



NEPS WORKING PAPERS

Katarina Weßling und Alexandra Wicht

# MÖGLICHKEITEN DER NUTZUNG REGIONALER DATEN IM NATIONALEN BILDUNGSPANEL – EIN ÜBERBLICK

NEPS Working Paper No. 54  
Bamberg, April 2015

**Working Papers of the German National Educational Panel Study (NEPS)**

at the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg

The NEPS Working Papers publish articles, expertises, and findings related to the German National Educational Panel Study (NEPS).

The NEPS Working Papers are edited by a board of researchers representing the wide range of disciplines covered by NEPS. The series started in 2011.

Papers appear in this series as work in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

Any opinions expressed in this series are those of the author(s) and not those of the NEPS Consortium.

The NEPS Working Papers are available at

<https://www.neps-data.de/projektübersicht/publikationen/nepsworkingpapers>

**Editorial Board:**

Jutta Allmendinger, WZB Berlin

Cordula Artelt, University of Bamberg

Jürgen Baumert, MPIB Berlin

Hans-Peter Blossfeld, EUI Florence

Wilfried Bos, University of Dortmund

Claus H. Carstensen, University of Bamberg

Henriette Engelhardt-Wölfler, University of Bamberg

Frank Kalter, University of Mannheim

Corinna Kleinert, IAB Nürnberg

Eckhard Klieme, DIPF Frankfurt

Cornelia Kristen, University of Bamberg

Wolfgang Ludwig-Mayerhofer, University of Siegen

Thomas Martens, DIPF Frankfurt

Manfred Prenzel, TU Munich

Susanne Rässler, University of Bamberg

Marc Rittberger, DIPF Frankfurt

Hans-Günther Roßbach, LifBi

Hildegard Schaeper, DZHW Hannover

Thorsten Schneider, University of Leipzig

Heike Solga, WZB Berlin

Petra Stanat, IQB Berlin

Volker Stocké, University of Kassel

Olaf Struck, University of Bamberg

Ulrich Trautwein, University of Tübingen

Jutta von Maurice, LifBi

Sabine Weinert, University of Bamberg

**Contact:** German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg – Germany – [contact@lifbi.de](mailto:contact@lifbi.de)

# Möglichkeiten der Nutzung regionaler Daten im Nationalen Bildungspanel – ein Überblick

*Katarina Weßling & Alexandra Wicht*

## **Autoreninformationen**

**Katarina Weßling** Universität Tübingen, Institut für Soziologie,  
E-Mail: katarina.wessling@uni-tuebingen.de.

**Alexandra Wicht** Universität Siegen, Seminar für Sozialwissenschaften,  
E-Mail: wicht@soziologie.uni-siegen.de

Die Autorinnen sind in alphabetischer Reihenfolge genannt.

Diese Arbeit wurde gefördert durch die Deutsche Forschungsgemeinschaft unter den Fördernummern HI 767/6-1 und LU 486/8-1.

## **Bibliographische Angaben**

K. Weßling & A. Wicht (2015). *Möglichkeiten der Nutzung regionaler Daten im Nationalen Bildungspanel – ein Überblick* (NEPS Working Paper No. 54). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.

# Möglichkeiten der Nutzung regionaler Daten im Nationalen Bildungspanel – ein Überblick

## Zusammenfassung

Um den Einfluss regionaler Kontexte auf Prozesse des Bildungserwerbs analysieren zu können, ist die Zusammenführung von Individualdaten mit raumstrukturellen Indikatoren unerlässlich. Besonders die Befragungsdaten des Nationalen Bildungspanels (NEPS) stellen hier eine vielversprechende Ressource dar. Um ein vereinfachtes Matching von Befragungsdaten mit Kontextdaten im Zeitreihenformat zu ermöglichen, gibt das vorliegende Working Paper einen Überblick über Veränderungen administrativer Gebiete in Deutschland und über Möglichkeiten, Gebietsstände und Indikatoren anzupassen und zu Zeitreihen zu vereinheitlichen. Für diesen Zweck ist dem Working Paper eine harmonisierte Tabelle der Kreiskennziffern ab 1980 im Stata- und Excelformat beigelegt. NEPS Nutzer\_innen können auf diese Tabelle über RemoteNEPS zugreifen. Darüber hinaus wird eine Übersicht über kommerzielle und nicht-kommerzielle Regionaldaten gegeben, die für das NEPS fruchtbar gemacht werden können.

## Schlagwörter

Regionale Kontextanalysen, Regionaldaten, Harmonisierung von Kreiskennziffern

## Abstract

Analyzing the impact of regional contexts on educational attainment processes requires the combination of individual data with structural locational indicators. The National Educational Panel Study (NEPS) represents a promising source of individual survey data. In order to allow for a simplified matching of survey and contextual time series data, this working paper provides, on the one hand an overview over changes in administrative regions in Germany. On the other hand, it demonstrates possibilities for adjusting territorial statuses and indicators to consistent time series data. For this purpose a harmonized table of administrative district codes from 1980 onwards in Stata and Excel is attached to the paper. NEPS user have access to this table via RemoteNEPS. Finally, an overview of commercial and non-commercial regional data is given that can be made fruitful for the NEPS.

## Keywords

Regional contextual analyses, regional data, harmonization of district codes

## 1 Einführung und Hintergrund des Arbeitspapiers

Die Zusammenführung von Daten des Nationalen Bildungspanels (NEPS) mit administrativen Regionaldaten stellt eine bedeutende Ressource für die Analyse von Kontexteinflüssen auf Bildungsprozesse dar. Voraussetzung für die Zusammenführung ist die Kodierung der Adressdaten der Studienteilnehmer\_innen des NEPS sowie deren Schul-, Arbeits- und Ausbildungsorte. Dies geschieht auf Basis raumbezogener Identifikationsvariablen. Über solche „Schlüssel“ lassen sich räumliche Kontextinformationen verschiedener Aggregationsniveaus an die Individualdaten des NEPS anspielen. Auf diese Weise besteht für Datennutzer\_innen der Startkohorte 6 („Erwachsene“) die Möglichkeit, eigenständig recherchierte Daten auf der Ebene von Landkreisen und kreisfreien Städten via „Remote NEPS“ über die Kreiskennziffer (KKZ) an die Befragungsdaten anzuspielen (siehe Datenzugangswege des NEPS). Im Falle anderer Startkohorten unterliegen Regionalanalysen stärkeren Restriktionen.

