschutzkonforme Bereitstellung derartiger Daten könnte über einen Datentreuhänder in einer zentralen Informationsstelle für Forschungsdaten erfolgen und Wissenschaftler über einen Remote Access mit Outputprüfung oder Gastwissenschaftleraufenthalt zur Verfügung gestellt werden.
4. Amtlichen Daten mit Regionalinformationen, die nicht bereits über [www.regionalstatistik.de](http://www.regionalstatistik.de) für alle Bundesländer angeboten werden, sollten standardmäßige bei Statistischen Bundesamt vorgehalten und der Wissenschaft kostenlos zur Verfügung gestellt werden, damit die derzeit für alle Beteiligten kosten- und zeitintensiven koordinierten Abfragen bei allen Statistischen Landesämtern entfallen können.
5. Zu existierenden Datenbeständen mit kleinräumigen Informationen, wie bspw. den Daten der regionalen Gutachterausschüsse oder georeferenzierten Versionen von Sozialversicherungsdaten, sollten der Wissenschaft zur Verfügung gestellt bzw. der Zugang erleichtert werden. Schließlich wäre eine stärkere Förderung der Identifikation, Digitalisierung und maschinenlesbaren Bereitstellung von statistischen Archivmaterialien wünschenswert.
6. Bedeutende Lücken in der Bereitstellung von Regionalinformationen sollten schnellstmöglich behoben werden. Dies beinhaltet insbesondere die Erhebung jährlicher regionaler Preisindizes, aber bspw. auch die Erfassung des Handels von Gütern und Dienstleistungen zwischen Regionen innerhalb Deutschlands.

## Literaturverzeichnis

Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) (Hrsg.) (2009): *Regionaler Preisindex*.  
Berichte, Bd. 30, Bonn.

Möller, J., Hohmann, E., & Huschka, D. (Hrsg.). (2010). *Der weiße Fleck - zur Konzeption und  
Machbarkeit regionaler Preisindizes* (IAB-Bibliothek (Gutachten), 324). Bielefeld: W.  
Bertelsmann Verlag. <https://doi.org/10.3278/300713w>

# Impressum

## Kontakt:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Geschäftsstelle

Am Friedrichshain 22

10407 Berlin

[office@ratswd.de](mailto:office@ratswd.de)

<https://www.ratswd.de>

Die Geschäftsstelle des RatSWD wird als Teil von KonsortSWD im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.

Berlin, Dezember 2023



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

DOI: [10.17620/02671.91](https://doi.org/10.17620/02671.91)

## Zitationsvorschlag:

Bauer, T. K., Gathmann, C., Helm, I., Raffer C., Siegloch, S., Schaffner S. & Lichter, A. (2023). *Für einen besseren Datenzugang im Bereich Regionaldaten. Stellungnahme aus dem Verein für Socialpolitik*. (RatSWD Working Paper 287/2023). Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.91>