Die Kodierung der Adressdaten erfolgt nicht direkt über das Forschungsdatenzentrum des NEPS, sondern über das Datenerhebungsinstitut infas. Grundlage hierfür ist eine Ortsliste, die dem Amtlichen Gemeindegrenzenverzeichnis (AGS) im Jahr 2001 entspricht. Der räumliche Bezug aller im NEPS verfügbaren Individualdaten wurde somit gewissermaßen auf den Gebietsstand im Jahr 2001 „eingefroren“. Administrative räumliche Einheiten sind jedoch über die Zeit variabel: Im Rahmen von Gebietsreformen werden Gebietsgrenzen neu gezogen und auch Schlüssel neu vergeben. Dies erschwert die Zuweisung (öffentlich oder kommerziell verfügbarer) raumstruktureller Daten an jegliche Art von Befragungsdaten. Aus diesem Grund spielt gerade der zeitliche Bezug der räumlich definierten Kontexte eine zentrale Rolle in regionalen Kontextanalysen (vor allem im Längsschnitt).

Die meisten öffentlich zugänglichen Regionaldaten werden entweder mit den KKZ zum aktuellen Gebietsstand veröffentlicht oder (insbesondere im Falle weit zurückliegender Daten) zum jeweiligen Gebietsstand. Um regionale Kontextdaten (im Zeitreihenformat) an das NEPS anzuspielen, ist folglich eine Harmonisierung der im NEPS verfügbaren Schlüssel mit denen zum Gebietsstand des jeweils relevanten Analysezeitpunkts oder -fensters erforderlich.

Das vorliegende Working Paper soll dazu beitragen den Nutzer\_innen der NEPS-Daten das Anspielen regionaler Kontextinformationen zu erleichtern. Zunächst werden wir einen Überblick über Gebietsreformen in Deutschland geben und eine mögliche Vorgehensweise zur Harmonisierung von KKZ auf Basis von Informationen aus dem Gemeindeverzeichnis-Informationssystem (GV-ISys) vorstellen. Einen Datensatz (Stata) mit den harmonisierten KKZ für den Zeitraum von 1980–2013 (Westdeutschland) bzw. 1996–2013 (West- und Ostdeutschland) stellen wir begleitend zu diesem Working Paper zur Verfügung. Dieser Datensatz ermöglicht ein einfaches Anspielen von Regionaldaten auf der Ebene von Kreisen und kreisfreien Städten zu verschiedenen Zeitpunkten bzw. für verschiedene Gebietsstände. Darüber hinaus erläutern wir, wie inhaltliche Indikatoren in sich veränderten Gebieten mittels öffentlich zugänglicher Gewichtungsfaktoren korrigiert werden können, um Zeitreihen spezifischer Kontextdaten zu erhalten. Anschließend geben wir einen Überblick über administrative und kommerzielle Regionaldaten, die für das (NEPS) fruchtbar gemacht werden können.

## 2 Administrative Grenzen im Zeitverlauf

Kreisgrenzen sind dynamisch und unterliegen einem fortlaufenden politisch und wirtschaftlich motivierten Wandel. Im Rahmen von Gebietsreformen – die potenziell zu jeder Zeit und auf allen administrativen Ebenen stattfinden können – werden Grenzen neu gezogen und Gebiete umbenannt. Tabelle 1 gibt einen ersten Überblick über die Veränderungen auf Kreisebene seit 1980. Aktuell existieren in Deutschland 107 kreisfreie Städte und 295 Landkreise.

Tabelle 1: Gebietsveränderungen auf Kreisebene seit 1980 in Deutschland  
(BBSR 2010, Wikipedia 2014)

Bundesland	Jahr	Kreisfreie Städte	Landkreise	Anzahl Landkreise/ Kreisfreie Städte
Niedersachsen	1980	–	+1	328
Neue Bundesländer	1990	+26	+189	543
Brandenburg	1993	–2	–24	517
Mecklenburg-Vorpommern	1994	–	–19	498
Sachsen-Anhalt	1994	–	–16	482
Thüringen	1994	–	–18	464
Sachsen	1994	–	–20	444
Sachsen	1996	+1	–6	439
Thüringen	1998	+1	–	440
Niedersachsen	2001	–1	–	439
Sachsen-Anhalt	2007	–	–10	429
Sachsen	2008	–4	–12	413
Nordrhein-Westfalen	2009	–1	–	412
Mecklenburg-Vorpommern	2011	–4	–6	402

Solche Gebietsreformen bestehen entweder in einfachen Zusammenlegungen von Kreisen (kreisscharfe Gebietsänderungen) oder aber in Umverteilungen von Gemeinden auf bereits bestehende oder neugebildete Kreise (nicht kreisscharfe Gebietsänderungen). Im Rahmen der umfassenden Gebietsreform in Sachsen-Anhalt im Jahr 2007 wurden beispielsweise die Kreise „Bördekreis“ und „Ohrekreis“ zum Landkreis Bördekreis zusammengefasst. In diesem Fall handelt es sich um eine einfache, kreisscharfe Gebietsreform, die anhand sich ändernder Kreisschlüssel nachvollzogen werden kann. Hingegen wurde der Landkreis „Anhalt-Zerbst“ zu ungleichen Teilen auf vier (zum Teil neue) Landkreise verteilt (siehe Abbildung 1). In diesem Fall handelt es sich um eine gemeinde-, nicht aber um eine kreisscharfe Gebietsänderung, bei der zusätzliche Informationen auf der Ebene von Gemeinden herangezogen werden müssen, um die Änderungen nachvollziehen zu können.

Bei der Zusammenführung von Individual- und Kontextdaten stellen sich im Zusammenhang mit Gebietsreformen häufig drei zentrale Probleme: Mit jeder Gebietsreform ändert sich häufig

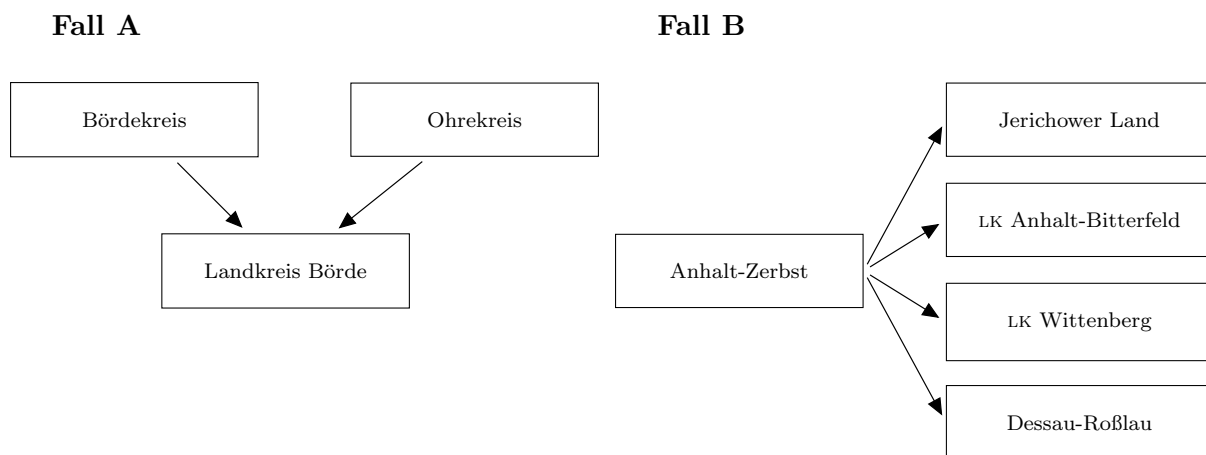


Abbildung 1: Reform der Kreisgebiete in Sachsen-Anhalt im Jahr 2007

(1) der Schlüssel, der eine Zuspiegung regionaler Indikatoren erlaubt, und (2) auch der sozio-ökonomische Kontext, in dem Individuen leben, lernen und arbeiten. Aus diesem Grund ist eine möglichst korrekte Zuordnung der befragten Personen zu dem regionalen Kontext erforderlich, der zum Zeitpunkt der Befragung relevant ist (oder, im Falle von Längsschnittdaten, dem Zeitfenster der Analysen entspricht). Aus diesem Grund ist nicht nur (a) eine Harmonisierung sich verändernder Kreisschlüssel erforderlich, sondern zudem (b) eine Harmonisierung der Zuordnung der Studienteilnehmer\_innen des NEPS auf den jeweils relevanten Gebietstand. Aktuell liegen den Nutzer\_innen des NEPS allerdings keine AGS für die Befragungsdaten vor, so dass der Harmonisierungsschritt einer korrekten Zuordnung der Studienteilnehmer\_innen im Falle nicht kreisscharfer Gebietsreformen (noch) nicht möglich ist (siehe Fall B in Abbildung 1). Ein weiteres Problem betrifft (3) die Regionalindikatoren selbst, die sich nicht selten auf einen anderen Gebietstand beziehen als den, der den Befragungsdaten zugrunde liegt. Sofern Gemeinden im Zuge von Gebietsreformen umverteilt werden, ändern sich nämlich auch die relevanten Populationen für die Berechnung regionaler Indikatoren.

## 2.1 Änderungsarten und Möglichkeiten der Harmonisierung

Für die Anpassung der in den Befragungsdaten des NEPS verwendeten Schlüssel (AGS 2001) an einen anderen Gebietstand sollten Gebietsveränderungen idealerweise auf der Ebene von Gemeinden nachvollzogen werden. Auf diese Weise können bei der Harmonisierung der Kreisschlüssel nicht nur Schlüsseländerungen berücksichtigt, sondern es kann auch die korrekte Zuordnung der Studienteilnehmer des NEPS zu Kreisen bzw. kreisfreien Städten sichergestellt werden. Gemeinden bilden die kleinste administrative Einheit und werden von den statistischen Ämtern des Bundes und der Länder entweder über Amtliche Gemeindegemeinschaften (AGS (achtstellig) oder Regionalschlüssel (zwölfstellig) kodiert. In beiden Fällen folgt die Kodierung der hierarchischen Gebietsordnung: Die ersten beiden Ziffern entsprechen dem Bundesland, die dritte Ziffer dem Regierungsbezirk, die vierte und fünfte Ziffer dem Kreis und die letzten drei Ziffern der Gemeinde. Regionalschlüssel geben zusätzlich über Gemeindeverbände Auskunft, die der sechsten bis neunten Ziffer entsprechen. So lautet beispielsweise der aktuelle Regionalschlüssel der Gemeinde Universitätsstadt Siegen 05 9 70 0040 040; diese Gemeinde (040) befindet sich im Kreis Siegen-Wittgenstein (70), des Regierungsbezirks Arnsberg (9) und des Bundeslands Nordrhein-Westfalen (05). Es handelt sich um eine verbandsfreie Gemeinde, identifizierbar über den mit der Zahl 0 beginnenden Verbandsschlüssel in Kombination mit dem dreistelligen Gemeindegemeinschaften.

Das Statistische Bundesamt dokumentiert Gebietsänderungen im Gemeindeverzeichnis-Informationssystem (GV-ISys) und veröffentlicht jährlich eine Gebietsänderungsliste (u. a. Statistische Ämter des Bundes und der Länder 2008) mit Informationen zu abgebenden, aufnehmenden und neugebildeten Gemeinden und deren übergeordneten Verwaltungsebenen. Die Gebietsänderungslisten enthalten, neben den AGS und den Gemeindegemeinschaften der abgebenden oder aufgelösten Gemeinde und der aufnehmenden oder neugebildeten Gemeinde, die Art der Änderung: (1) Auflösung, (2) Teilausgliederung, (3) Schlüsseländerung oder (4) Namensänderung. Nicht alle diese Änderungen können bei einer Harmonisierung von KKZ auf der Basis von AGS berücksichtigt werden:

**Auflösung oder Schlüsseländerung** Bei einer Gemeindeauflösung (1) wird die aufgelöste Gemeinde einer anderen Gemeinde zugeordnet; damit geht immer auch eine Umkodierung der Schlüssel einher. Schlüsseländerungen (3) hingegen sind das Ergebnis einer Umverteilung von Gemeinden auf höherer Aggregatebene, u. a. auch auf der Ebene von Kreisen. In beiden Fällen können die AGS entsprechend der Informationen der jährlich veröffentlichten Gebietsänderungslisten rekodiert werden, was in Abschnitt 2.2 beschrieben wird.

**Teilausgliederung und Namensänderung** Teilausgliederungen (2) von Gemeinden können bei der Harmonisierung nicht berücksichtigt werden, da keine Informationen darüber vorliegen,

welchem Teil der gesplitteten Gemeinde die Studienteilnehmer\_innen des NEPS zugeordnet sind. In diesen Fällen wird der AGS der abgebenden Gemeinde beibehalten. Da wir uns für eine Harmonisierung auf Kreisebene interessieren, ist dieses Vorgehen weitestgehend unbedenklich, da Teilausgliederungen von Gemeinden zumeist keine Auswirkungen auf die KKZ haben. Schließlich werden auch Namensänderungen (4) von Gemeinden nicht berücksichtigt.

## 2.2 Rekodierung der AGS und Ableitung von KKZ

Für die Rekodierung der AGS müssen diese als numerische Variable vorliegen, wohingegen die Generierung der KKZ aus den AGS eine String-Variablen mit acht Zeichen voraussetzt – im Falle von Gemeinden, die den Bundesländern 01–09 angehören, ist der numerische AGS siebenstellig, für die Bundesländer mit den Kennungen 10–16 achteinstellig. Beispielsweise ist die Konvertierung der numerischen Variable „AGS06“ in eine solche String-Variablen über folgenden Befehl möglich:

```
tempvar help06
tostring ags06, gen('help06')
replace 'help06' = "0" + 'help06' if length('help06') == 7
```

Schließlich kann zur Erzeugung der KKZ die achteinstellige String-Variablen nach den ersten fünf Stellen abgeschnitten und wieder in eine numerische Variablen überführt werden, um eine Zuspiegelung administrativer Daten zu ermöglichen.

```
tempvar help06_kkz
gen 'help06_kkz' = substr('help06', 1, 5)
destring 'help06_kkz', gen(kkz06)
lab var kkz06 "Kreiskennziffer '06"
```

Für die Rekodierung des „AGS06“ bzw. der „KKZ06“ entsprechend des Gebietsstands 2007 werden alle Auflösungen (Änderungsart 1) und Schlüsseländerungen (Änderungsart 3) auf Gemeindeebene berücksichtigt. Da Gemeinden theoretisch mehrfach im Jahr reformiert werden können und praktisch zum Teil auch werden, ist es erforderlich, das exakte Datum der Reform bei der Rekodierung der Schlüssel in Rechnung zu stellen. Es wird zwischen dem juristischen und dem statistischen Wirksamkeitsdatum unterschieden; zu beachten ist, dass die Zuordnung zu den einzelnen Gebietsänderungslisten uneinheitlich ist, d. h. teilweise nach dem juristischen, teilweise nach dem statistischen Wirksamkeitsdatum vorgenommen wird. Da wir an der Zuspiegelung von Regionalinformationen interessiert sind, müssen wir uns bei der Zuordnung zu den Gebietsänderungslisten an dem statistischen Wirksamkeitsdatum orientieren. Nichtsdestoweniger muss die Abfolge der Rekodierung gemäß dem juristischen Wirksamkeitsdatums erfolgen.

```
clonevar ags07 = ags06
lab var ags07 "Amtlicher Gemeindeschlüssel '07"
recode ags07 15101000 = 15001000 // Dessau > Dessau-Roßlau (1), 01.07.07 (jur.)
recode ags07 15151053 = 15001000 // Roßlau > Dessau-Roßlau (1), 01.07.07 (jur.)
recode ags07 15154010 = 15082015 // Greppin > Bitterfeld-Wolfen (1), 01.07.07 (jur.)
recode ags07 15154041 = 15082015 // Thalheim > Bitterfeld-Wolfen (1), 01.07.07 (jur.)
recode ags07 15367029 = 15089370 // Zens > Zens (3), 01.07.07 (jur.)
...
```



```
recode ags07 15089370 = 15089042 // Zens > Bördeland (3), 29.12.07 (jur.)
```

Die rekodierten AGS von 2007 bilden wiederum die Grundlage für den nächsten Harmonisierungsschritt, der die Gebietsreformen im Folgejahr berücksichtigt. Das zuvor beschriebene Verfahren kann so oft wiederholt werden, bis die AGS den gewünschten Gebietsstand repräsentieren. Abschließend können, wie oben beschrieben, die KKZ aus den AGS abgeleitet werden.

### 2.3 Zusammenführung der NEPS Daten mit Indikatoren auf Kreisebene

Eine Zusammenführung der NEPS-Daten mit regionalen Zeitreihendaten auf Kreisebene ist aktuell nicht über die AGS möglich, sondern ausschließlich über die KKZ. Anhand der diesem Working Paper beigefügten Umschlüsselungstabelle von KKZ zeigen wir im Folgenden beispielhaft, wie Zeitreihen von Arbeitslosenquoten auf der Ebene von Kreisen an die Individualdaten der Startkohorte 6 des NEPS angespielt werden können (die Tabelle ist für NEPS Nutzer\_innen über RemoteNEPS zugänglich). Dabei gehen wir von dem einfachen Fall aus, dass die Kreisindikatoren bereits als aufbereitete Zeitreihen mit einem einheitlichen Gebietsstand vorliegen (im folgenden Beispiel dem des Jahres 2007). Eine Möglichkeit jahresbezogene Indikatoren auf Kreisebene auf einen einheitlichen Gebietsstand zu bringen, wird in Kapitel 2.4 dargestellt.

Im NEPS liegen verschiedene Variablen mit Raumbezug vor, wie beispielsweise die KKZ des Schulortes (Variable `ts11202_g3R`) oder die des Ausbildungsortes (Variable `ts15207_g3R`). Der Gebietsstand aller dieser im NEPS verfügbaren Variablen entspricht der Variable `kkz2001` in der bereitgestellten Umschlüsselungstabelle. Die Variable `kkz2001` bildet daher den Ausgangspunkt für das Daten-Merging. Tabelle 2 zeigt einen Auszug aus der Umschlüsselungstabelle erneut mit Bezug auf die Kreisgebietsreform in Sachsen-Anhalt; exemplarisch führen wir lediglich eine Regionalvariable auf, nämlich die Arbeitslosenquote im Jahr 2009.

Tabelle 2: Auszug aus der bereitgestellten Tabelle mit den harmonisierten Kreiskennziffern von 1980–2011

KKZ2001	Kreisname2001	KKZ2007	Kreisname2007	Unemp2009	...
15151	Anhalt-Zerbst	15001	Dessau-Roßlau, Krf. S.	15,4	
15151	Anhalt-Zerbst	15082	Anhalt-Bitterfeld	14,0	
15151	Anhalt-Zerbst	15086	Jerichower Land	12,2	
15151	Anhalt-Zerbst	15091	Wittenberg	12,9	
:					
15154	Bitterfeld	15082	Anhalt-Bitterfeld	14,0	
15352	Aschersleben-S.	15085	Harz	13	
15352	Aschersleben-S.	15089	Salzlandkreis	13,7	
15355	Bördekreis	15083	Börde	10,3	
15362	Ohre-Kreis	15083	Börde	10,3	

In den Spalten `kkz2001` und `kkz2007` fällt auf, dass diese jeweils mehrere Einträge mit gleichen Kreiskennziffern enthalten, anhand derer die Änderungen durch die Kreisgebietsreform in Sachsen-Anhalt nachvollzogen werden können. Für ein Anspielen von Kreisdaten an die Individualdaten des NEPS ist jedoch ein einheitlicher Merging-Key erforderlich (hier die Variable `kkz2001`). Die teilweise nicht-kreisscharfe Gebietsreform in Sachsen-Anhalt führt dazu, dass nicht für alle Kreise eine einheitliche Zuordnung möglich ist. Im dargestellten Beispiel sind ausschließlich die kreisscharf reformierten Gebiete ‚Bördekreis‘, ‚Ohre-Kreis‘ und ‚Bitterfeld‘ eindeutig zuordbar. Die Studienteilnehmer\_innen im NEPS, die aus einem dieser Kreise kommen, können einfach der jeweils neuen Kreiskennziffer zugeordnet werden. Im Fall der nicht kreisscharfen Gebietsreform

‚Anhalt-Zerbst‘ und ‚Aschersleben-Staßfurt‘ können die Befragten jedoch, ohne Rückgriff auf die AGS, nicht exakt einer neuen Kreiskennziffer zugeordnet werden. Eine Möglichkeit, daraus resultierende fehlende Werte in den Kontextdaten zu vermeiden, besteht in der Verwendung sogenannter Umsteigeschlüssel. Bei Umsteigeschlüsseln handelt es sich entweder um flächen- (FPS), bevölkerungs- (BevPS) oder beschäftigtenproportionale (BeschPS) Gewichte, die darüber Auskunft geben, wie groß der Anteil eines Kreises ist, der in einen anderen Kreis überführt wurde, siehe exemplarisch Tabelle 3 mit den Gewichten für die Kreisgebietsreform in Sachsen-Anhalt.

Tabelle 3: Kreisänderungen mit Umsteigeschlüsseln zur Gewichtung regionaler Indikatoren (vgl. Laufende Raumbearbeitung des BBSR)

<b>Kreise 2001</b>	<b>Name</b>	<b>FPS</b>	<b>BevPS</b>	<b>BeschPS</b>	<b>Kreise 2007</b>	<b>Name</b>
15101	Dessau, Krf. S.	1	1	1	15001	Dessau-Roßlau, Krf. S.
15151	Anhalt-Zerbst	0,160	0,076	0,034	15086	Jerichower Land
15151	Anhalt-Zerbst	0,372	0,355	0,347	15091	Wittenberg
15151	Anhalt-Zerbst	0,413	0,366	0,480	15082	Anhalt-Bitterfeld
15151	Anhalt-Zerbst	0,055	0,203	0,139	15001	Dessau-Roßlau, Krf. S.
15154	Bitterfeld	1	1	1	15082	Anhalt-Bitterfeld
15159	Köthen	1	1	1	15082	Anhalt-Bitterfeld
15171	Wittenberg	1	1	1	15091	Wittenberg
15352	Aschersleben-S.	0,157	0,066	0,071	15085	Harz
15352	Aschersleben-S.	0,843	0,934	0,929	15089	Salzlandkreis
15358	Jerichower Land	1	1	1	15086	Jerichower Land

Auf Basis dieser Informationen können wir eine mehr oder weniger fundierte Entscheidung darüber treffen, welchem Teil der ehemaligen Kreise ‚Anhalts-Zerbst‘ und ‚Aschersleben-Staßfurt‘ die Studienteilnehmer\_innen des NEPS zugeordnet werden sollen. Am besten eignen sich hierfür die bevölkerungsproportionalen Gewichte. Bei der im Folgenden dargestellten Routine werden die Studienteilnehmer\_innen immer dem Kreis mit dem größten Bevölkerungsgewicht zugeordnet. Dieses Vorgehen ist im Falle der Gebietsreform ‚Anhalt-Zerbst‘ freilich problematisch, da der Kreis nahezu paritätisch auf die Kreise ‚Wittenberg‘, ‚Anhalt-Bitterfeld‘ und ‚Dessau-Rohrlau‘ aufgeteilt wurde.

```
keep KKZ2001 KKZ2007 Unemp2009
duplicates drop KKZ2001 KKZ2007, force
merge 1:1 KKZ2001 KKZ2007 using "\dataDir\Gewichte_2001_2007.dta"
tempvar rank
gsort KKZ2001 -BevPS
by KKZ2001, sort: gen 'rank' = _n
drop if 'rank' > 1
save "\dataDir\USING.dta"
```

Die auf diese Weise aufbereiteten Kreisdaten können schließlich über einen einfachen Merge-Befehl an die Befragungsdaten des NEPS angespielt werden. Exemplarisch können Kreisinformationen zum Schulort wie folgt hinzugespielt werden:

```
ren ts11202_g3R KKZ2001
merge m:1 KKZ2001 using "\dataDir\USING.dta", keep(master match) nogen
```

Das dargestellte Beispiel basiert auf der diesem Working Paper beigelegten Tabelle über Kreisgebietsänderungen von 1980 bis 2011. Zusätzlich wird eine Datei mit Umsteigeschlüsseln benötigt (hier: Gewichte\_2001\_2007.dta). Bevölkerungs- und flächenproportionale Gewichte für diverse Zeiträume seit 1990 sind auf Nachfrage beim Bundesamt für Bau-, Stadt- und Raumforschung (BBSR) erhältlich. Für frühere Zeitpunkte müssen die Gewichte selbst berechnet werden. Für eine detaillierte Beschreibung der Problematik bei Kreisänderungen und der Vorgehensweise bei der Erstellung der Gewichte siehe: BBSR-Berichte KOMPAKT 6/2010.

## 2.4 Gewichtung von Indikatoren auf Kreisebene

Umsteigeschlüssel können auch verwendet werden, um kreisscharf abgegrenzte Zeitreiheninformationen zu erhalten. Im Folgenden zeigen wir beispielhaft, wie anhand von Umsteigeschlüsseln eine auf einen einheitlichen Gebietsstand harmonisierte Zeitreihe erstellt werden kann. Wir veranschaulichen dies anhand von Arbeitslosenzahlen zum Gebietsstand im Jahr 2001 (alo01), die auf den Gebietsstand im Jahr 2007 (alo07) gebracht werden sollen. Da Arbeitslosenzahlen am Wohnort erhoben werden, verwenden wir für die Harmonisierung bevölkerungsproportionale Umsteigeschlüssel, mit denen zunächst alle Arbeitslosenzahlen aus dem Jahr 2001 auf Kreisebene multipliziert werden, siehe Tabelle 3. Im Falle kreisscharfer Gebietsänderungen entspricht das Gewicht dem Wert 1. Im Falle nicht-kreisscharfer Gebietsänderungen ergibt sich der konkrete Wert des jeweiligen Gewichts aus dem Anteil, zu dem ein Kreis aus dem Jahr 2001 einem anderen (ggf. neugebildeten) Kreis aus dem Jahr 2007 zugerechnet wird.

```
gen alo07weight = alo01 * BevPS
```

Um nun für jeden Kreis eine Arbeitslosenzahl zum Gebietsstand 2007 zu erhalten, müssen die für den jeweiligen Kreis relevanten (gewichteten) Werte aufsummiert werden. Für den Landkreis „Jerichower Land“ existieren beispielsweise zwei Einträge. Der erste Eintrag bezieht sich auf den Anteil des ehemaligen Kreises „Anhalt-Zerbst“, der im Zuge der Gebietsreform im Jahr 2007 dem Kreis „Jerichower Land“ zugewiesen wurde (BevPS 0,076). Der zweite Eintrag bezieht sich auf den Kreis „Jerichower Land“, der bereits vor der Kreisgebietsreform im Jahr 2007 bestand (der Gewichtungsfaktor beträgt 1). Im Falle des Kreises „Anhalt-Bitterfeld“ wird es etwas komplizierter. Dieser Kreis setzt sich im Jahr 2007 aus den beiden Kreisen „Bitterfeld“ und „Köthen“ sowie anteilig aus dem Kreis „Anhalt-Zerbst“ (BevPS 0,366) zusammen. Nun können alle doppelten Einträge für das Jahr 2007 aus dem harmonisierten Datensatz entfernt werden.

```
bysort kkz07: egen alo07 = total(alo07weight)
```

```
duplicates drop kkz07, force
```

## 3 Übersicht über raumstrukturelle Daten in Deutschland

Nachdem wir die technischen Voraussetzungen für das Anspielen von Kontextdaten an die Befragungsdaten des NEPS erläutert haben, geben wir im Folgenden einen Überblick über kommerziell und nichtkommerziell verfügbare raumstrukturelle Daten. Dabei konzentrieren wir uns insbesondere auf solche Daten, die als Zeitreihen vorliegen und bei denen eine Vergleichbarkeit über die Zeit gewährleistet ist. Letzteres ist dann gegeben, wenn Datenanbieter Regionalindikatoren in bereits aufbereiteten Zeitreihen zur Verfügung stellen, bei denen der Gebietsstand zu einem bestimmten Zeitpunkt „eingefroren“ wird (bei Fortsetzungsreihen handelt es sich häufig um das aktuelle Jahr). Regionale Zeitreihendaten sind insbesondere für Analysen im Längsschnitt von Bedeutung, bei denen individuelle Prozesse über ein längeres Zeitfenster hinweg beobachtet werden. In solchen Analysen ist es häufig sinnvoll, die jeweils regionalspezifische Makrostruktur

abzubilden, die sich parallel zu individuellen Veränderungen im Lebensverlauf wandelt (Wicht 2011 und insbesondere Windzio 2006).

Tabelle 4 gibt einen Überblick über kommerzielle und frei zugängliche Daten. Diese Übersicht kann aufgrund der Vielzahl von Anbietern (vor allem im kommerziellen Bereich), aber auch aufgrund der Veränderungen des Datenmaterials und -angebots, keinen Anspruch auf Vollständigkeit erheben. Die Recherche befindet sich auf dem Stand von September 2014.

Tabelle 4: Überblick über nicht-/teilkommerzielles und kommerzielles Datenangebot auf unterschiedlichen Aggregatebenen in Deutschland, Stand der URL 20.11.2014

Nicht-/teilkommerziell	Statistische Ämter des Bundes und der Länder: Genesis-Online Datenbank, URL	Kommerziell	Microm Consumer, URL
	Kommunale Bildungsdatenbank, URL		Nexiga, URL
	Bundesinstitut für Bau-, Stadt und Raumforschung (BBSR): Indikatoren zur Raum- und Stadtentwicklung in Deutschland und Europa (INKAR), URL		Acxiom, URL
	Statistik der Bundesagentur für Arbeit, URL		GfK, URL
	Bertelsmann Stiftung: Wegweiser Kommune, URL		Lutum+Tappert, URL
	Institut für Arbeitsmarkt- und Berufsforschung (IAB): Betriebs-Historik-Panel, URL		
	Deutsches Jugendinstitut (DJI): Regionaldatenbank, URL		
	Verband Deutscher Städtestatistiker (VDSt): Arbeitsgemeinschaft Kommunalstatistik (KOSTAT), URL		
	Bundesamt für Kartographie und Geodäsie (BKG): Kartenmaterial, URL		

**Statistische Ämter des Bundes und der Länder** Die statistischen Ämter des Bundes und der Länder sind die zentrale Anlaufstelle für regionale Daten. Die Genesis-Online-Datenbank enthält Jahresergebnisse der amtlichen Statistik für Gemeinden, Stadt- und Landkreise, Bundesländer und den Bund. Der Gebietsstand bezieht sich auf das Jahr für welches die Information jeweils vorliegt. Für eine Übersicht über die thematische Gliederung siehe Ditton (2014).

**Kommunale Bildungsdatenbank** Die kommunale Bildungsdatenbank ist ein vergleichsweise junges Projekt der statistischen Ämter des Bundes und der Länder (online seit Juni 2013). Die Datenbank trägt Informationen aus der Kinder- und Jugendhilfestatistik, Schulstatistik, Berufsbildungsstatistik sowie der Hochschulstatistik zur Darstellung zentraler Kennzahlen des regionalen Bildungsgeschehens zusammen. Aktuell liegen Informationen für die Schuljahre 2005/2006 bis 2012/2013 vor.

Über das frei zugängliche Angebot hinaus besteht die Möglichkeit, Anfragen an das statistische Bundesamt zu spezifischen Indikatoren zu richten. Diese werden i.d.R. an das zuständige Landesamt weitergeleitet und von diesem (ggfs. gegen Gebühr) bearbeitet. Die Verfügbarkeit von Indikatoren auf regionalen Gliederungsebenen variiert deutlich zwischen den Bundesländern – besonders für weit zurückliegende Zeitpunkte.

**Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR)** Das BBSR stellt jährlich „Indikatoren und Karten zur Raum- und Stadtentwicklung in Deutschland und in Europa“ (INKAR) als kostenpflichtige CD-Rom zur Verfügung. Die Kosten für eine Einzelplatzlizenz liegen bei 45 Euro. INKAR umfasst folgende Themenbereiche: Arbeitslosigkeit, Bauen und Wohnen, Beschäftigung und Erwerbstätigkeit, Bevölkerung, Bildung, Privateinkommen, Flächennutzung und Umwelt, medizinische und soziale Versorgung, öffentliche Haushalte, raumwirksame Mittel, Siedlungsstruktur, Sozialstruktur, Verkehr und Erreichbarkeit und Wirtschaft. Die Indikatoren zu den verschiedenen Themenfeldern liegen mehrheitlich als Zeitreihe ab dem Jahr 1995 (aktuell bis 2012) vor. Die Zeitreihen beziehen sich immer auf den aktuellen Gebietsstand und sind auf verschiedensten Aggregatebenen verfügbar. Die kleinste räumliche Gliederungseinheit bildet die der Gemeinden.

**Statistik der Bundesagentur für Arbeit** Die Statistik der Bundesagentur für Arbeit stellt im Rahmen ihres Online-Angebots regionale Informationen zu arbeitsmarktbezogenen Themen unentgeltlich zum Download zur Verfügung. Die räumlichen Gliederungsebenen, auf denen Daten vorliegen, umfassen politische Gliederungseinheiten (Kreise, Länder, Bund) sowie Gliederungseinheiten der Bundesagentur für Arbeit (Regionaldirektionsbezirke, Arbeitsagenturbezirke). Alle abrufbaren Informationen beziehen sich auf den jeweils aktuellen Gebietsstand. Die Informationen können entweder monatsgenau abgerufen werden oder als Zeitreihen. Zeitreihendaten sind für Westdeutschland ab 1984 und für Gesamtdeutschland ab 1996 (jeweils bis 2008) auf Kreisebene verfügbar. Arbeitsmarktinformationen sind auf den verschiedenen regionalen Gliederungsebenen zudem für frühere Zeitpunkte in Papierform in den „Amtlichen Nachrichten der Bundesagentur für Arbeit“ (ANBA) erhältlich.

**Wegweiser Kommune** Der Wegweiser Kommune ist ein Angebot der Bertelsmann Stiftung, das als Informationsplattform für Kommunen gedacht ist und u. a. die Abfrage sozio-struktureller Indikatoren für (einzelne) Städte, Gemeinden und Kreise ermöglicht. Wie weit diese Informationen zurückreichen, ist abhängig vom gewählten Indikator, teilweise sind Angaben ab 2003 (bis 2011) verfügbar. Über einen Datennutzungsvertrag sind diese Informationen zu Forschungszwecken im Excel-Format erhältlich. Die kleinste Gliederungseinheit stellen Gemeinden und Gemeindeverbände mit Gemeinden über 5000 Einwohner dar.

**Institut für Arbeitsmarkt- und Berufsforschung (IAB)** Das IAB stellt das Betriebs-Historik-Panel (BHP) zusammen. Dieses setzt sich aus Querschnittsdatensätzen zusammen; für Westdeutschland seit 1975 und für Gesamtdeutschland seit 1991. Enthalten sind Informationen über alle Betriebe des gesamtdeutschen Raums. Für Regionalanalysen sind insbesondere die Informationen über die Standorte von Betrieben interessant. Stichproben des Gesamtdatenbestandes können im Rahmen eines Gastaufenthaltes oder via Fernzugriff verarbeitet werden.

**Regionaldatenbank des Deutschen Jugendinstituts (DJI)** Die Regionaldatenbank des DJI enthält nach thematischen Schwerpunkten zusammengefasste Informationen aus der amtlichen Statistik. Besonders hervorzuheben sind hier die Informationen zu bildungsbezogenen Themenstellungen. Die bereits 1987 aufgebaute Datenbank wird laufend aktualisiert.

**Arbeitsgemeinschaft Kommunalstatistik (KOSTAT)** Die Arbeitsgemeinschaft KOSTAT sammelt kleinräumig gegliederte Daten und bietet inhaltliche Indikatoren (z.B. Einwohnerzahlen für spezifische soziale Gruppen) sowie Grenzgeometrien der statistischen Bezirke für eine Vielzahl von KOSTAT-Projekten verschiedener Städte (z.B. Straßenverzeichnis) an. Der Datenbestand umfasst jährliche Fortschreibungen seit 1988. Externe Nutzer können gegen ein Nutzungsentgelt, welches sich nach dem gewünschten Datenumfang richtet, Zugriff erlangen.

**Bundesamt für Kartographie und Geodäsie (BKG)** Das BKG stellt in seinem Downloadbereich Kartenmaterial (sog. Shapefiles) auf der Ebene von Kreisen, Regierungsbezirken und Bundesländern für Gesamtdeutschland bereit. Diese Datensätze können mit Indikatoren auf den entsprechenden Gliederungsebenen verknüpft werden und ermöglichen die Erstellung von

Karten zur illustrativen Darstellung von Forschungsergebnissen. Kartenmaterial kann mittels spezieller Geodaten-Software (z.B. ArcGIS), aber auch mit statistischer Software (z. B. R und Stata) bearbeitet werden (Drukker, Prucha & Raciborski 2013; Pisati 2008).

Die Datenpakete, die über kommerzielle Anbieter erhältlich sind, liegen zumeist auf sehr kleinräumigen Ebenen vor, wie beispielsweise Postleitzahlenbezirke, Wohnquartiere oder Straßenabschnitte. Die räumliche Gliederung entspricht zumeist einem eigens vom jeweiligen Anbieter entwickelten Konzept. Die zur Zusammenstellung der Daten genutzten Quellen sind vielfältig: Informationen statistischer Ämter und Verwaltungsbehörden sowie verschiedener Wirtschafts- und Marktforschungsinstitute fließen hier ein. Indikatoren und Kartenmaterial für verschiedene räumliche Gliederungsebenen sind bei den jeweiligen Anbietern erhältlich. Die Kosten für diese Daten sind i.d.R. hoch. Zahlreiche Indikatoren von Microm Consumer Marketing sowie von RegioInfas werden in bereits aufbereiteter Form vor Ort am FDZ des NEPS für die verschiedenen Startkohorten des NEPS zur Verfügung gestellt. Allerdings ist eine regionale Identifizierung bzw. Verortung der Studienteilnehmer\_innen des NEPS ausgeschlossen. Weitere Informationen zu den im NEPS enthaltenen microm Indikatoren sind in den Studiendokumentationen der einzelnen Startkohorten zu finden. Für weitere Informationen zu den RegioInfas Daten im NEPS gibt Koberg (2012) einen Überblick.

## 4 Ausblick

Für die Bildungs- und Arbeitsmarktforschung stellt die Verfügbarkeit regionaler Kontextdaten (auch als Zeitreihen) eine bedeutende Ressource dar, insofern sich damit soziale Phänomene im zeitlichen Verlauf und unter Berücksichtigung ihrer räumlichen Verortung untersuchen lassen. Aus dieser Perspektive sind Räume als Handlungsräume aufzufassen: sie bilden Gelegenheitsstrukturen, die Bildungsprozesse ermöglichen oder behindern. Das NEPS zeichnet sich in diesem Zusammenhang als einzigartige Datenquelle aus, die es erlaubt, einerseits individuelle Bildungserwerbsprozesse und Kompetenzentwicklungen über die Zeit umfassend zu untersuchen und andererseits gleichzeitig Befragungsdaten mit räumlichen Kontextdaten auf verschiedenen Aggregatenebenen zusammenzuführen. Das Ziel des vorliegenden Arbeitspapiers ist es, Datennutzer\_innen des NEPS einen ersten Überblick über Probleme und Möglichkeiten der Nutzung regionaler Daten insbesondere auf der Ebene von Kreisen und kreisfreien Städten zu geben, da diese dem derzeit kleinsten im NEPS verfügbaren administrativen Aggregationslevel entspricht. Im Rahmen dieser Übersicht haben wir die Problematik von Gebietsänderungen im Zeitverlauf dargestellt und Möglichkeiten der Harmonisierung von Gebietsständen sowie von Regionalindikatoren erläutert. Zudem haben wir einen Überblick über die in Deutschland (frei) verfügbare raumstrukturelle Daten gegeben, die sich zum Anspielen an die Befragungsdaten des NEPS besonders eignen.

Grundlegend stellt sich bei der Verwendung raumstruktureller Daten zur Analyse von Kontexteinflüssen sowohl theoretisch als auch methodisch die Frage, inwieweit regionale Kontexte Handlungskontexte tatsächlich abbilden. Anders ausgedrückt, welche räumlichen Dimensionen ein theoretisch vermuteter Mechanismus annimmt. Administrative Gebietseinheiten, wie etwa Landkreise und kreisfreie Städte, stellen sich für die Untersuchung solcher Fragestellungen als eher unzureichend heraus, insofern zwischen derartigen administrativen räumlichen Einheiten Interdependenzen bestehen können, die in den Analysen unberücksichtigt bleiben. Bei speziellen statistischen räumlichen Analyseverfahren (u. a. Elhorst 2014, Anselin 1988) rücken gerade Interdependenzen zwischen Gebietseinheiten und Personen in verschiedenen Gebietseinheiten in den Vordergrund. Dabei werden Handlungsräume weniger als administrativ vordefinierte Gebiete mit starren Grenzen aufgefasst, sondern vielmehr als Kontinuum, dessen Einfluss sich über die Berücksichtigung räumlicher Distanzen und regionaler Nachbarschaften abbilden lässt (Weßling, Hartung & Hillmert 2014) und/oder über empirisch beobachtete Infrastrukturen, wie beispielsweise Pendlerverflechtungen (Wicht 2015). Regionale Analysen können in beträchtlichem

Maße davon profitieren, wenn sie bestehende räumliche Interdependenzen reflektieren.

## 5 Literatur

- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- Bundesinstitut für Bau-, Stadt- und Raumforschung im Bundesamt für Bauwesen und Raumordnung (BBSR). 2010. *Gebietsreformen – politische Entscheidungen und Folgen für die Statistik*. BBSR-Berichte KOMPAKT (6/2010).
- Ditton, H. 2014. *Bildung und Region* (NEPS Working Paper No. 44). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Drukker, D. M., I. R. Prucha & R. Raciborski. 2013. *Creating and Managing Spatial-Weighting Matrices With the Spmat Command*. *The Stata Journal* 13(2): 242–386.
- Elhorst, J. P. 2014. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. London: Springer.
- Koberg, T. 2012. *Starting Cohort 2, 3, 4 and 6 of the National Educational Panel Study: Regio-Infas (infas geodaten), Technical Report (Data Manual [Supplement], v.0.3)*. Bamberg: Nationales Bildungspanel (NEPS), Universität Bamberg.
- Pisati, M. 2008. *SPMAP: Stata Module to Visualize Spatial Data*. EconPapers. Online verfügbar.
- Statistische Ämter des Bundes und der Länder (Hrsg.). 2008. *Daten aus dem Gemeindeverzeichnis. Gebietsänderungen vom 01.01. bis 31.12.2007*. Wiesbaden: Statistisches Bundesamt.
- Weßling, K., A. Hartung & S. Hillmert. 2014. *Spatial Structure Counts: The Relevance of Regional Labour-market Conditions for Educational Transitions After Compulsory Schooling*. Working-Paper, Universität Tübingen. Online verfügbar.
- Wicht, A. 2011. *Zwischen Ausbildung und Arbeitsmarkt. Regionale Unterschiede beim Übergang von der beruflichen Erstausbildung in Beschäftigung*. Siegen: Universität, MA Thesis. Online verfügbar.
- Wicht, A. 2015. *Von Orten zu Räumen: Regionale Kontexte in der Bildungssoziologie*. In: T. Coelen, A. J. Heinrich, A. Million (Hrsg.). *Stadtbaustein Bildung*. Wiesbaden: vs.
- Wikipedia. 2014. *Gebietsreform*. Online verfügbar (14.04.2014).
- Windzio, M. 2006. *The Problem of Time Dependent Explanatory Variables at the Context-Level in Discret Time Multilevel Event History Analysis. A Comparison of Models Considering Mobility Between Local Labour Markets as an Example*. *Quality & Quantity* 40(2), 175–185.